Enzymes chimériques pour la catalyse énantiosélective de réactions en cascade

Anna Christine Eid

THÈSE DE DOCTORAT

NNT : 2022UPASF035

Chimeric enzymes for the catalysis of enantioselective cascade reactions

Enzymes chimériques pour la catalyse énantiosélective des réactions en cascade

Thèse de doctorat de l'université Paris-Saclay et de l'université Libanaise

École doctorale n° 571 : Sciences Chimiques : Molécules, Matériaux, Instrumentation et Biosystèmes (2MIB)
Spécialité de doctorat : Chimie
Graduate School : Chimie. Référent : Faculté des sciences d'Orsay

Thèse préparée dans les unités de recherche : Institut de Chimie Moléculaire et des Matériaux d'Orsay (Université Paris-Saclay, CNRS) LG2E (Université Libanaise) sous la direction de **Jean-Pierre Mahy**, Professeur, Université Paris-Saclay, le co-encadrement de **Wadih Ghattas**, Chargé de recherche CNRS, Université Paris-Saclay, et la co-direction de **Bassam Badran**, Professeur, Université Libanaise

Thèse soutenue à Paris-Saclay, Orsay, le 3 juin 2022, par

Anna Christine EID

Composition du Jury

Agathe URVOAS Professeure, Université Paris Saclay	Présidente
Marianne ILBERT Chargée de recherche CNRS, HDR, Université Aix-Marseille	Rapporteuse & Examinatrice
Thierry TRON Directeur de recherche, Université Aix-Marseille	Rapporteur & Examinateur
Mira MAALOUF Maître de conférences, Université Catholique de Lyon	Examinatrice
Jean-Pierre MAHY Professeur, Université Paris-Saclay	Directeur de thèse

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences Chimiques: Molécules, Matériaux, Instrumentation et Biosystèmes (2MIB)

Titre : Enzymes chimériques pour la catalyse énantiosélective de réactions en cascade.

Mots clés : Métalloenzymes ; métalloenzymes artificielles ; enzyme chimérique ; réactions en cascade.

Résumé : Dans le cadre de la recherche de procédés chimiques plus respectueux de l'environnement, les enzymes artificielles constituent des catalyseurs de choix pour la synthèse de produits d'intérêt pour la chimie fine dans des conditions éco-compatibles (P, T, H_2O). En s'inspirant des réactions enzymatiques multiples, naturelles et artificielles, l'objectif de ce projet était de préparer une enzyme chimérique possédant deux sites actifs, par fusion des domaines catalytiques de deux enzymes apparentées de la super-famille des Cupines : La Pirin Like Protein (PLP) et la Mannose-6-Phosphate Isomérase (PMI). Ces deux bicupines sont constituées de deux domaines composés chacun d'un tonneau de quatre feuillets β, qui sont reliés par un seul inter-domaine composé d'hélices α. Nous avons édité les vecteurs codant pour les protéines ciblées His_6 taguées en y introduisant le site de clivage de la protéase TEV.
PMI et PLP ont pu être exprimées et purifiées. Après l'extraction de leurs cations métalliques natifs par de l'EDTA, leur transformation en de nouvelles métalloprotéines artificielles par fixation de cuivre(II) a été réussie et Cu-PMI et Cu-PLP ont pu être caractérisées par ITC, RPE et Fluorescence. L'enzyme chimérique combinant les domaines catalytiques de la PMI et de la PLP a ensuite pu être préparée. La stratégie suivie a commencé par des études de modélisation structurale permettant d'identifier le site de fusion adéquat des deux domaines. Ensuite, le vecteur d'expression codant pour la chimère a été construit puis la chimère a été exprimée dans E.*coli* BL21. Cependant, une optimisation de la purification reste à faire. Les perspectives de l'utilisation de l'enzyme chimérique PMI-PLP résident dans la catalyse eco-compatible de réactions en cascade avec une économie d'atome.

Title : Chimeric enzymes for the catalysis of enantioselective cascade reactions.

Keywords : Metalloenzymes; artificial metalloenzymes; chimeric protein; cascade reactions

Abstract : In the context of ecofriendly chemistry, artificial enzymes are promising catalysts that should enable the synthesis of fine chemicals. Inspired by natural and artificial enzymatic catalysis, we aimed to prepare an artificial chimeric enzyme that possesses two active sites by the fusion of catalytic domains of two related enzymes that belong to the cupin superfamily of proteins. Pirin Like Protein (PLP) and Mannose-6-Phosphate Isomerase (PMI) are similar bicupins characterized by a structure formed of two beta barrel domains. In this project, the vectors encoding for PMI and PLP comprising a His_6 tag were edited by introducing a TEV protease cleavage site. PMI and PLP proteins have been successfully expressed and purified. After removal of their native metal cations by EDTA, their transformation into a copper metalloprotein was successful. Cu-PMI and Cu-PLP were characterized by ITC, EPR and Fluorescence. Then, the chimeric enzyme was designed by combining the catalytic domains of PMI and PLP. The strategy we followed to prepare the PMI-PLP bicupin chimera started first by structural modeling studies that allowed us to identify the most favorable site for the fusion of the two domains. Then, the expression vector containing the DNA encoding for the PMI-PLP chimera was constructed, and the chimera was successfully expressed in E.*coli* BL21 cells. However, an optimization of the purification remains to be done. As perspectives, the chimeric PMI-PLP enzyme would be used to catalyze reactions in cascade with maximized atom economy under ecofriendly conditions.

ACKNOWLEDGMENTS

First and foremost, I would like to express my special appreciation and thanks to my supervisor Dr. Wadih GHATTAS for his warm welcome, his patient guidance, his encouragement and his support. I have been extremely lucky to have a clever and unique supervisor who cared so much for my work. His dynamism, vision, sincerity, motivation and expertise have deeply inspired me. It was a great honor to work under his guidance. I would also like to thank him for his friendship, empathy and great sense of humor.

The completion of this study could not have been possible without the expertise of Prof. Jean-Pierre MAHY without whom I would never have succeeded. He is the kindest and the smartest director I have ever had. Thank you for taking the time to discuss with me and helping me.

I must also express my deep and sincere gratitude to Prof. Bassam BADRAN who made this internship possible. Special thanks are due to Dr. Ziad ABI KHATTAR for his friendly guidance, constant encouragement and expert advice.

I take the opportunity to acknowledge the committee members: Prof. Thierry TRON, Dr. Marianne ILBERT, Prof. Agathe URVOAS and Dr. Mira MAALOUF who accepted to examine my thesis. I thank you for giving me the opportunity to discuss my work with you and I am grateful for your comments and suggestions.

I would like to thank ICMMO, Université Paris-Saclay and in particular Prof. Laurent Salmon, for providing excellent working conditions, as well as to all the members of LCBB for their kindness, fun and support.

Completing this work abroad would have been more difficult without the support, keen interest and friendship provided by all my friends here and overseas. **Special Thanks** are due to *Bassel, Mariam, Abed, Valène, Joe, Mimi, Reem, Ghenwa, Samantha and Eddy*. Thank you for always cheering me up.

Last, but not least, a special thanks to my family: *MOM & DAD, Lello, Anto, Antoinette, Juju, Imad, Zaza, Jiji and Maro*. Words cannot express how grateful I am for all the sacrifices. Thanks for encouraging me to be independent and having confidence in myself. Your prayers were what sustained me thus far.
I LOVE YOU VERY VERY VERY VERY MUCH.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABTS	2,2'-azino-bis 3-ethylbenzothiazoline-6-sulfonic acid
ACCO	1-amino-cyclopropane oxidase
ACC	1-Aminocyclopropane-1-carboxylic acid
AcTEV	Active Tobacco Etch Virus protease
ADH	Alcohol dehydrogenase
Aa	Amino acids
AS	Enzyme specific activity
Asp	Aspartic acid
BLA	Betta lactamase
CaM	Calmodulin domain
CPA	Carboxypeptidase A
CD	Circular dichroism
Cf	Final concentration
Ci	Initial concentration
Co-S	Co-substrate
CP	Cyclopentadiene
CYP51	Cytochrome P450
DmPMI	Glucose-6-phosphate isomerase of *Dehalococcoides mccartyi*
Equiv.	Equivalent
EDTA	Ethylenediaminetetraacetic acid
EPR	Electron paramagnetic resonance
ETBr	Ethidium bromide
F6P	Fructose-6-phosphate
FPMOD	Fusion Protein MODeller
G6P	Glucose-6-phosphate
G6PDH	Glucose-6-phosphate dehydrogenase

Glu	Glutamic acid
Gln	Glutamine
GFP	Green Fluorescent Protein
GDP	Guanosine diphosphate
GMP	Guanosine monophosphate
HEI	High energy intermediate
His	Histidine
His_6-LcZ-PLP	His_6-PLP protein fused to a portion of LacZ
His_6-MBP-TEV	His_6-Maltose Binding Protein-TEV
His_6-	His_6 tagged
HGO	Homogentisate dioxygenase
IL-12	Interleukin-12
Inf	Insoluble fraction
ITC	Isothermal titration calorimetry
KLD	Kinase, Ligase and Dpn1
LB	Lysogeny Broth
MBP	Maltose binding protein
M6P	Mannose-6-phosphate
MW	Molecular weight
NEB	New England Biolab
PGI	Phosphoglucose isomerase
PMI	Phosphomannose isomerase
PLP	Pirin Like Protein
SD	Shine Dalgarno
Sf	Soluble supernatant fraction
SUMO	**S**mall **U**biquitin-like **Mo**difier
TB	Terrific Broth
TMB	3,3',5,5'-tétraméthylbenzidine

$TEV_{c/site}$	TEV protease cleavage site
TEV	Tobacco Etch Virus
TBE	Tris Borate EDTA
TNF	Tumor necrosis factor
6PGL	6-phosphogluconolactone
5PAHz	Phospho-D-arabinonhydrazide

INTRODUCTION

INTRODUCTION

A. GENERAL CONTEXT

A.1. Searching for new bio-compatible biocatalysts

In the context of sustainable green chemistry, enzyme-catalyzed chemical reactions represent the best eco-friendly scenario in organic synthesis. Enzymes can be used under eco-friendly conditions (in aqueous solution, at room temperature, under atmospheric pressure,) to obtain specific chiral products and avoid the formation of undesirable side-products.[1–3]

Natural enzymes are classified as the most powerful catalysts of natural organisms. They can catalyze efficiently different kinds of reactions. They have evolved for millions of years to reach efficient catalysis for the transformation of very specific substrates which at the same time limits their applications when using non-natural substrates. Additionally, enzymes can catalyze only limited types of reactions and desired product stereoisomers are not always accessible.[4] To overcome these major limitations of enzymes, the field of artificial metalloenzymes has evolved. [4]
In the context of ecofriendly chemistry, artificial metalloenzymes result from the incorporation of a catalytically competent metal cofactor into a protein scaffold. [4] The construction of these hybrid biocatalysts combining biological components and synthetic chemical components include genetic mutation, chemical modification and/or combination of these two methods.[5] Hybrid biocatalysts have several advantages such as: broad range of substrates and reactions, possibility of working in a chiral environment provided by the residue of the protein under mild operating conditions. [4]

With the fast advances of artificial metalloenzyme's research, as well as their advantages, they have the potential to become legitimate contenders to natural enzymes for their use in a wide range of fields, from chemical and food industries to diagnostics and medicine.[6] Thus, artificial metalloenzymes are categorized in the production of fuels, in biological reactions (Ribonuclease A mimics, ATPase mimics) and in green and sustainable chemistry (Diels Alder, Suzuki-Miyaura and others for chemical synthesis). [6] Different strategies were described over the years to construct artificial metalloenzymes. These strategies will be discussed in chapter 1, section 1.1.1. But in this project, a novel strategy seemed most appealing to us: the construction of chimeric enzymes.

A.2. Chimeric enzymes

Over the previous decade, many fields like biochemistry, protein chemistry, molecular cloning, random and site-directed mutagenesis, directed evolution of biocatalysts… have all made it possible to use a wide range of enzymes and microbial cultures as tools for organic synthesis.[7] The utilization of various classes of enzymes for the catalysis of a wide range of chemical reactions can result in a wide range of chiral molecules.[7]

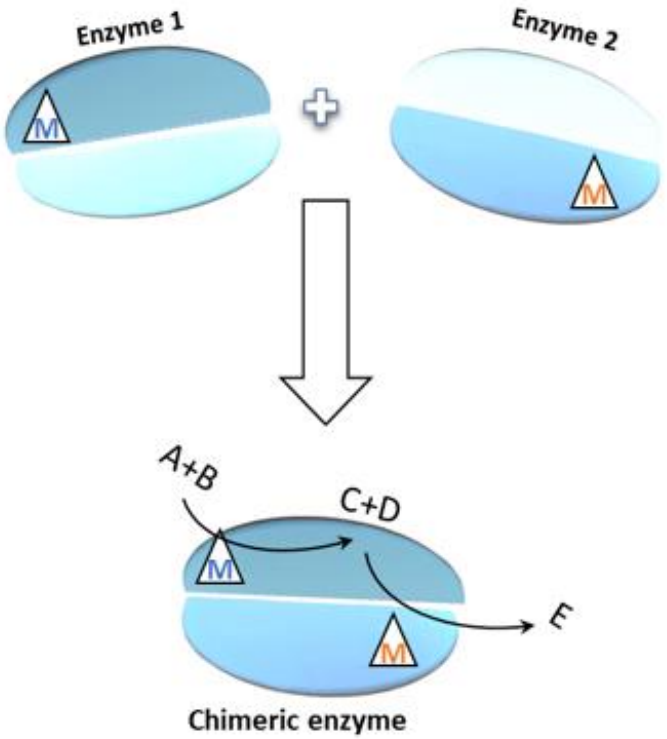
Fusion proteins, also known as chimeric proteins, are produced from the translation of a gene that combines two or more genes that previously coded for different or similar proteins. Indeed, when this fusion gene is translated, it produces a single polypeptide with functional features that are derived from those of the original proteins.[7]

Furthermore, chimeric proteins are advantageous because they reduce costs by replacing multiple enzymes by a single enzyme able to catalyze two or more reactions. Hence, they are advantageous in atom economy which is the second of the 12 green chemistry principles that aims to optimize the incorporation of all the atoms of the starting materials into the desired final product without generating side products.[8–10]

B. AIM OF THE PROJECT

In the context of the preparation of artificial enzymes for applications in chemical synthesis and inspired by natural and artificial enzymatic catalysis, the ambitious challenge of this project is to fuse structurally related enzymes possessing different but complementary activities to catalyze different reactions and reactions in cascade.

We aim to prepare artificial chimeric metalloenzymes that possess multiple active sites by the fusion of catalytic domains and the deletion of non-catalytic domains of two related enzymes. This chimeric enzyme will be used to catalyze reactions in cascade with maximized atom economy under ecofriendly conditions (scheme 1).



Scheme 1: Preparation of artificial chimeric metalloenzyme containing multiple active sites. M=Metals

To prepare the chimeric enzyme, we chose to rely on the cupin superfamily of proteins. Cupins are proteins characterized by a structure formed of a barrel of four β sheets containing the active site.[11]

However, artificial metalloenzymes comprising an active site formed by β sheets are rare and to understand the effect of active sites formed by β sheets on catalysis, it is

thus necessary to transform other cupins into artificial metalloenzymes and to study their activity.[12]

In this project, we selected two related enzymes from the bicupin family: Mannose-6-Phosphate Isomerase (PMI) (Figure 1A) and Pirin Like protein (PLP) (Figure 1B) are similar bicupins characterized by a structure formed of two β barrel domains, each possessing one active site in one of their two β-barrel domains. So, they are formed only of two domains composed each of one barrel of four β-sheets and connected by a single inter-domain consisting of α helixes.[13,14]

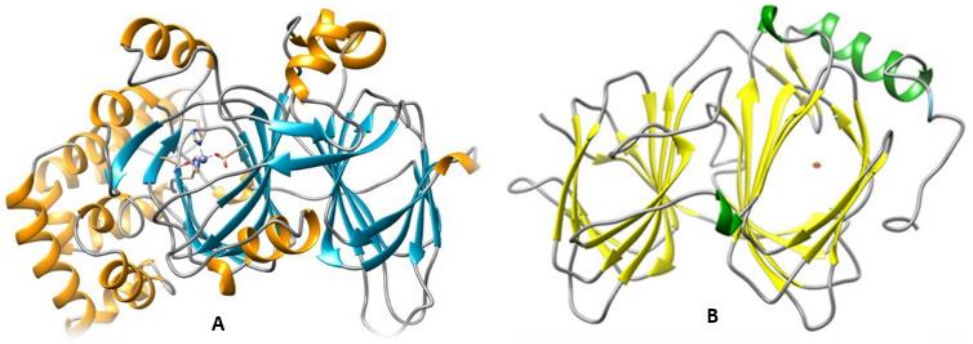


Figure 1: A) 3D structure of PMI B) Swiss model of PLP

The bicupin chimera PMI-PLP (Figure 2) will therefore be fused at the inter-domains of the two original bicupins.

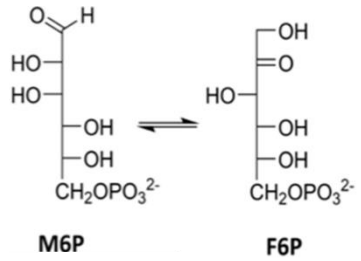
Figure 2: Predicted structure of PMI-PLP chimeric enzyme. In red: catalytic domain of PMI; in yellow: catalytic domain of PLP.

Then, inspired by natural and artificial enzymatic catalysis, the chimeric PMI-PLP enzyme will be used with their native and non-native metal cations to catalyze different

reactions and different cascades of reactions, involving for example Diels-Alder cycloadditions and hydrolysis.

B.1. Starting enzymes: PMI & PLP

PMI is a metal dependent aldose-ketose isomerase containing one atom of zinc in its active center that catalyzes the reversible isomerization of D-Mannose-6-phosphate (M6P) to D-Fructose-6-phosphate (F6P) in eukaryotic and prokaryotic cells (scheme 2).[13]



Scheme 2: Reversible isomerization of D-mannose-6-phosphate (M6P) to D-fructose-6-phosphate (F6P) catalyzed by PMI.

Recently in our laboratory, the crystal structure of mannose-6-phosphate isomerase of *Candida albicans* was obtained.[13] Aiming at the preparation of artificial enzymes, we substituted the Zn(II) cation of its active site by a Cu(II) cation. This substitution transformed the isomerase into an artificial metalloprotein.

PLP is present in Pseudomonas *stutzeri zobelli* and it is thought to contain one atom of iron in its active center. Its biological role remains unclear.[14] Its structure resembles that of the PMI. PLP has also been transformed into an artificial enzyme via the substitution of its Fe(III) cation by a Cu(II) cation, which provided it with quercetinase activity.[14]

The PMI active site is found in the β-sheet barrel domain of the C-terminal side of PMI, while the PLP active site is located in the β-sheet barrel domain of the N-terminal side of PLP. The superimposition of both starting enzymes has shown that the zinc-containing PMI's active center aligns with the empty cupin domain of PLP, while the iron-containing PLP 's active center aligns with the empty cupin domain of PMI (Figure 3).

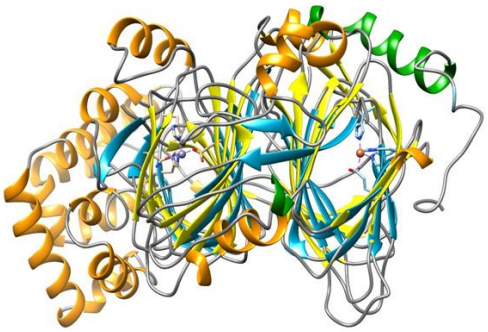
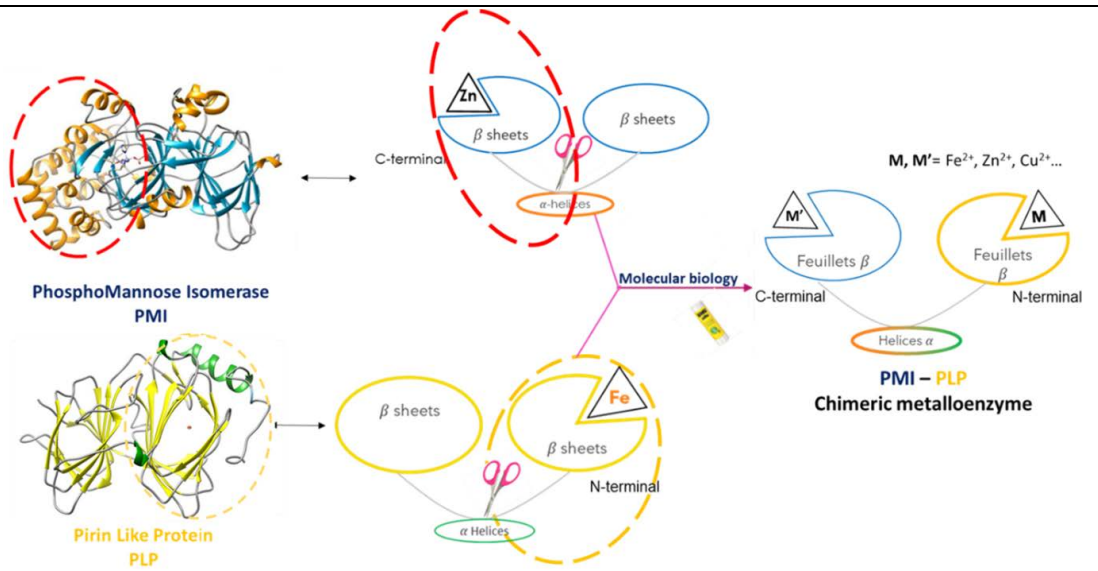


Figure 3: Superposition of PMI and PLP.

Thus, we plan to design a functional chimeric enzyme by combining the catalytic domains of PMI and PLP (C-terminal side of PMI containing the active site and the N-terminal side of PLP also containing its active site), which should avoid denaturation of the produced protein and should lead to a correct folding.

B.2. The strategy to produce the Chimeric enzyme

The strategy we followed to prepare the PMI-PLP bicupin chimera started first by structural modeling studies that allowed us to identify the linking point. We used *elNémo* to simulate the normal modes of vibration of all the atoms of both PMI and PLP, except hydrogen atoms, starting from the lowest frequency. We studied the movements of all atoms/domains in PMI and PLP and we visualized their movements. Molecular biology techniques were then used to design the plasmid for the production of the protein (scheme 3).



Scheme 3: Design of a chimeric PMI-PLP enzyme that would be able to catalyze cascade reactions.

CHAPTER 1

STATE OF THE ART

CHAPTER 1: STATE OF THE ART

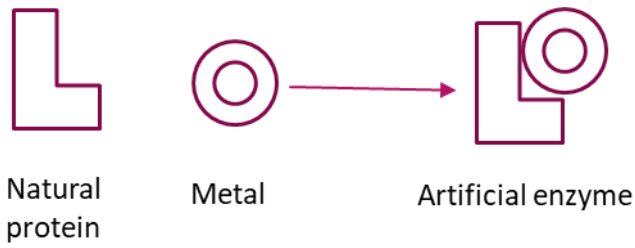
1.1. ARTIFICIAL METALLOENZYMES

Enzymes are proteins folded into three-dimensional structures in most of which hydrophobic cores are surrounded by hydrophilic shells. This fold allows reactions to occur inside the hydrophobic portions, which gives the specificity of enzymes.[6] In most cases, enzymes require the use of co-substrates *i.e.* cofactors which often causes an increase in the cost of production. [6]

As mentioned in the introduction, enzymes can catalyze only limited types of reactions and desired stereoisomers are not always accessible. To overcome these hurdles, **artificial metalloenzymes** have been developed.[4]

Artificial metalloenzymes result from the incorporation of a catalytically competent metal cofactor into a protein scaffold (Scheme 4). The construction of these hybrid catalysts includes genetic mutations, chemical modifications and/or combination of these two methods.[5]

These powerful catalysts are conveyed with extraordinary catalytic properties provided by the non-native metal cofactor and with a protective and stereoselective environment provided by the hosting protein.[16]



Scheme 4: Preparation of artificial enzyme by incorporating synthetic transition metal catalysts into natural protein.

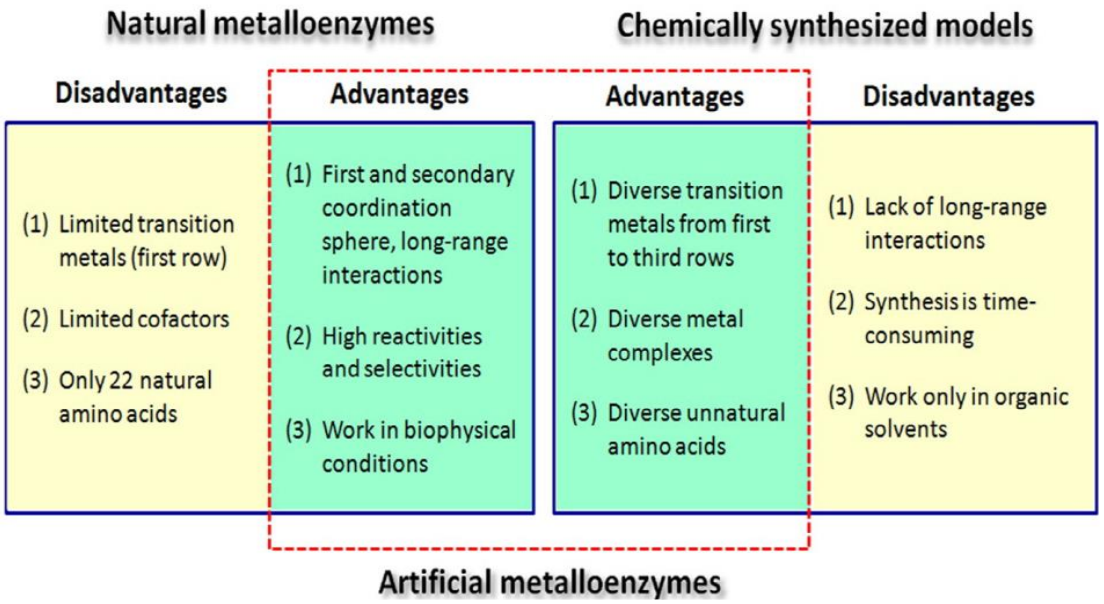
Highly sophisticated catalysts can be chemically synthesized and can act like enzyme models that can catalyze desired reactions. Such catalysts -namely biomimetic catalysts- can be designed to mimic the active site of natural metalloenzymes. Their preparation however requires environmentally unfriendly protocols. Since chemically synthesized catalysts usually work in organic solvents, artificial metalloenzymes were

designed to overcome the limitations of both natural enzymes and chemically synthesized models (see Table 1).[4,17]

A chiral environment can be easily provided by biomolecule hosts under mild conditions in aqueous solution and which can host simple metal complexes to catalyze desired reaction under enzyme-like conditions.[4]

Artificial metalloenzymes have thus several advantages such as: they can perform asymmetric catalysis to obtain pure enantiomers, they can catalyze various reactions, using a broad range of substrates, and can allow to obtain large scale products for a practical use while working under eco-compatible conditions.[4]

Table 1: Features and limitations of natural metalloenzymes and chemically synthesized models, in comparison to those of artificial metalloenzymes. [4]



1.1.1. Different strategies to construct artificial metalloenzymes:

Initial challenges to construct artificial metalloenzymes have been explored by Kaiser and co-workers[3] as well as Wilson and Whitesides[18] since the late 1970's. Different strategies have thus been developed[3,19] such as:

- *Covalent attachment*

This strategy requires the presence of a reactive group (such as maleimide) on the metal complex that will interact with a specific amino acid of the protein such as the SH chemical function of cysteine (Figure 4). Kaiser *et al.* were the first to demonstrate that novel catalytic activity can be created by covalent modifications of an amino acid residue with appropriately modified metal complex.[3,4]

- *Supramolecular anchoring*

This strategy relies on the covalent attachment of the metal complex moiety to an inhibitor or substrate analogue that possesses high affinity for the hosting protein, and that is then used as a "Trojan horse" to drive it inside the active site (Figure 4). This method was historically introduced by Whitesides *et al.*[18] and a variety of enantioselective transformations were performed with such artificial enzymes, including sulfoxidation, epoxidation, reduction and Diels–Alder cycloaddition reactions.[4]

- *Metal exchange*

This strategy relies on the intrinsic affinity of a protein for a metal cation (Figure 4). Kaiser *et al.* were the first to modify a protein by this strategy to afford an artificial metalloenzyme with novel catalytic properties by substitution of zinc by copper in carboxypeptidase A (CPA). [4]

A separate bioconjugation step for introducing a synthetic ligand is not necessary. This strategy is nevertheless challenging, because the design must be so "perfect" that the protein binds the respective transition metal solely at the envisioned site. If alternative binding sites in the protein compete for the metal, then an undesired mixture of catalysts will be formed. In the worst case, some of the catalytically active metal salt will not bind at all, leading to catalysis outside the protein active site thus racemic products.[20]

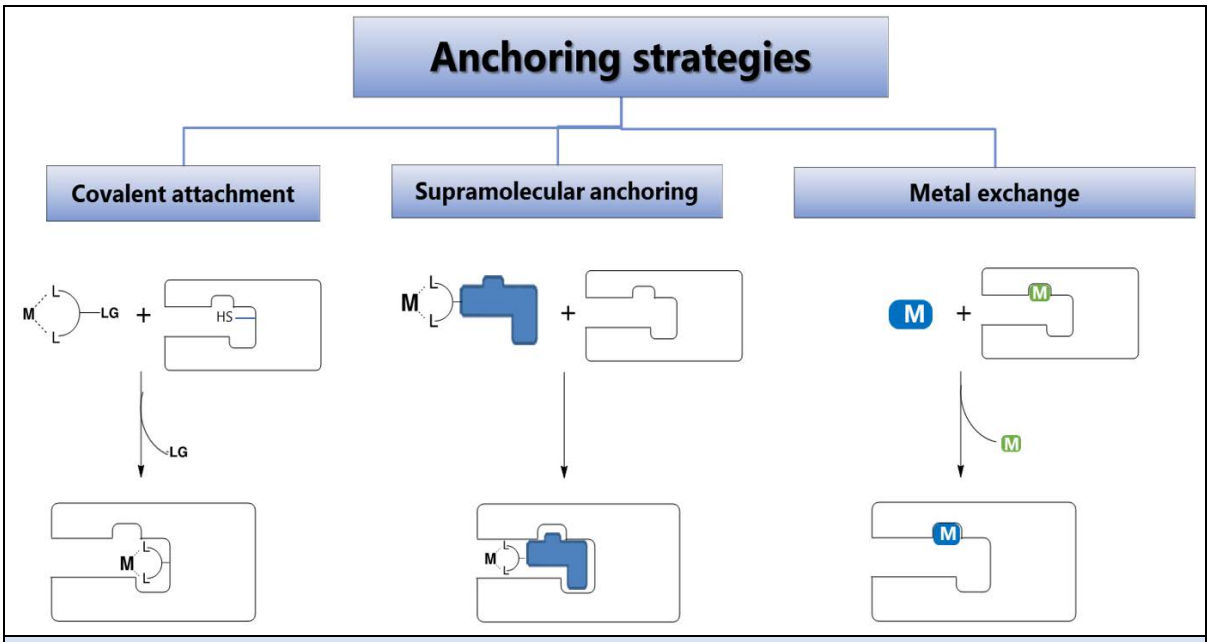


Figure 4: Three strategies for site-specific introduction of transition metals or their complexes into in protein hosts (L = ligand; M = transition metal)

- *Directed evolution*

Directed evolution is a protein engineering technique increasingly used to improve enzymatic activity for different types of substrates and to enhance their enantioselectivity.[21] It also allows for non-biological activity to be conferred on an existing protein or enzyme. Thus, directed evolution induce an artificial activity and can help improve the activity by increasing the conversion and selectivity of the chosen reaction. In a first round of mutation, the gene encoding for the enzyme will be randomly mutated to obtain a library of mutant genes by random PCR mutagenesis. Then, the mutant genes are expressed in microorganisms to produce various mutated enzymes. The catalytic activity of these mutants for the target reaction is then screened and the mutants with activity are selected (Figure 5). A new random mutagenesis step is performed on the mutated genes that provide enzymes the best activity to build a new library.[22]

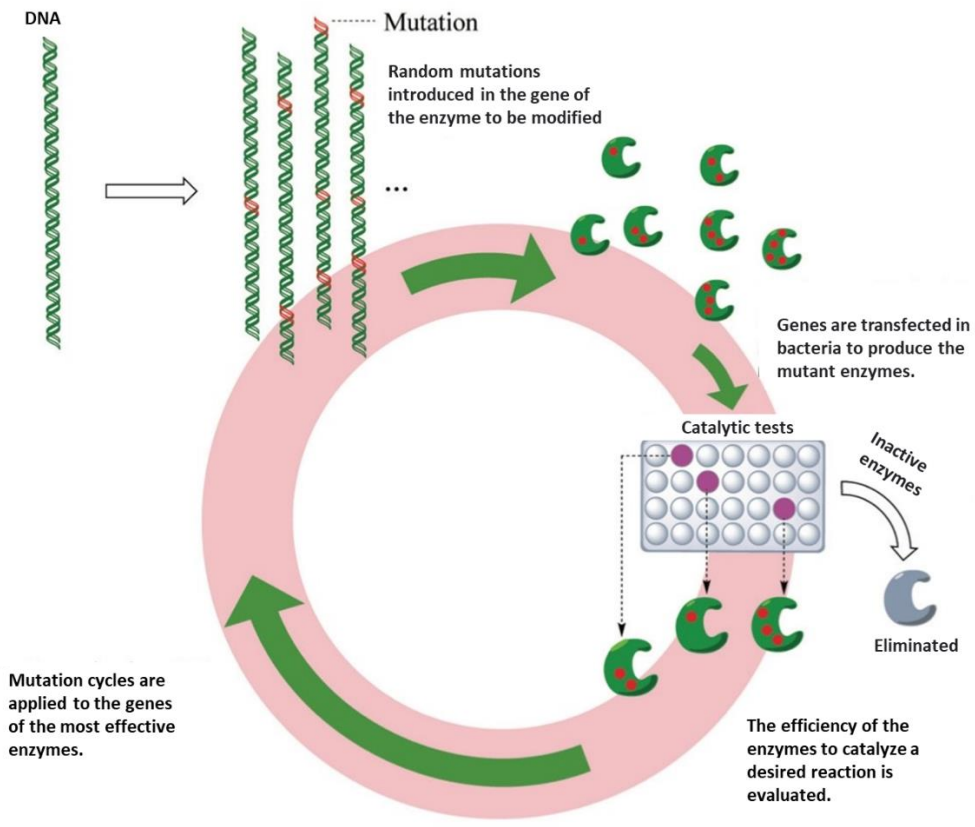


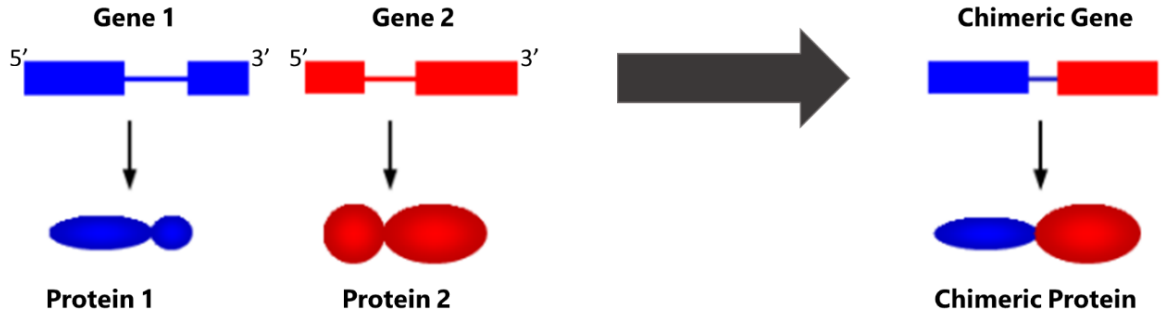
Figure 5: Production of enzymes with an optimized catalytic activity following the directed evolution strategy.[23]

Several research groups, and in particular that of Frances Arnold whose work was rewarded by a Nobel prize in 2018, have developed the directed evolution technique and applied it to hemoproteins in order to make them capable of catalyzing new non-biological activities such as nitrene and carbene transfer reactions. [22]

- *Fusion of multiple proteins*

The formation of chimeric genes can result in proteins with modified functions and sometimes entirely new functions.[8] In some instances, these chimeric gene products are selectively favored and then rapidly spread through populations.[24]

Chimeric proteins can be prepared in vitro by fusing the structural genes of desired proteins into a suitable expression vector. The 3 'translational term of the first gene is deleted, as is the promoter of the 5' term of the second structural gene. The two genes are linked in the frame and expressed in an appropriate host (scheme 5).[24]



Scheme 5: Formation of chimeric gene by deleting 3′ translational term of the gene 1 and the 5′ translational term of the gene 2.

In this project we were interested in this strategy to prepare an artificial chimeric metalloenzyme that possess a double active site by the fusion of the catalytic domains and the deletion of non-catalytic domains of two related enzymes. The development of such multifunctional chimeras is beneficial as not only these contain the same number of atoms as the starting enzymes do, but, in addition they can catalyze two or more reactions.[8]

1.1.2. Some examples of reactions catalyzed by artificial metalloenzymes

Many artificial metalloenzymes prepared by the strategies described above have been reported in literature, in particular by Ward *et al.*[25] They have been reported to catalyze a large variety of reactions such as:[4,19] Redox reactions, Hydrogenation, ketone reduction, alcohol oxidation, epoxidation, dihydroxylation, peroxidation, hydrolysis, transamination, DNA-cleavage, C-C bond forming artificial metalloenzymes, Diels-Alder cycloaddition, hydroformylation, suzuki coupling, etc… Among the reported artificial metalloenzymes, an artificial Diels-Alderase was developed recently in the host laboratory from 1-amino-cyclopropane oxidase (ACCO), which belongs to the cupin family of protein.

ACCO is a non heme iron containing enzyme (Figure 6) that catalyzes the final step of the ethylene biosynthesis in plants.[26] In 2013, Thierry Tron *et al.* characterized for the first time a Cu^{2+} reconstituted ACCO using experimental and theoretical approaches. They have shown that Cu^{2+} coordinated to the protein with a good affinity and it bound two histidines and two water molecules. [26]

ACCO was transformed into a Diels-Alderase by exchanging the iron(II) cation of its active site by a copper(II) cation.[27] The artificial enzyme obtained was more effective and more selective than any other known artificial Diels Alderase.

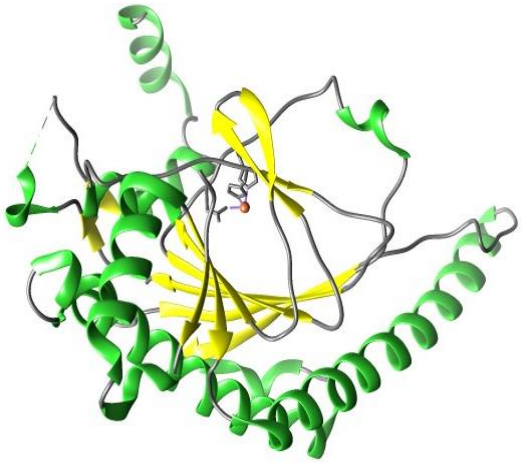
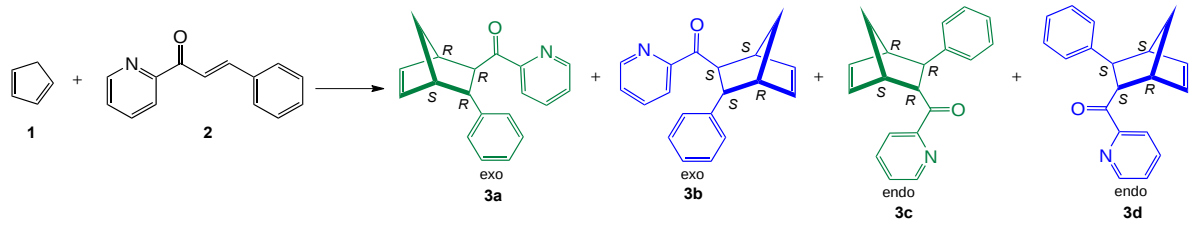


Figure 6: Crystal structure of ACCO proving that it belongs to the cupin superfamily of enzymes and revealing its iron(II) containing active site.

The Cu(II)-ACC artificial enzyme could catalyze the cycloaddition of 40 equivalents of 2-azachalcone with an excess of cyclopentadiene and formed exclusively the (1*R*,2*S*,3*S*,4*S*) *endo* product with 100 % enantiomeric excess (scheme 6).



Scheme 6: Diels-Alder cycloaddition of cyclopentadiene and 2-azachalcone.

1.2. CHIMERIC ENZYMES

According to Greek etymology, the term "chimera" originates from the "Chimaera" a monstrous creature with the head of a lion, body of goat, and tail of serpent.

According to National Cancer Institute (NCI) dictionary of cancer terms, ''Chimeric'' means having parts of different origins. In medicine, it refers to a person, organ, or

tissue that contains cells with different genes than the rest of the person, organ, or tissue. This may happen because of a mutation that occurs during development, or results of a transplant of cells, organs, or tissues from another person or from a different species. In the laboratory, a chimeric protein can be made by combining two different genes. The formation of chimeric genes can result in proteins with modified functions and sometimes entirely new functions. In some cases, chimeric proteins do not fold in the same way as the parental proteins.[8]

As mentioned before, the development of multifunctional chimera is beneficial and has many advantages especially in atom economy. [8,10] However, it is important to engineer a chimeric protein with improved catalytic efficiency to generate bifunctional or multifunctional enzymes and to enhance the consecutive enzyme reaction rate. [8,28]

Fusion proteins can be obtained by different strategies mimicking the nature's strategy such as: tandem fusion, domain insertion and post translational conjugation. [28]

Starting from partner proteins with similar working temperatures and pH profiles appear to be especially important for designing fusion enzymes, as mismatched parameters could affect overall bioactivity. For fusion enzymes, the tandem fusion approach is commonly utilized, and suitable linker peptides should be chosen to avoid severe interference between fusion partners. [28] The orientation in which the fusion partners are positioned has a significant impact on their performance. The transfer of substrates, cofactors, and electrons can be considerably facilitated by a favorable orientation and distance between the active sites. When constructing fusion enzymes, it is always a good idea to take into consideration the known tertiary structures of each protein, the domain order on the polypeptide chain, and the length of linkers if any are used. This should favor the good folding of the design before it is built physically. [28]

Many fusion enzymes were important in the biotechnology and biopharmaceutical fields. For example, pyrroline-5-carboxylate synthase and acetyl-coA carboxylase have been characterized as efficient biocatalysts in metabolic pathways; green fluorescent protein (GFP) is considered as a biosensor for monitoring signaling molecules.[28]

1.2.1. Fundamental types of chimeric proteins: Type 1 & Type 2

Chimeric genes are essential evolutionary forces because they provide a source of new genes, allowing organisms to create novel phenotypes and adapt to their environment. [24] Chimeric genes, as opposed to direct gene duplication, can result in proteins with altered and sometimes altogether novel functions. One of the most well-known chimeric protein genes has been identified in Drosophila and has been named

jingwei.[24,29] Its chimerical gene sequence contained a retro-sequence of alcohol dehydrogenase (ADH) and a non ADH parental gene sequence.[29] Biochemical analysis has revealed that the jingwei-encoded protein evolved a new functional role in the metabolism of the hormone.[24]

Chimeric proteins can be also prepared in vitro by fusing the structural genes of the proteins in question into a suitable expression vector. Fusion proteins are divided into two categories: Type 1 and type 2 proteins (Figure 7).

In type 1, the 3 'translational term of the first gene is deleted, as is the promoter of the 5' term of the second structural gene. The two genes are linked in the frame and expressed in an appropriate host. Plant, mammal, bacteria and insect cells have been used as hosts but the most commonly used hosts are *Escherichia coli* strains.[8]

After transcription and translation, the cell will produce a single chain of polypeptide with the properties from both original genetic products. Fusion can be performed at one or both ends of a protein. Thus, it should be assayed whether one end or the other is more favorable for the sought-after biological activity. [24]

In type 2, the sequences of two related proteins are interlaced to create a distinct protein, generally with a novel activity. Unlike the first type of fusion protein, which has two types of activity that are not novel, this type of fusion protein usually only has one type of activity that is however novel. [8]

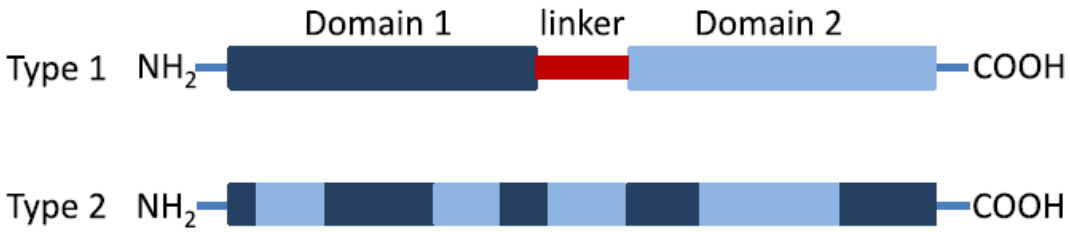


Figure 7: Two fundamental types of chimeric proteins: Type 1: The two proteins or protein subunits are united end-to-end and usually linked by a linker; type 2: the amino acids from both donors are interspersed in the fusion protein product.[8]

The growing number of sequenced genes and genomes, together with the use of polymerase chain reaction (PCR) to recover intact genes, open up a plethora of possibilities for structural gene fusion from many sources. The nearly infinite number of conceivable fusion partner combinations has made this technology extremely adaptable in many fields of biochemistry and biotechnology. [8]

Genetic engineering has made it possible to create chimeric proteins via the fusion of genes encoding separate enzymes into a chimeric gene coding for a single multifunctional polypeptide.[24] For example, fatty acid synthases and the enzymes involved in aromatic amino acid biosynthesis exist as discrete enzymes in many plants and in most prokaryotic organisms. In other species, various combinations of multifunctional enzymes are present. Thus, they are considered to be examples of enzymes produced through an evolution process involving gene fusion. [24]

1.2.2. Choice of the linker

The main problem in the production of chimeric proteins is the retention of activity following fusion. Hence, it's crucial to find the suitable fusion point to retain the original folding and thereafter the original activity. [24]

The capacity of individual units to fold independently of the rest of the polypeptide chain after fusion is critical to the design of a chimeric protein. Misfolding otherwise often results in inactive structures. When developing and examining a linker region, there are numerous factors to consider. The first criterion is to avoid proteolytic cleavage within the linker, which restricts the linker's amino acid composition. [24]

Linkers are found in all naturally occurring multidomain proteins, and their role is to maintain the required distance between two protein domains to avoid steric hindrance and/or to allow for favorable domain–domain interactions. As a result of this observation, researchers have used numerous sorts of naturally occurring linkers in their synthetic fusion structures.[28] For example, in many recombinant therapeutic proteins, such as tailored antibody constructions, the immunoglobulin hinge region serves as a linker. [28,30] Natural enzyme linker sequences are also attractive possibilities for joining fusion partners: O-glycosylation, proline- and hydroxylamine-rich cellulase and xylanase linkers maintain an extended shape and are protected from proteolysis.[28] Fusion protein linkers have lately been studied and explored in terms of their features, functions, and instances. In general, various linker qualities should be carefully considered: length, amino acid composition, hydrophobicity, protease sensitivity, secondary structure, and probable fusion partner interaction. The linkers appear to be a critical component of a good fusion structure. Furthermore, because each sort of linker has its own unique properties, it's difficult to assume that a linker that works in one circumstance would work in another. [28]

Researchers have developed a variety of artificial linkers in addition to natural linkers, which can be categorized into flexible, rigid, and in vivo cleavable linkers. [28]

(Gly)n and (Gly-Gly-Gly-Gly-Gly-Ser)n are the most commonly used flexible linker sequences in which the linker length is controlled by the copy number "n." [28] The high homologous repeats in this family of linkers' DNA encoding sequences, may limit protein expression levels and necessitate suitable codon pair selection. Rigid linkers may be desirable when sufficient separation of protein domains is desirable. A helix-forming linker with the (EAAAK)n motif, for example, was created for this purpose. The intrinsic Glu-Lys+ salt bridges in this linker stabilize it, resulting in a rigid and prolonged shape. [28]

When the recombinant fusion protein is subjected to reducing agents or proteases, cleavable linkers are used to liberate free functional domains in vivo. [28]
In natural chimeric proteins, bulky amino acids as well as hydrophobic residues are avoided in linkers, excepted the smaller hydrophobic residues, alanine and proline. On the contrary, Glycine, serine and threonine are the preferred amino acids in natural linkers whereas other common residues such as asparagine, glutamine and lysine can also be found.[24]

It can be useful to arrange the catalytic centers of various enzymes in a close vicinity, to increase the overall kinetics and yield of an enzymatic process, especially if it is based on successive stages. In this way, the intermediate product can be efficiently channeled to the second active site of the fusion protein. In nature, this has been achieved through the evolution of multifunctional proteins and multienzyme complexes with small linkers.[24]

Interesting bioinformatics tools have been created to facilitate the logical design and selection of linkers. LINKER, an online application for automatically creating linker sequences for fusion proteins, was created. It should be noted that the server page is no longer accessible, most likely owing to a lack of updates. [28]
Similarly, in 2002, R. A. George and J. Heringa created a web-based linker database (http://www.ibi.vu.nl/programs/linkerdbwww/) that returns a list of linker possibilities that meet user-specified criteria including linker length, sequence, and secondary structure preferences.[31]

In 2012, Kai Yu et al. developed ‘‘LinkerDB’’, (http://bioinfo.bti.a-star.edu.sg/linkerdb) a novel tool to help with linker selection and fusion protein modeling. It included about 2200 linkers taken from the most recent multi-domain structures in the Protein Data Bank (PDB), as well as empirical recombinant fusion proteins reported in the literature and patents.[28]

1.2.3. Design of chimeric enzymes

The structure's prediction is an important factor to be taken into consideration before proceeding to the construction and the expression of a chimeric enzyme.

1.2.3.1. Structure's prediction:

Although significant progress has been made in the development of synthetic fusion proteins, the structures of these proteins are difficult to predict. This is mostly due to their vast size and inherent complexity and flexibility. [28] As a result, only a few structural features of synthetic fusion proteins, such as inter-domain interactions, have been well elucidated thus far. [28] The interactions between domain pairs, can be crucial in the design of fusion proteins that need proximal domain placement for quicker electron transfer, substrate channeling, or cofactor regeneration. As a result, a greater understanding of the structure of the fusion protein could potentially speed up its rational design and increase its success rate. In comparison to experimental determination, computational protein structure prediction can be regarded as a cost-effective option. Significant progress has been achieved in this field to better understand the structure of fusion proteins by utilizing individual domain structural information either from known structural data or from homology modeling.[28]

- Fusion Protein MODeller (FPMOD) is a modeling tool developed by Chiang *et al.* [32] in 2006 that can construct fusion protein models using the structures of the fusion partners. The realistic models can be constructed by designating regions of flexible linkers and then rotating rigid-body domains around them to build random structures.[28]

- Rosetta software is used in computational modeling to analyze protein structures, to anticipate the structure of designed domain insertion proteins from scratch, and to create such fusion structures.[28]

- MODELLER is a program that is used to model protein three-dimensional structures in terms of homology or comparison.[33,34] It automatically calculates a model comprising all non-hydrogen atoms when the user aligns a sequence to be modeled with known related structures. MODELLER performs de novo modeling of loops in protein structures, optimization of various protein structure models with respect to a flexibly defined objective function, multiple alignment of protein

sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, and many other tasks.[35]

- I-TASSER is a hierarchical approach to automated protein structure prediction and structure-based function annotation. Multiple threading alignment techniques are used in I-TASSER to recognize structural templates from the PDB. Iterative fragment assembly simulations are then used to build full-length structural models.[36]

- AlphaFold network predicts the 3D coordinates of all atoms for a given protein. It relies on artificial intelligent by learning from multiple core amino acid sequences/structures and aligns sequences of homologues as inputs.[37]

The choice of component proteins to be fused is generally simple when the specific application is in mind. Fusion partners, in theory, should have been thoroughly researched and reported to be physico-chemically compatible (e.g., optimum pH, temperature, ionic strength, and inhibitor effects).[28]

The order in which fusion partners appear in the polypeptide chain is important, as the location of one domain can have an impact on the localization and functionality of another.[28] As a result, unless the possible effect of the order is known, two sequence choices should be tested in the event of a two-component fusion. An end-to-end genetic fusion, in which their coding genes are fused together and produced as a single peptide chain in a suitable host organism, is the easiest method of combining them after they've been chosen.[28]

1.2.3.2. *Construction of fusion proteins*

Fusion proteins can be obtained by different strategies mimicking nature's strategy such as tandem fusion, domain insertion and post translational conjugation (Figure 8).[28]

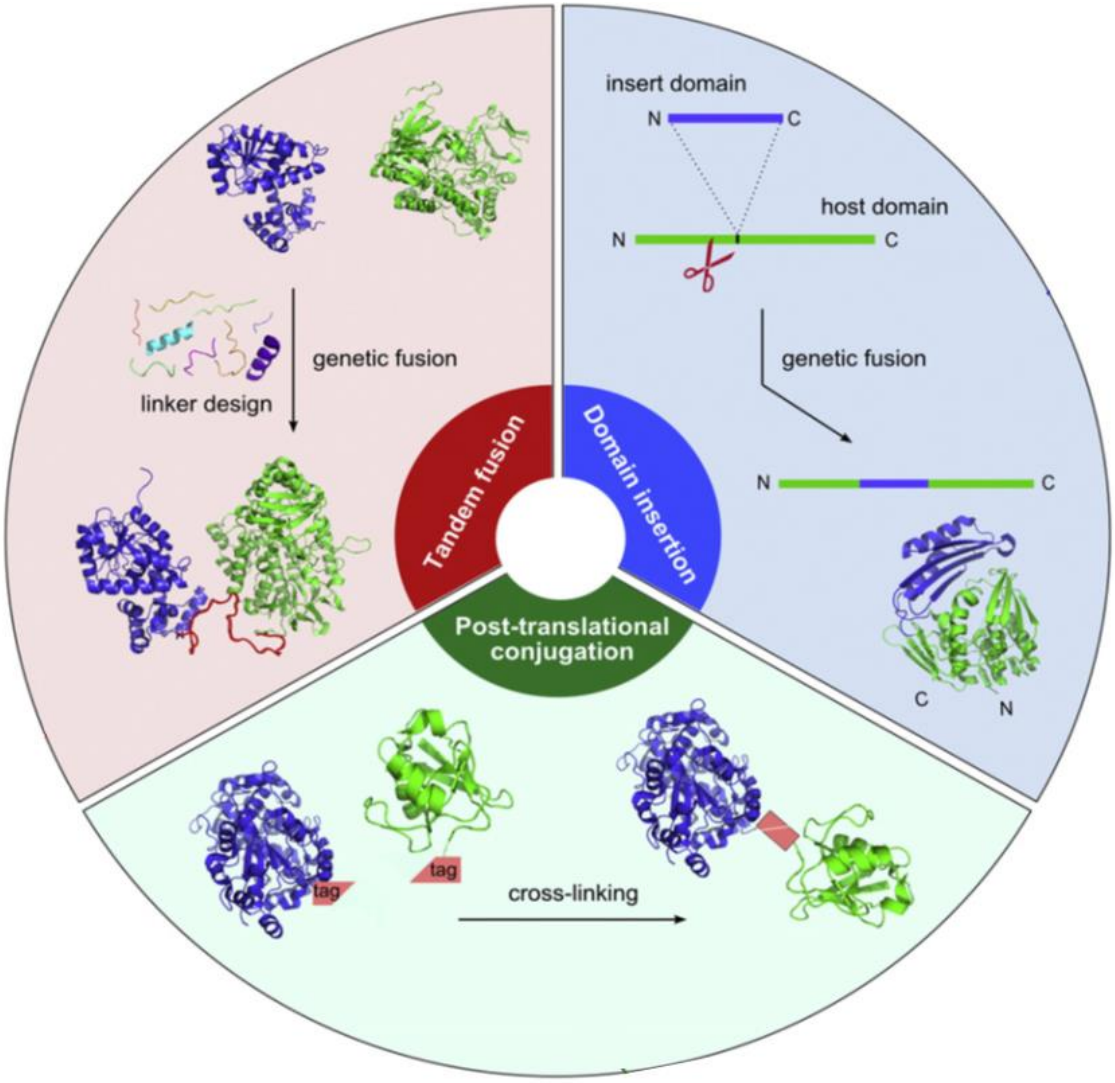


Figure 8: Common strategies for the construction of fusion proteins.[28]

- Tandem fusion

Direct tandem fusion is easy and effective in some circumstances where the N- or C-terminal regions are unstructured and flexible. These regions can operate as a "bridge" to give enough space between protein domains for proper folding.[28] This strategy fails, however, when either the free N- or C- terminus to be tethered is required for the function of the parent protein and/or is not flexible or long enough to avoid steric hindrance, which reduces the degrees of freedom in protein dynamics and can result in undesirable consequences such as protein misfolding, aggregation leading to inclusion body formation, low protein yield, and impaired bioactivity.[28] As a result, linker peptides are commonly used to connect fusion partners to preserve their distinct structures and functions (Figure 8).[28]

- Domain insertion

The majority of multidomain proteins are made up of domains encoded in sequential sequences on a single polypeptide chain; however, there are exceptions, such as when one domain is enclosed within the other.[28]

In 2004, Aroul-Selvam *et al.*[38] reported that insertions are found in 9% of multidomain proteins with known structures, particularly in the α/β classes of proteins. This shows that the polypeptide sequence's continuity isn't required for proper domain folding. As a result, a functional fusion protein can be achieved by introducing one domain into a host protein (Figure 8).

It's worth mentioning that domain insertion is more difficult to design than tandem fusion because it requires finding a "cut" site on the host protein to accept the insert, which, if chosen incorrectly, could affect the folding of both domains.[28] As a rule of thumb, insertion spots should be surface loops or turns that are not directly involved in the catalytic sites, to avoid clashes between the host and the inserted domains.[28] The insert domain is usually shorter than the host, with its exposed N- and C-termini ligated to the host protein's internal regions through two ligation sites. In most cases, a successful fusion design using this method involves two conditions:[28] 1) the insert domain's N- and C-termini must be close together, and 2) the parent domain must be able to withstand the discontinuity induced by domain insertion. To preserve their individual functions, the structure of both domains should be intact and mutually permitted after the insertion without significant disruption.[28] Another important step in this method is to choose suitable ligation sites. In this regard, the program SCHEMA can assist in the selection of proper boundaries for chimera building without affecting the three-dimensional structure's integrity.[39] SCHEMA scans the protein structure and finds the interactions between residues that may be broken in the resulting chimeric protein, allowing for the prediction of possible chimera misfolding.[28]

Figure 9 shows an example of engineering a chimeric enzyme using SCHEMA tool to find the insertion sites.

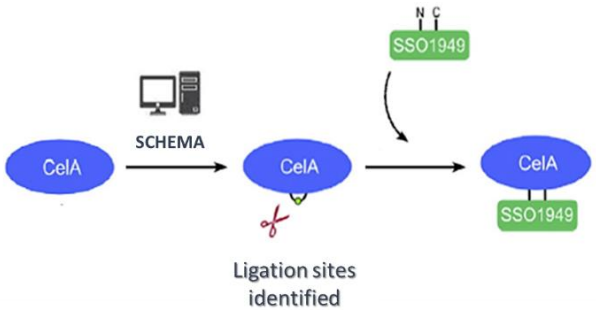


Figure 9: The engineering of a chimeric enzyme comprising two cellulases, CelA and SSO1949 using the 'domain insertion' approach. The SCHEMA tool was used to find the insertion sites on CelA.[28]

- Post translational protein conjugation:

Recombinant proteins can be produced via an alternate approach (Figure 8): individual proteins can be biochemically joined after they have been individually expressed. This approach is a protein-level ligation in which protein expression and fusion are separated into two phases. Although the two genetic fusion mechanisms outlined above are the most common approaches for chimera construction, a given pair of protein domains may fail to express in a tolerable yield or lack correct folding and function, limiting the range of domains that can be joined.[28]

When working with recombinant proteins derived from various hosts, this problem becomes even more serious. Post-translational protein fusion, on the other hand, allows for modular pairing and assembly of component domains after their expression in optimum hosts, effectively bypassing the genetic fusion's diversity-limiting cloning processes.[28]

Bifunctional reagents capable of covalently binding to reactive functional groups on proteins (e.g., –SH, $–NH_2$, and –COOH), which are routinely employed for protein immobilization and stabilization, are typically used for intermolecular protein conjugation. Although direct chemical cross-linking can result in functional fusion proteins, it is not always site-specific and can lead to protein aggregation.[28]

Because the genetic fusion strategy may result in undesirable domain interactions or low protein production, post-translational fusion can be used in multiple-component systems.[28] Many examples have been reported such as the fusion of bovine serum

albumin with a lipase, the fusion of two immunoglobulin-binding protein G domains and the conjugation of two Green Fluorescent Proteins (GFP).[28]
Aside from utilizing the reactive side chains of natural amino acids, the introduction of non-natural amino acids with unusual functional groups, such as azide, ketone, and phosphate, into target proteins via stop codon suppressor tRNA and in vitro translation technology would provide a new level of flexibility in the creation of post-translational protein fusions.[40]

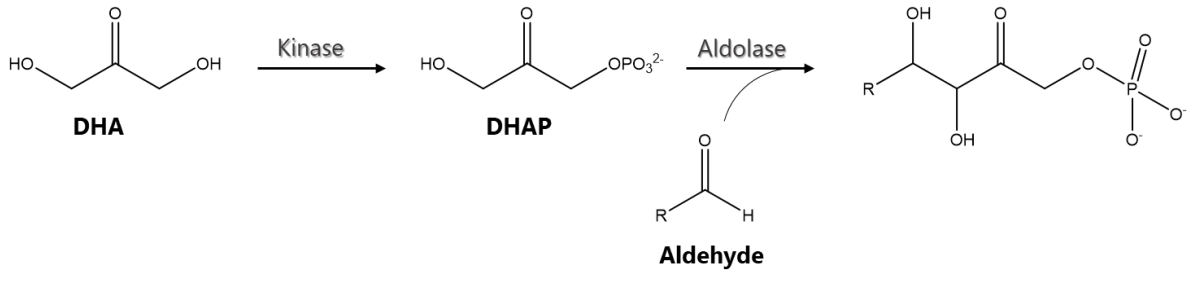
1.2.4. Applications of the fusion proteins:

Artificial fusion proteins have been used in different technologies, so they have a broad range of applications.[28] Some applications are described below:

1.2.4.1. Multifunctional biocatalytic activities:

Engineered fusion enzyme may lead to improved capabilities vs the parent proteins as merging functional properties from different enzyme moieties in a single one, facilitates the channeling of the substrates in a sequential process, reduces the effective reaction volume, and improves electron transfer.[28]

- In 2010, Iturrate *et al.* engineered a bifunctional enzyme composed of dihydroxyacetone (DHA) kinase and fructose-1,6-bisphosphate (FBP) aldolase. DHA was transformed to the unstable dihydroxyacetone phosphate (DHAP) by the DHA kinase part of the chimera. DHAP was then used as a substrate in the aldol condensation reaction catalyzed by (FBP) aldolase part of the chimera to form phosphorylated aldol adduct, resulting in a one-pot carbon–carbon bond synthesis technique (scheme 7) with 20-fold increase in the aldol reaction rate.[28,41]

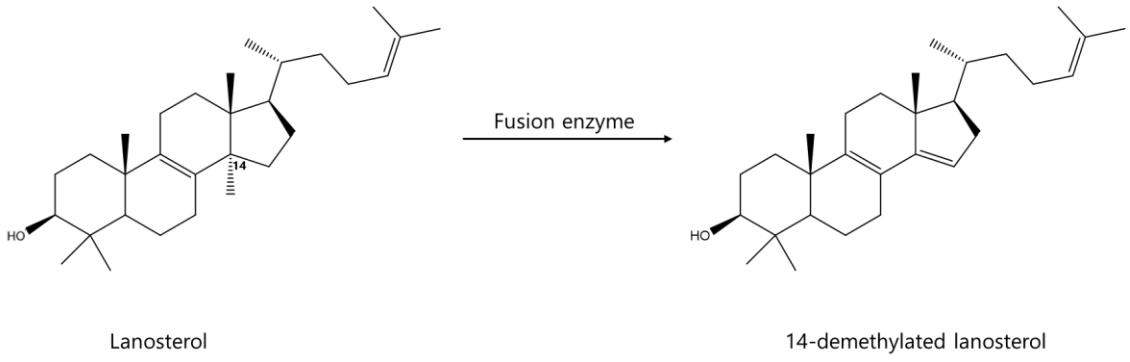


Scheme 7: Reaction catalyzed by the bifunctional enzyme composed of DHA kinase and FBP aldolase.[41] Benzyloxyacetaldehyde, acetaldehyde and 3-(methylthio)propionaldehyde were the assayed aldehydes.

- In 2010, a chimeric protein was constructed by Wang *et al.* combining arabinosidase, xylanase and xylosidase. This trifunctional protein promoted the degradation of complex agricultural byproducts with an obvious increase in activity.[28,42]

- In 2012, Choudhoury *et al.* developed an artificial chimeric enzyme via the fusion of an RNA cleavage domain and Pumilio/fem-3 mRNA-binding factors. The two domains were linked via the tandem strategy, and the RNA endonucleases that resulted were able to detect RNA substrates and cleave them efficiently near their binding sites.[28,43]

- In 2012, Choi *et al.* engineered a chimeric enzyme composed of Cytochrome P450 (CYP51) and Fe-S containing NADPH reductase. This chimeric enzyme was able to catalyze oxidative demethylation of lanosterol more efficiently than CYP51 and the reductase separately, culminating in the development of a self-contained P450 biocatalysis system (scheme 8).[28,44]



Scheme 8: Reaction catalyzed by the fusion enzyme composed of CYP51 and Fe-S containing NADPH reductase.[44]

1.2.4.2. Biosensing and bioactivity regulation:

The binding of specific molecules by some proteins induces conformational changes. Researchers use such proteins as switches that can operate as biosensors or regulators upon conformational change. The phenomenon of allostery, in which conformational changes at a remote regulatory site can be carried through to the active site, influencing bioactivity in either a negative or a positive manner, is one technique for developing such switches.[28]

In 2013, Meister and Joshi[45] inserted the calmodulin domain (CaM) into Beta lactamase (BLA). They took advantage of the CaM's capacity to change conformation (Figure 10).

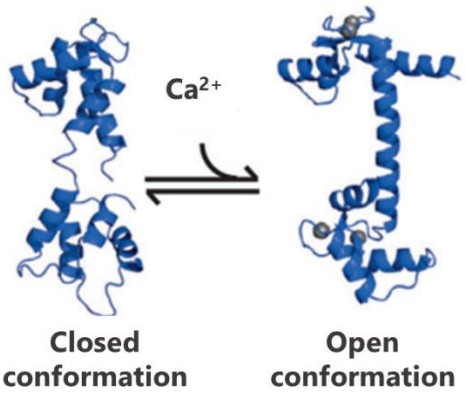


Figure 10: Structures showing the conformational changes of CaM in the closed state (pdb: 1CFD) and bound to Ca^{2+} (pdb: 1CLL).[45]

When compared to the inactive state, the activated state of the allosteric enzyme demonstrated a 120-fold increase in BLA catalytic activity.[28] This engineered calmodulin-based allosteric switch fusion protein can be used as biosensor to detect toxins and other biomarkers.[45]

1.2.4.3. Creating peptide libraries

Creating protein libraries on the surface of cells and bacteriophages is considered an important application of chimeric proteins.[24]

Recently, R.M. Lu *et al* [46] worked on developing several chimeras of nucleases. They have designed a Cas12a-type nuclease library by fusing different Cas12a-like nucleases and successfully screened functional chimeras and identified several dozen that showed basal editing.[46] Then, they demonstrated the use of this strategy as a platform for rapid generation of nucleases. Some chimeras showed better specificity than all other parent-wild type nucleases. This demonstration opens-up the possibility of generating new nuclease sequences with implications across biotechnology.[46]

1.2.4.4. Bio-therapeutics applications

Therapeutic fusion proteins are categorized based on their incorporated protein domains such as: Fc (fragment crystallizable region) fusions, antibody-enzyme fusions, immunotoxins-cytokine fusions, and many other categories.[24]

- Immunotechnology

Recently, chimeric proteins were shown to be promising tools for developing new antibodies that are used in a variety of analytical, therapeutics and clinical applications.[24,47] For example, fusions involving antibodies as one fusion partner, have made affinity-based analytical procedures such as enzyme-linked immunosorbent assay (ELISA) and Western blotting easier to design.[24]

- Creating new drugs with targeted delivery to tumor cells

Recombinant immunotoxins are developed via the fusion of a toxin with an antibody or its fragment that has an affinity for the target tumor cells.[24]

i. In 2003, Halin *et al.*[48] fused two cytokines: Tumor necrosis factor (TNF) and interleukin-12 (IL-12), with an scFv human antibody (Figure 11). This recombinant protein showed strong tumor targeting properties, which were attributed to the cytokine activity and antibody binding capacity.[28]

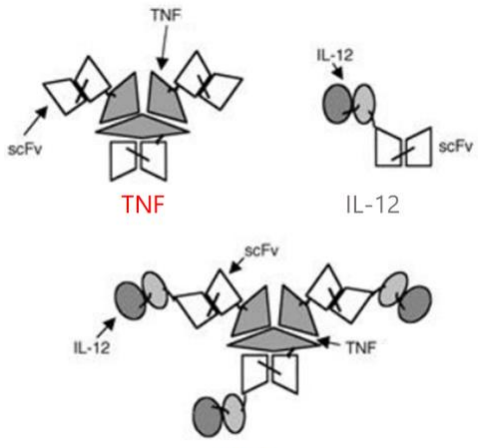


Figure 11: Schematic representation of the recombinant TNF – IL-12 protein comprising the IL-12, TNF and scFv human antibody.[48]

ii. In 2004, a highly stable conjugate with potent tumor activity was created via the fusion of BLA enzyme with the heavy chain of a single-domain antibody. This recombinant enzyme has shown a good biodistribution profile in vivo and a cancer therapy. It can be used to cure tumor xenografts.[28,49]

- Humanized antibodies

Many antibodies were and still are developed based in mice. These antibodies evoke an immune response when administered to humans. Chimeras wherein human antibody constant domains replaced the corresponding mouse moieties eliminated most of the immunogenic part of the antibodies without modifying specificity.[24]

- Some examples of fusion therapeutic's proteins:

There are currently a number of fusion protein-based therapeutics on the market.[28] Some examples are cited below:

i. Etanercept: a drug formed by a tumor necrosis factor receptor (TNFR) linked with an immunoglobulin G1 (IgG1) Fc segment.[24,28] It is indicated to treat autoimmune diseases such as psoriasis, rheumatoid, polyarthritis…
ii. Ontak: developed via the fusion of the human IL-2 and the catalytic and transmembrane domains of the toxin of diphteria.[28] It is indicated to treat recurrent cutaneous T-cell lymphoma.
iii. Amevive: results from the fusion of the Fc domain of IgG1 and the human LFA-3. It is used to treat symptoms of psoriasis.[28]

1.3. CUPIN PROTEINS

As we mentioned before, to prepare the chimeric enzyme, we chose to rely on the cupin superfamily of proteins. We selected two related proteins from the bicupin family: PMI and PLP. For this, we will describe below the cupin superfamily of proteins.

The cupin superfamily is a diverse protein superfamily whose name is derived from that of its conserved β-barrel domain, as cupa is the Latin term for a small barrel. This protein superfamily is amongst the most versatile of all protein folds in activity, structure and distribution. It comprises a wide variety of enzymes besides non-enzymatic proteins.[11]

The first cupin to be studied was discovered to be a germination marker in higher plants' seeds by Thompson and Lane[50] in 1980 and was thus given the name "Germin". Initial studies showed that it had a homopentameric structure[51,52] and that it possessed an oxalate oxidase (ox-ox) activity.
Another protein from the cupin family i.e., spherulin, involved in the response of the blob *Physarum polycephalum* to stress was discovered by Bernier *et al.* In 1987. Spherulins have about 44% sequence similarity with germin[53] but show no ox-ox

activity. Part of this sequence similarity comes from a common decapeptide sequence PH(I/T)HPRAIEI named the Germin Box. In 1998, Dunwell and Gane proved that this box was part of a much larger conserved region in numerous proteins, that they named the cupin domain.[54]

Finally, Woo and coworkers[55] published in 2000 a crystal structure at a resolution of 1.6 Å that proved that germin was a homohexameric cupin and shed light on the catalytic mechanism of its ox-ox activity.

The characteristic cupin domain comprises two conserved motifs 1 and 2, each corresponding to two β-strands, separated by a less conserved inter-motif region composed of another two β-strands with an intervening variable loop (Figure 12).[11]

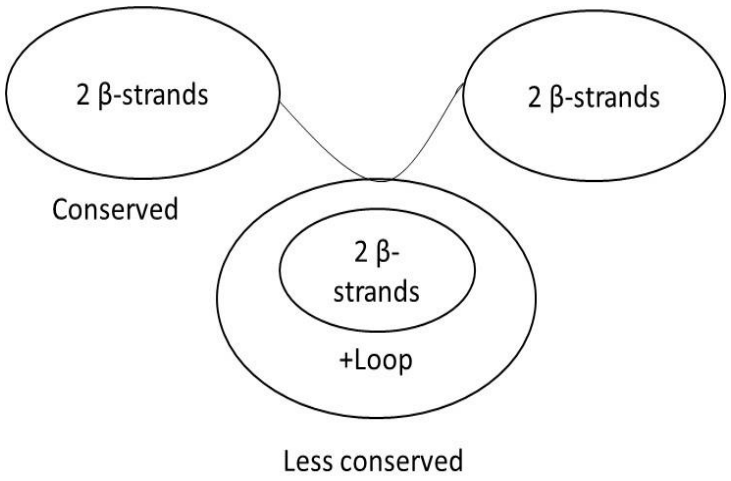


Figure 12: Characteristic structure of cupin proteins showing two conserved 2-β-strands domains separated by a less conserved inter-motif region containing another 2-β-strands and a loop.

The total size of the inter-motif region varies from a minimum of 11 amino acids in some microbial enzymes, to 50 amino acids in the non-enzymatic seed storage proteins, and up to 100 and more amino acids in certain eukaryotic transcription factors and dioxygenases.

Initially, the conserved characteristic sequence of Motif 1 was identified as **G(X)5HXH(X)3,4E(X)6G** and that for motif 2 as **G(X)5PXG(X)2H(X)3N**. [55] However, in 2004, another study [11] showed that the primary sequence of the two motifs was much less conserved. For example, one of two His residues in Motif 1 that are supposedly conserved can be substituted by Gln or Asp/Glu in some isomerases and dioxygenases. The characteristic features of these proteins with this fold include resistance to

proteases as well as thermal stability. These characteristics are attributed to their high degree of subunit contacts, hydrophobic interactions, and short loops.[56]

According to the Protein Data Bank (PDB) (https://www.rcsb.org), 183 structures represented unique members of the Cupin superfamily. Most of them contained iron as an active site metal (49 PDB structures) but others could contain multiple metals such as copper, nickel, zinc, manganese, cadmium, cobalt, mercury as cofactors allowing various kinds of chemical reactions to occur within the active site of the conserved tertiary structure. For example, each monomer of Germin binds one Mn(II) which confers the oxalate oxidase (ox-ox) activity. Many proteins have been identified as cupins through sequence and structure alignments, sometimes without biochemical studies, and thus many are of unknown function.[11]

In Protein database (https://www.ncbi.nlm.nih.gov/protein/), around 1,830,200 entries are identified as cupins. Cupins can be classified according to species, source databases, genetic compartments, enzyme types, sequence length, molecular weight... They can be found in bacteria, animals, fungi, plants, archaea, protists, viruses and homo sapiens. Although, there are cupins whose function remains unknown. According to the Protein database, cupins can be classified according to 6 major enzyme types: Hydrolases, isomerases, ligases, lyases, oxidoreductases and transferases (Table 2).

Surprisingly, among the about 2 million cupins identified on the databases, only a small proportion are classified as known enzyme function.

Table 2: Classification of Cupins according to the Protein database.

Total number found: 1,830,199 cupins					
Isomerases (18,209)	Lyases (2,363)	Oxidoreductases (10,996)	Transferases (10,986)	Hydrolases (1,764)	Ligases (46)

Analysis of the structure-activity relationships among the cupin family had allowed to identify that they were involved in some 18 different functional classes that range from single domain bacterial enzymes (isomerases and epimerases) through two-

domain bicupins (desiccation-tolerant seed storage globulins) to multidomain cupins (transcription factors). [11]

In 2004, proteins from the cupin family have been classified according to the number of cupin domains they contained, so 3 subgroups of cupins: Monocupins, bicupins and multicupins could be distinguished. [11]

a) Monocupins:

These are proteins that comprise a single cupin domain, either as the center of a simple protein or as one of several domains in a complex protein. Monocupins include some microbial transcription factors (for example AraC) and enzymes such as Dioxygenases 2-oxoglutarate iron dependant dioxygenases such as ACCO (Figure 13), Acireductone dioxygenases, Auxin binding protein…

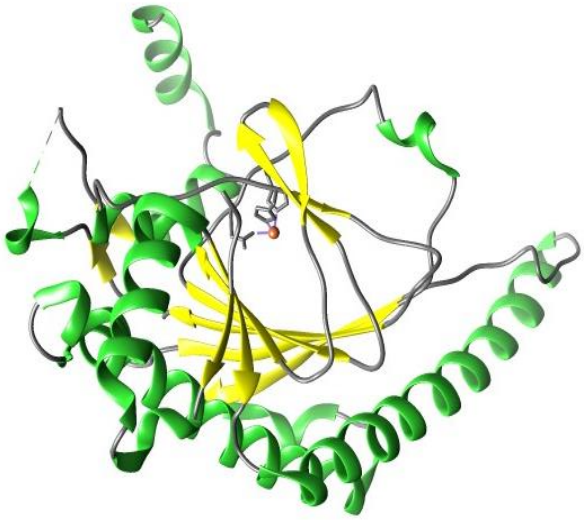


Figure 13: Crystal structure of ACCO showing its unique cupin domain and revealing its Fe^{2+} containing active site.

b) Bicupins:

These proteins are most likely to have evolved from the fusion of two identical single domains issued from the duplication of a single domain monocupin ancestor, though the fusion of two different monocupin precursors may also be possible. Some examples of bicupins are dioxygenases such as Quercetin dioxygenase, Human Pirin protein (Figure 14), Gentirate 1,2-dioxygenase and Oxalate decarboxylase, as well as seed storage globulins and storage protein / sucrose binding protein.

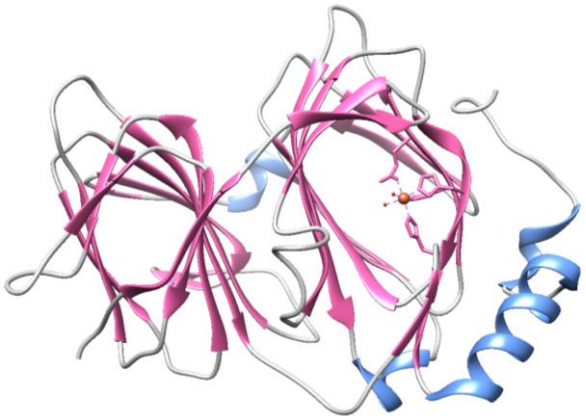


Figure 14: Crystal structure of Human Pirin (PDB: 1J1L) revealing its Fe^{2+} containing site.

c) Multicupins:

These are proteins that contain multiple (>2) cupin domains. This group includes HomoGentisate diOxygenase (HGO, Figure 15). Other potential examples include the Arabidopsis protein that comprises four 2-oxoglutarate-Fe(II) domains.

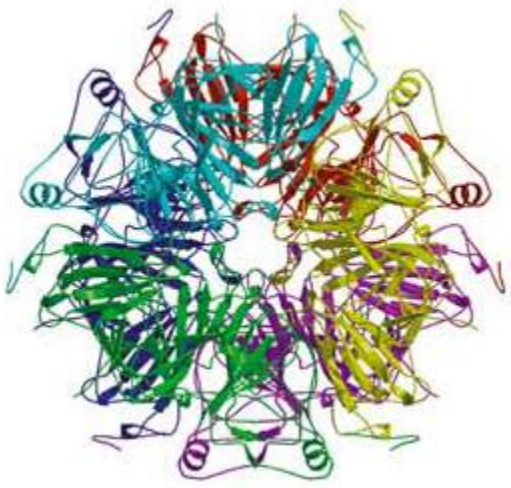


Figure 15: An example of multicupin: Human homogentisate dioxygenase (HGO).

Although the majority of enzymatic cupins contain iron as a metal cation in their active site, other members contain either cations of copper, zinc, cobalt, nickel or manganese as cofactors allowing the catalysis of various kinds of chemical reactions to occur within the conserved tertiary structure.[11]

1.4. CONCLUSION

Biocatalysis using natural enzymes is mostly limited to the natural substrate of the employed enzyme and to biotic reactions. Starting from novel substrates, artificial enzymes catalyze chemist-invented reactions leading to industrially relevant products under ecofriendly conditions.[5]
Metal complexes are the most efficient man-made catalysts that are routinely used to form the catalytic sites of artificial metalloenzymes. The latter combine both the advantages of natural metalloenzymes such as stereoselective catalysis in aqueous solution, under mild conditions and the advantages of chemically synthesized metal complexes catalyst such as chemist-invented catalysis and compatibility with artificial substrates.[16]

During this thesis, several artificial metalloproteins have been constructed by different strategies. This chapter presented an overview of artificial enzymes by discussing their advantages and their construction strategies. In particular, Chimeric enzymes are introduced as they are further advantageous in atom economy.

In this project, we followed a new strategy to develop artificial metalloenzymes: We aim to prepare an artificial chimeric enzyme that possesses multiple active sites by the fusion of catalytic domains of 2 related enzymes. This chimeric enzyme would be used to catalyze reactions in cascade under ecofriendly conditions.

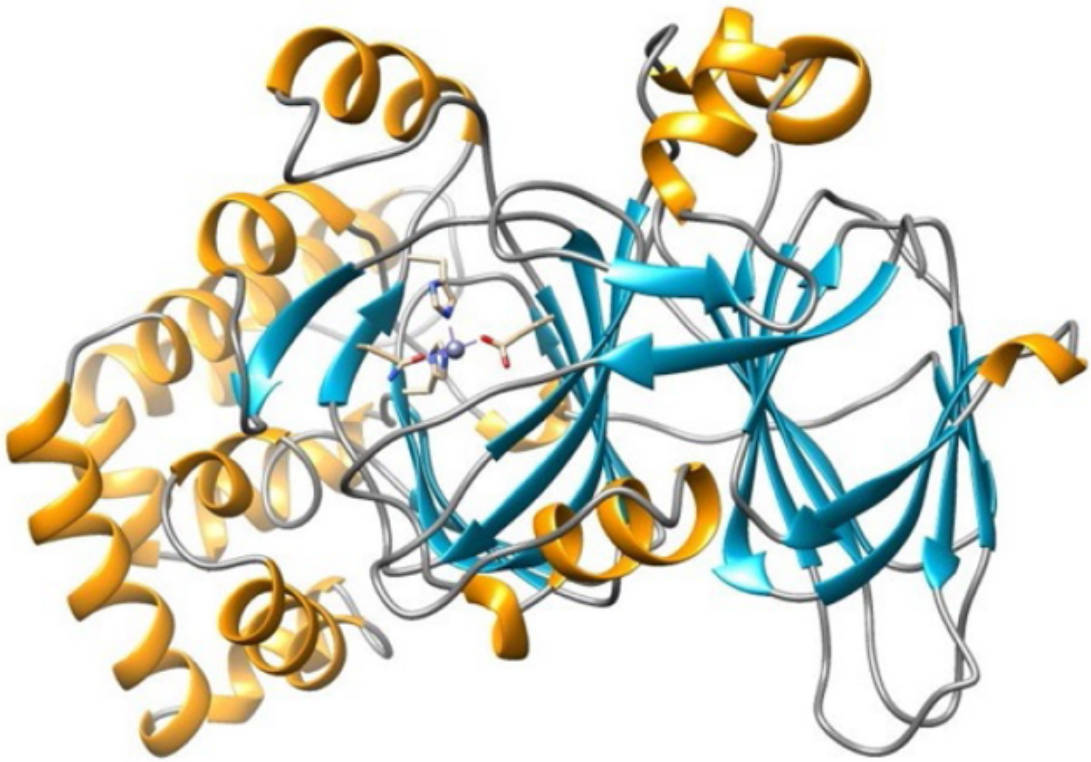
To prepare the chimeric enzyme, we chose to rely on the cupin superfamily of proteins. Cupins are proteins characterized by a structure formed of barrels of four β sheets containing the active site and are surrounded by alpha helixes.[11] Pirin Like protein (PLP) and Mannose-6-Phosphate Isomerase (PMI) are similar bicupins characterized by a structure formed of two β barrel cupin domains but each of the two enzymes possess one active site in only one of the cupin domains.[13,14]

Therefore, we plan to design a functional chimeric enzyme by combining the catalytic domains of PMI and PLP, which should avoid denaturation and lead to a correct folding.

CHAPTER 2

PHOSPHOMANNOSE ISOMERASE

PMI



CHAPTER 2: PHOSPHOMANNOSE ISOMERASE

2.1. INTRODUCTION:

Prior to the preparation of the chimeric PMI-PLP metalloproteins we proceeded to express and purify each of the two starting enzymes i.e. PMI and PLP and to test their activities before and after exchanging the metal cations of their respective active site. We plan to compare these activities with those of the chimeric PMI-PLP metalloproteins.

In this chapter, after a short presentation of the structures and properties of the various variants of PMIs, we will then describe the preparation and the characterization of a novel metalloprotein based on PMI from *Candida albicans* by substituting the native Zn^{2+} of its active site by a Cu^{2+}.

The construction of the expression vector encoding for PMI and His_6 tag flanking TEV protease cleavage site ($TEV_{c/site}$) will first be explained. The cloning and the overexpression of the His_6-$TEV_{c/site}$-PMI in *E. coli*. will be presented. After exposing the purification and characterization of the obtained protein, its PMI activity will be discussed. Subsequently, PMI transformation into an artificial metalloprotein by exchanging the native Zn^{2+} by Cu^{2+}will be described. The preparation and characterization of the artificial metalloprotein by various spectroscopic techniques will be detailed and finally, an activity study will be presented.

2.2. STATE OF THE ART

PMI's are metal-dependent aldose-ketose isomerases catalyzing the interconversion of *D*-mannose-6-phosphate (M6P) and *D*-fructose-6-phosphate (F6P) in eukaryotic and prokaryotic organisms.[13] In 1968, PMI was suggested, for the first time, to be a metalloenzyme on the basis of studies that showed that chelating agents inhibited its activity and that this activity was regenerated by addition of metal cations, such as Co^{2+}.[57] It was also found that this metalloenzyme was a bicupin containing one atom of zinc per enzyme monomer (MW = 45 kDa).[58] This metal was located at the active center of the enzyme and it appeared to be essential for the catalysis since a substrate can protect the enzyme from inactivation by metal chelators.[57–59] PMI was thus the first isomerase for which a metal was established as an integral part of the enzyme structure.[58,59]

Different studies have shown that PMI is required for the survival, virulence, and/or pathogenicity of several bacteria and protozoan parasites[60,61], as well as for the yeast cell wall integrity in Saccharomyces cerevisiae[62] , Candida albicans[63], Cryptococcus neoformans[64], and Aspergillus nidulans.[65] In addition, PMI was found an essential enzyme for E.*coli*:[66] Mutants lacking PMI could not multiply on a medium where mannose was the only source of carbon and could not synthesize their capsular completely neither.[66] As a result, PMI is being explored as a possible target against fungal infections that cause significant disease or death, particularly in immunocompromised people.[59,67]

2.2.1. Different types of PMI: classification of Proudfoot

Based on their amino acid sequence identity, PMIs have been classified into three unrelated families as proposed by Proudfoot *et al.*[68]

The organisms that produce the most common Type I, Type II and Type III PMI enzyme families are summarized in Table 3.[58,68]

Table 3: Different organisms providing different types of PMI.

PMI Type I	**PMI Type II**	**PMI Type III**
Saccharomyces cerevisiae	*Escherichia coli*	*Rhizobium meliloti*
Aspergillus nidulans	*Salmonella typhimurium*	
Candida albicans	*Acetobacter xylinum*	
Leishmania mexicana	*Acinetobacter calcoaceticus*	
Escherichia coli	*Helicobacter pylori*	
Salmonella enterica	*Pseudomonas aeruginosa*	
Cryptococcus neoformans	*Rhodospirillum rubrum*	
Caenorhabditis elegans	*Xanthomnas campestris*	
Streptococcus mutans		
Homo sapiens		

2.2.1.1. Type I PMIs:

This group contains PMIs of identified eukaryotic organisms including human PMI, PMIs of several yeasts as well as PMIs of most bacteria and parasites.[57] Type I PMIs are monofunctional enzymes that catalyze only the reversible isomerization of M6P into F6P. They are monomeric enzymes, with a molecular weight of about 45 kDa that include a zinc cation (Zn^{2+}) within their active site. Their optimal pH varies between 7 and 8.5. [67]

Type I PMI are homologous proteins with greater than 30 % sequence identity. The ligands of the zinc cation within the active site are conserved from one enzyme to another.[57,67] The replacement of Zn^{2+} with other metal cations such as Ni^{2+}, Cd^{2+}, Mn^{2+}, Cu^{2+} or Fe^{2+} causes a loss of their PMI activity. [57] However, an increase in the isomerase activity was noted when Zn^{2+} was replaced with Co^{2+}.[57] Thus, these enzymes poorly tolerate changes in their metal cofactor. [57]

2.2.1.2. Type II PMIs:

Type II PMIs are classified as bifunctional enzymes with a molecular weight of 55 kDa. They are found exclusively in bacteria. They are involved at two distinct levels of the metabolism of mannose: in addition to their isomerase activity, these enzymes also possess a guanosine diphosphate (GDP)-D-mannose pyrophosphorylase (GMP)

activity, the last phase of the production of GDP-mannose (GDP-D-mannose).[68] Each activity occurs in a separate domain.[58] GDP-D-mannose is a precursor of manno-conjugated substrates (GDP-L-fucose, GDP-colitose, GDP-perosamine, and GDP-D-rhamnose) which are important for cell survival of both eukaryotes and prokaryotes and for the pathogenicity of most microorganisms.[58]

Type II PMIs require the presence of a metal in their active site to be functional. Unlike type I PMIs, these enzymes tolerate variation in the metal cofactor. They are active not only with zinc, but also with a wide range of divalent cations such as Co^{2+}, Mg^{2+}, Ni^{2+}, Mn^{2+} and Ca^{2+}.[60]

These enzymes present a low sequence similarity with those of type I PMIs that is limited to a small number of amino acids that belong to their PMI domain.[7] The conservation of these residues between type I and type II PMIs thus suggest a common catalytic process for their isomerase activity.[58]

Some bacteria, such as *E. coli* and *S. typhimurium,* possess the 2 types of PMI. The reason for this remains unclear. Proudfoot *et al.*[68] proposed a hypothesis assigning different roles to the two types of enzymes: PMIs type I could preferably be involved in the metabolism of D-mannose, whereas PMI type II could be involved in the synthesis of polysaccharide capsules.

2.2.1.3. Type III PMIs:

This type of enzymes is limited to the Gram-negative bacterium *Rhizobium meliloti* and presents a low sequence similarity with those of type I and type II PMIs.[66,68]

In 1992, Schmidt et al. isolated and established the complete nucleotide sequence of the PMI gene of R. *meliloti*. This nucleotide segment showed two open reading frames (ORFs) that code for 18- and 43-kDa polypeptides. According to gene disruption studies, the ORF2 encoded for PMI. Hence, there was no substantial resemblance between the deduced amino acid sequence and the equivalent sequences of the Pseudomonas aeruginosa and Escherichia coli PMIs, indicating that the isolated gene produces a novel type of PMI.[66]

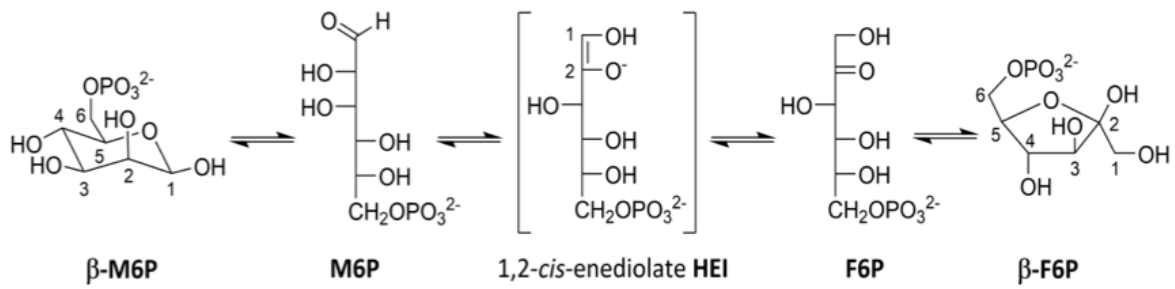
2.2.1.4. A novel type: Type IV PMIs:

These enzymes have been recently described in the literature.[69,70] They present proteins with dual specificity: PMI and Phosphoglucose isomerase (PGI).[70] In general, PGI has a crucial role in sugar metabolism and it catalyzes the interconversion of G6P into F6P.[69] These enzymes belong to the PGI superfamily because their folding is similar to the inner core of PGIs from eucaryotes and bacteria and they possess a PMI activity.[71]

They include DmPMI (*Dehalococcoides mccartyi* PMI) which was proposed to be a hypothetical protein and a bifunctional PMI/PGI as well as the PMIs of the actinobacteria, firmicutes and bacteroides families.[69] They share a low sequence similarity with those of other types of PMI and their activity is lower.[69] This lower in activity was suggested to be correlated to the fact that the substrate could be a different sugar than M6P.[69]

2.2.2. Reaction catalyzed by PMI

PMI specifically catalyzes the reversible isomerization of M6P into F6P. It has been demonstrated that this reaction proceeded through a proton transfer mechanism between the two carbon atoms C1 and C2 of the substrate, involving the formation of 1,2-cis-enediolate High Energy Intermediate (HEI),[72] as shown in Scheme 9.



Scheme 9: Reversible isomerization of D-mannose-6-phosphate (M6P) to D-fructose-6-phosphate (F6P) catalyzed by PMI.

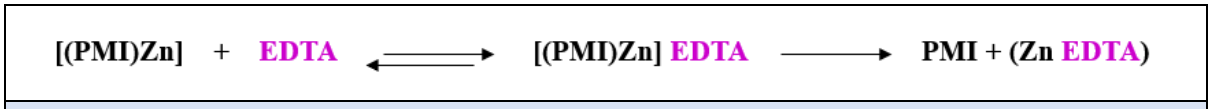
PMI is an enzyme widely used in nature as it is involved in the metabolism of mannose allowing the synthesis of mannosylated structures and thus glycoconjugates.[13] However, it is less widespread in plants and the gene coding for the PMI of *Escherichia coli* was recently used as a marker gene for the selection of genetically modified plants. Only plants that are transformed by the PMI gene are able to grow on media containing mannose as a single source of carbon.[73]

2.2.3. PMI: Characterization as a Zinc metalloenzyme

In 1958, Noltmann and Bruns[74] discovered that the activity of the PMI from pig erythrocytes was inhibited by chelating agents such as versene, dipyridyl and o-phenanthroline[74] and that this activity could be regenerated after addition of divalent metal cations such as Ca^{2+}, Zn^{2+}, Co^{2+} and Fe^{2+}. This led them to suggest that PMI is a metal-enzyme complex.[74]

In 1968, this observation was later confirmed by Gracy and Noltmann who purified for the first time the PMI from *Saccharomyces cerevisiae*.[75] The inhibition by chelating agents such as EDTA (Scheme 10), and o-phenanthroline was pH-dependent, occurring faster and easily at pH 6 than at pH 8.[57] The chelating agent was treated with a stoichiometric quantity of metal which results in the coordination of the agent's binding sites, reducing its ability to interact with the enzyme's metal.[57] Hence, the inhibition could be prevented by blocking the coordination sites of the metal-binding agents.[57] The regeneration was instantaneous and was dependent on the metal species as well as on its concentration: Metals binding tightly to EDTA such as Zn^{2+}, Co^{2+} and Fe^{2+} were the most powerful metals to reverse the inhibition of EDTA; although, metals which bind much lower to EDTA such as Ca^{2+} and Mg^{2+} were not capable to reverse the inhibition of EDTA.[57]

In addition, Bruns, Noltmann and Willemsen have shown that zinc was able to reverse completely the inhibition of the erythrocyte PMI by EDTA.[57]



Scheme 10: Reaction of EDTA with PMI.

Then, Gracy and Noltmann[57] showed via metal analyses performed by atomic absorption spectrophotometry that the PMI of yeast was a monomeric enzyme having a molecular weight of 45 KDa and containing a Zn^{2+} cation per protein. The authors reported that this divalent cation played an important role in the catalysis of the reversible isomerization of M6P into F6P. Indeed, the enzyme was protected against chelating agent's inhibition by the substrates, which confirmed first, that the metal was indeed in the active site of the enzyme and that it bound the substrate, and second, that this metal was essential for the catalysis.[57]

PMI was the first isomerase for which a metal was established as an integral part of the enzyme.[76] Its zinc cation maintains its structural conformity and confers a catalytic role.[13]

2.2.3.1. PMI from Candida albicans

PMI from *Candida albicans* is a type 1 PMI. It is a bicupin that contains two domains, each composed of a barrel of β-sheets, which are connected by a single small inter-domain composed of α-helixes (Figure 16). It also contains one atom of zinc in the active site, located in the β-sheet barrel domain at the C-terminal side of the protein (Figure 16).[11]

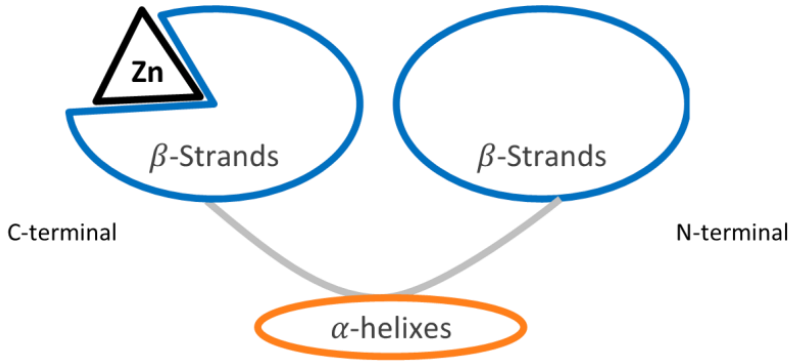


Figure 16: Schematic representation of the structure of PMI showing a large inter-domain between two conserved barrels containing Zinc in the active site located on the C-terminal side.

In 1996, Clesaby *et al*[59] reported the crystal structure of PMI. It showed a monomeric protein that has three domains: The C-terminal central β-barrel catalytic domain flanked by an helical domain on one side and a N-terminal β-barrel domain on the other side (Figure 17).

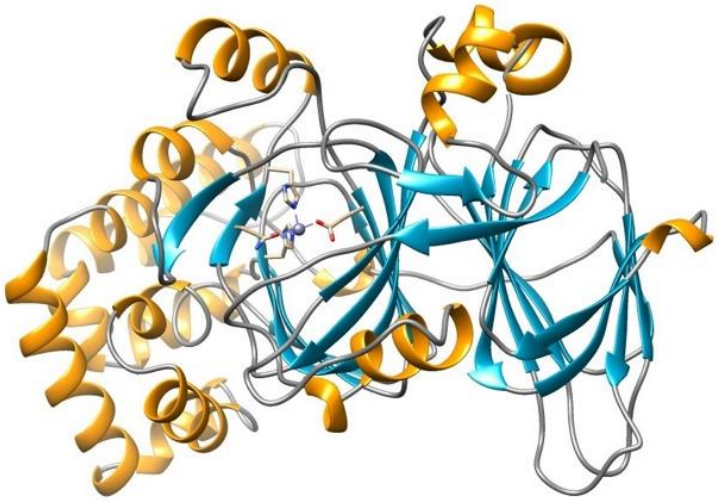


Figure 17: 3D structure of PMI showing a central β-barrel catalytic domain flanked by helical domain on one side and a N-terminal β-barrel domain on the other side.

The catalytic site is located within the central domain and forms a deep, open cavity whose dimensions are compatible with those of its substrates (M6P/F6P) in association with a zinc atom. The zinc is pentacoordinated: its coordination sphere comprises four amino acids including two Histidines H285 and H113, a glutamate E138, a glutamine Q111 and a water molecule. The coordination geometry is a trigonal bipyramid, that generally favors a catalytic role of zinc in metalloproteins (Figure 18).[59]

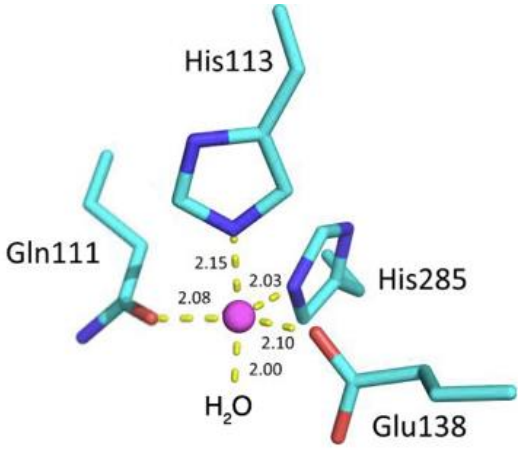
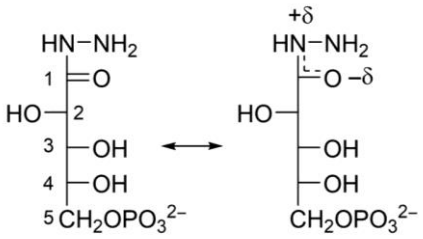


Figure 18: View of the metal binding site in PMI (pdb : 1PMI). Interatomic distances are given in Å.

F6P was located in the binding site in a position that suggested its coordination to Zn^{2+}. However, the electron density around the F6P was unclear, which was justified by the authors by the co-existence of the cyclic fructofuranose form of ß-F6P and its partially open form. [59] Nevertheless, this coordination proves that the zinc has not only a structural role but also a catalytic role in the reaction process.[13]

In the hosting laboratory, phospho-D-arabinohydrazide (5PAHz) was designed as an analogue of the enediolate high-energy intermediate in the mechanism of the M6P/F6P isomerization catalyzed by PMI (Scheme 11).



5PAHz

Scheme 11**:** Structure of the HEI analogue inhibitor 5PAHz.

This compound caused a strong inhibition of PMI (Ki = 1.7 µM). A His_6 tagged variant of PMI was also expressed, purified, and crystallized in the presence of 5PAHz. The structure of PMI complexed with 5PAHz at 1.85 Å resolution was obtained (Figure 19).[13]

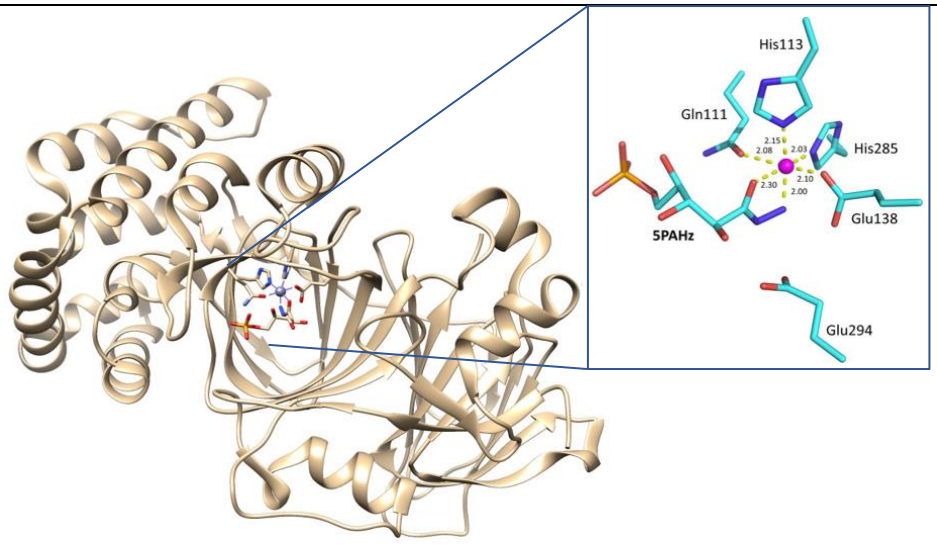


Figure 19: Crystal structure of His_6-PMI coordinated to the inhibitor 5PAHz.

This structure suggests that Glu294 was the catalytic base that transferred a proton between the C1 and C2 carbon atoms of the substrate (Scheme 11). It also showed the bidentate coordination of the inhibitor on Zn^{2+} which explained the stereoselectivity of the reaction (Figure 19). [13]

2.2.4. Depletion of PMI and its consequences

PMI constitutes a potential therapeutic target. The effects of the lack of PMI have been studied in most fungi and yeasts. These studies have shown that fungal microorganisms lacking PMI were unable to secrete the glycoproteins constituting their cell wall. However, supplementing the media with D-mannose allowed in some cases to compensate for PMI deficiency.[77] Nevertheless, mannose supplementation does not make it possible either to overcome PMI failure in *Aspergillus nidulans*, or to restore the morphological defects caused by the absence of this enzyme in *C. neoformans*.[64] These studies show the importance of PMI for the growth and survival of these microorganisms, it then appears as a potential therapeutic target for conceiving new antifungals.[64] A study by the team of Wills *et al*. supports this hypothesis: the absence of PMI in *C. neoformans* mutants not only causes an arrest of growth "*in vitro*", but also a loss of virulence of the pathogens "in vivo", showing for the first time the validity of PMI as a potential antifungal target.[64] As for human PMI, this latter has not been widely studied; it has been shown that patients suffering from leukemia present reduction in PMI activity.[64]

2.3. RESULTS AND DISCUSSION:

2.3.1. Redesign the expression vector for the PMI

The pRSET-C vector comprising an insert encoding for the His_6-tagged PMI (His_6-PMI) was available in host laboratory. This plasmid comprises a His_6 tag which should be cleaved before the metal substitution because it's known to be an excellent ligand for metal cations. For this, we planned to insert a specific $TEV_{c/site}$ flanked by the sequence of the enzyme and that of the His_6 tag.

The initial plasmid pRSET-C was first digested by the restriction enzyme NheI. The result of this reaction was analyzed by agarose gel electrophoresis and ethidium bromide (ETBr) staining (Figure 20). A single band around 8 Kb was visible after digestion by Nhe1 (lanes 1 and 2) which confirmed that the digestion of the plasmid was successful. For the undigested plasmid, three different bands can be seen (lane 3) which can be assigned to the three different structures of the plasmid DNA: nicked-relaxed, supercoiled and circular single-stranded.

The digested plasmid was extracted from the gel and then used to insert the DNA encoding for $TEV_{c/site}$ (ENLYQFS).

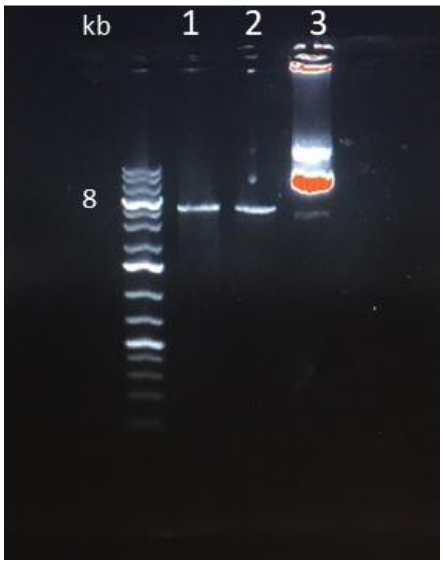


Figure 20 : Agarose gel stained with ETBr of the undigested (Lane 3) and digested plasmid (lanes 1 and 2). Lane 3 shows 3 possible plasmid conformations.

The DNA insert encoding for the $TEV_{c/site}$ was a synthetic double stranded DNA that was complementary to the remaining palindromic fragments of the Nhe1 restriction sites in the linearized plasmid (Figure 21). In the presence of ligase, the synthetic insert

could ligate in the vector in two directions as both ends were complimentary to the palindromic remaining fragments of the Nhe1 restriction site: either in the direction coding for $TEV_{c/site}$ or in the direction opposite to that coding for $TEV_{c/site}$ resulting in the amino acid sequence ASHALKIQVF. It is also noteworthy that, because of these palindromic fragments at the ends of the vector, the addition of ligase could also yield the starting circular plasmid thus without the incorporation of the insert.

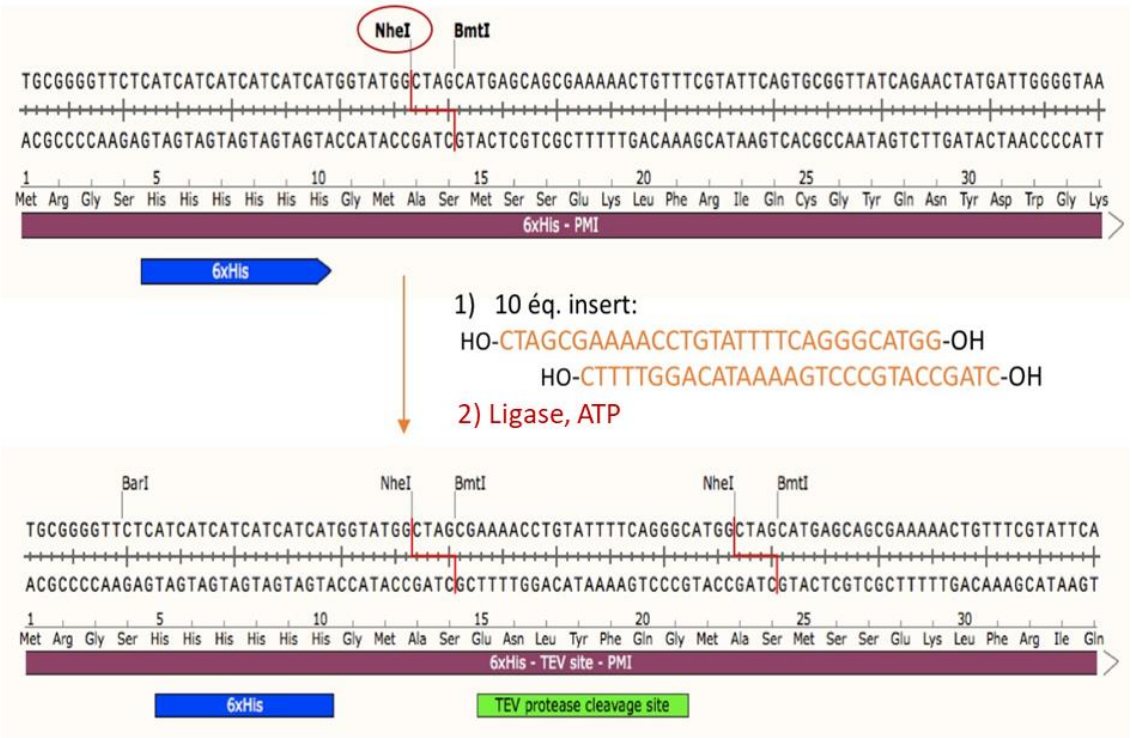


Figure 21: Ligation of the insert coding for $TEV_{c/site}$ and the vector linearized plasmid. The insert is complementary to the remaining Nhe1 fragments of both vector strands.

A mixture was formed by adding 10 equiv. of double stranded synthetic DNA insert encoding for TEV protease to 1 equiv. of the purified digested plasmid. Bacteriophage T4 DNA ligase and ATP were added to the mixture. The excess of insert was aimed at limiting the self-ligation of the vector without incorporation of insert.

The ligation mixture was then used to transform DH5alfa competent bacteria. The growth of bacteria following transformation in antibiotic containing medium allows the selection of recombinant cells if the newly acquired DNA conveys an antibiotic

resistance. In the present case, among the selected bacteria about 50 colonies were further screened by PCR for the correct insertion of the TEV protease DNA.

Colony screening using PCR is a convenient high-throughput method that can be used for determining the presence or absence of an inserted DNA in plasmid constructs and can also help to determine the orientation of the insertion. PCR amplification of a plasmid fragment using an insert specific primer paired with a vector specific primer provided amplicons of a specific size (700 bp) only if the insert is indeed present in the plasmid and in the correct direction. This was achieved using primer mix 1: CCTGTATTTTCAGGGCATGG and TTCAGGCCAACATGATTCAG (Figure 22).

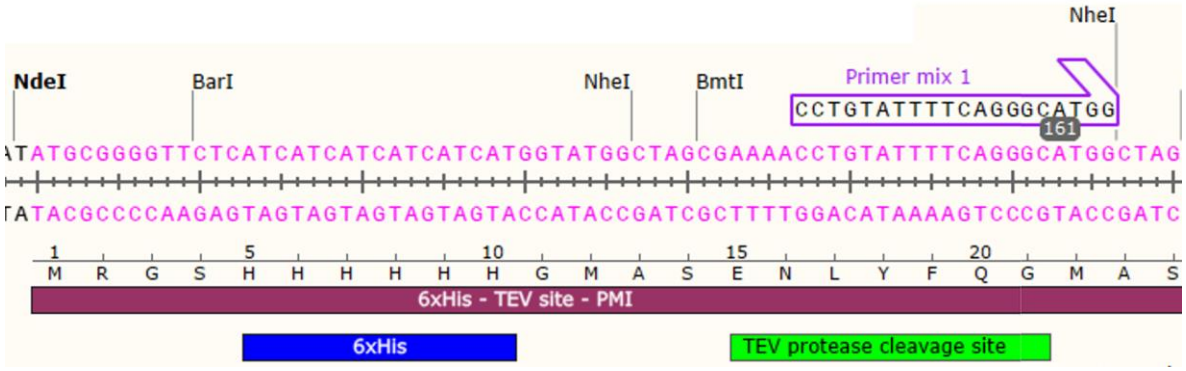


Figure 22: Fragment of the sequence encoding for His_6-PMI and comprising the inserted $TEV_{c/site}$ with the insert specific primer (violet frame, primer mix 1) that was used for bacterial colony-screening.

The success of the PCR is dependent on the presence of the expected insert and on the reaction conditions. To optimize PCR conditions and to avoid negative results due the experimental conditions, we relied on a set of primers that targeted a DNA sequence that was surely present in the plasmid. All selected colonies acquired ampicillin resistance by incorporating the transformation plasmid. Accordingly, the reaction conditions were optimized using primer mix 2 (positive control), that targeted the sequence of the gene encoding for ampicillin resistance to provide amplicons of about 700 bp and showed physical properties similar to those of primer 1 (Tm = 54.1-55 °C, Annealing T = 48-49 °C).

PCRs were performed and the samples were analyzed by electrophoresis/Syber safe staining. Figure 23 shows that using primer mix 2, a fragment of amplified DNA of about 700 bp was formed, proving that, as expected, all selected colonies contained the sequence of the gene encoding for ampicillin resistance. Using the same conditions with primer mix 1, a fragment of amplified DNA of 700 bp was detected only for

colonies of lanes 7 and 12 (Figure 24). The plasmid containing the insert encoding for the $TEV_{c/site}$ was thus only incorporated in the desired direction in the bacteria of these two colonies.

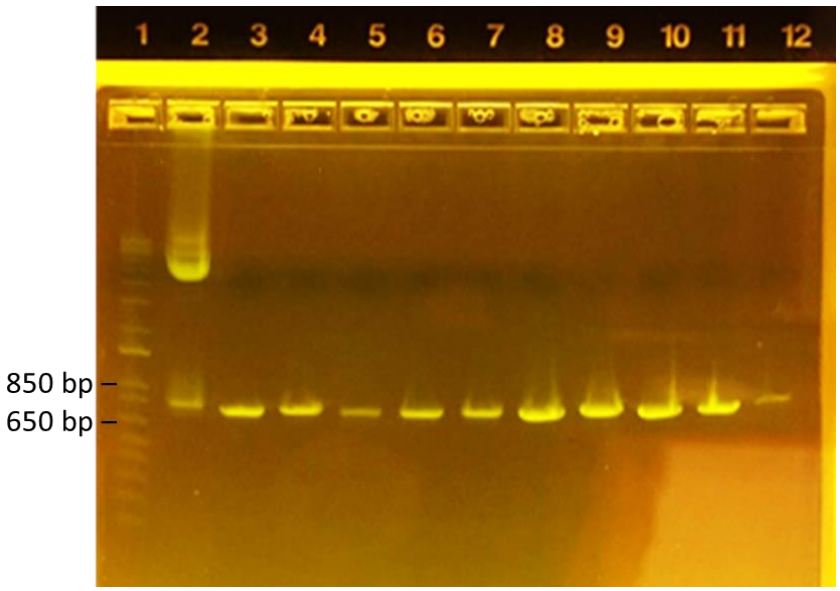


Figure 23: Agarose gel electrophoresis of PCR products obtained using the positive control primers (primer mix 2) which are complementary to the fragments of the sequence encoding for ampicillin resistance. Lane 1: DNA Ladder

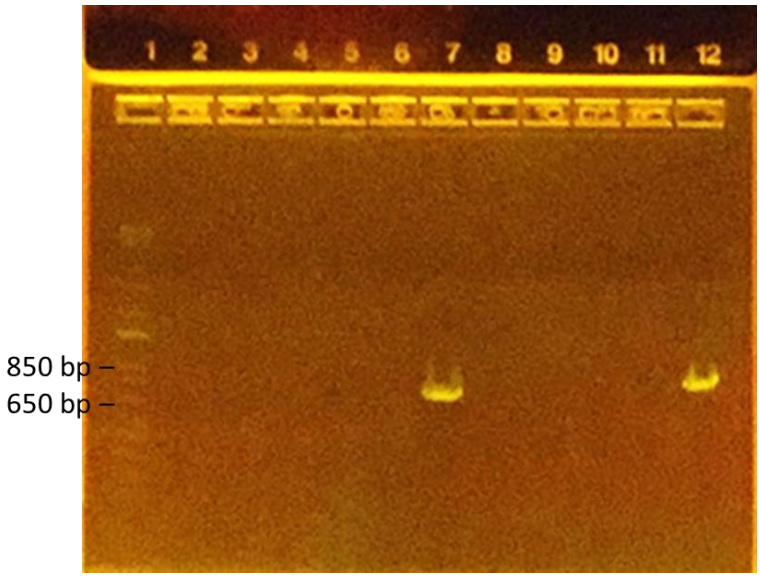


Figure 24: Agarose gel electrophoresis of PCR products obtained using the primer mix 1 which is complementary to fragments of the sequence of the insert i.e. $TEV_{c/site}$ and the vector. Lane 1: DNA Ladder

In order to confirm these results, and to examine the junctions between the plasmid and the inserted DNA, colonies that provided lanes 7 and 12 were grown and their plasmid DNA extracted and sequenced. Indeed, they show the expected sequence.

2.3.2. Expression and purification of the His_6-PMI:

E. coli BL21(DE3) strain was used for expressing the protein. B strains are deficient in lon protease (cytoplasm) and OmpT protease (outer membrane) therefore they are preferred for the expression of recombinant proteins. The strain used contained the λDE3 lysogen that carries the gene for T7 RNA polymerase under control of the lacUV5 promoter. IPTG addition is thus required to maximally induce expression of the T7 RNA polymerase in order to express recombinant genes cloned downstream of a T7 promoter. Accordingly, the PMI gene was located downstream of the T7 promoter. Purified and sequenced plasmid DNA containing the insert sequence was used to transform BL21(DE3) that were then selected on agarose/ampicillin.

A pre-inoculum from a single colony was grown in a starter culture LB-ampicillin. TB supplemented with ampicillin was inoculated from the preinocculum at 1/100 and grown at 37 °C until the OD at 600 nm reached a value between 0,6 and 0,8. 1 mM IPTG was then added and incubation was continued at 20 °C for 18 h. Bacterial cells were pelleted and kept at -20 °C. Analysis by SDS-PAGE, followed by Coomassie blue staining, showed a band of high intensity at a MW of about 50 kDa in all the lanes corresponding to the bacterial pellets, which was characteristic of the expression of His_6-PMI. By contrast, no intense protein band was observed in the lanes corresponding to the supernatant fractions (Figure 25).

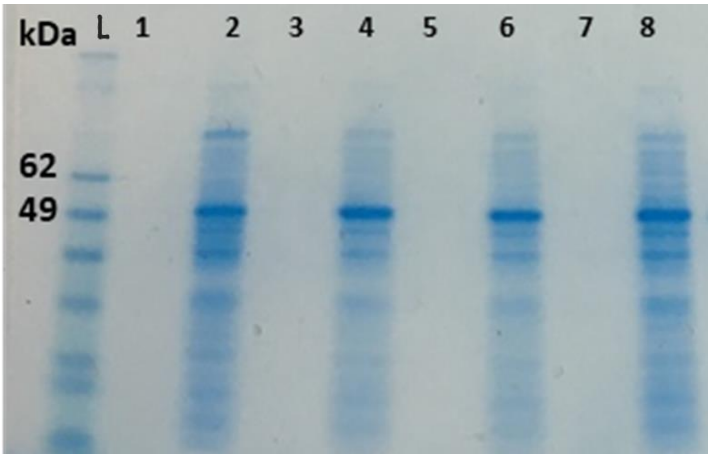


Figure 25: Coomassie blue stained SDS-PAGE of supernatant fractions and pellets after the expression of the PMI. Lanes 1, 3, 5 and 7: Supernatant fractions; Lanes 2, 4, 6 and 8: Pellet fractions L: Protein Ladder.

The bacterial pellets were resuspended in the presence of benzonase and lysozyme. The mixture was sonicated then centrifuged to eliminate all insoluble fractions. The clear solution was passed through Ni-NTA agarose column. The various fractions eluted from the column were analyzed by SDS-PAGE followed by Coomassie blue staining. As shown in Figure 26, a band corresponding to a protein with a molecular weight of about 50 kDa was observed for fractions 6 to 12, which suggested that those fractions contained the purified His_6-PMI.

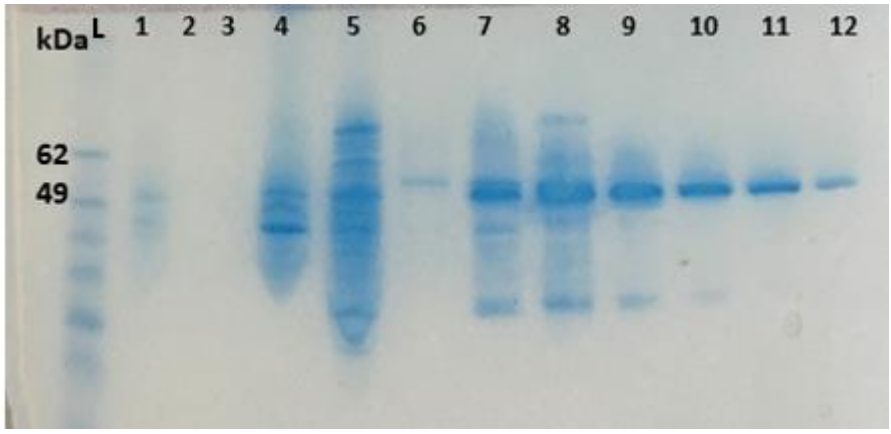


Figure 26: Coomassie blue stained SDS-PAGE of fractions collected after the purification on an Ni-NTA agarose column of His_6-PMI. Lanes 6, 7, 8 ,9, 10, 11 and 12 show bands corresponding to the expected molecular weight of pure His_6-PMI (of about 50 kDa); L: Protein Ladder.

To improve the expression of the His_6 – PMI, *E. coli* BL21(DE3) pLysS strain was used for expressing the protein. This strain allows high efficiency protein expression, express T7 RNA polymerase which can improve the expression and encode T7 lysozyme that suppresses basal expression of toxic target proteins prior to induction. The same procedure was followed as above to express and purify the protein. Figure 27 confirms that this strain is much better for the His_6-PMI's expression. A band with high intensity corresponding to a protein with a molecular weight of about 50 kDa was observed for fractions 3 and 4, which suggested that those fractions contained the purified His_6-PMI.

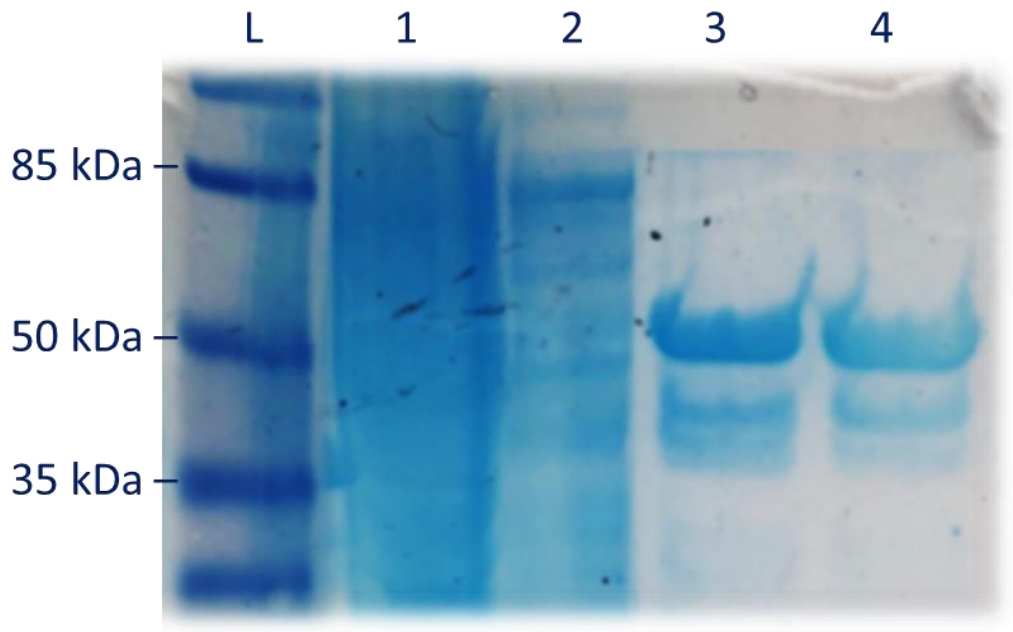


Figure 27: Coomassie blue stained SDS-PAGE gel of fractions collected after the purification. Lane 1: supernatant; Lane 2: washed fraction; Lanes 3 and 4 show bands with high intensity corresponding to the molecular weight of pure His_6-PMI (about 50 kDa); L: Protein Ladder.

2.3.3. Purification of His_6 tag free PMI:

The purified His_6-PMI protein fractions were incubated with His_6-AcTEV. Following digestion, His_6-AcTEV was easily removed from the reaction medium by purification on a Ni-NTA affinity column.

Indeed, SDS-PAGE analysis of the fractions eluted from the Ni-NTA affinity column, followed by coomassie blue staining revealed a characteristic band for the flow through at a molecular weight of about 50 KDa (Figure 28 left, lane 3), similar to those of the starting His_6-PMI protein (Figure 28 left, lane 1) and of the reaction mixture (His_6-PMI + His_6-AcTEV) (Figure 28 left, lane 2), while no band was found for the fraction eluted with 300 mM imidazole (Figure 28 left, lane 4).

The SDS-PAGE analysis was also followed by a Western blot using anti His_6-tag antibody (Figure 28). The comparison with the coomassie blue staining experiment confirms that the starting protein was indeed His_6 tagged (Figure 28 right, lane 1) while after incubation with AcTEV the His_6-PMI had lost its His_6 tag (Figure 28 right, lane 2). Similarly, no band was observed in the lane 3 of the western blot corresponding to the flow through of the Ni-NTA column, which suggested that it indeed contained the tag

free PMI. However, it appears that the fraction eluted with imidazole still contained a low amount of His_6-PMI that had not been digested (Figure 28 right, lane 4).

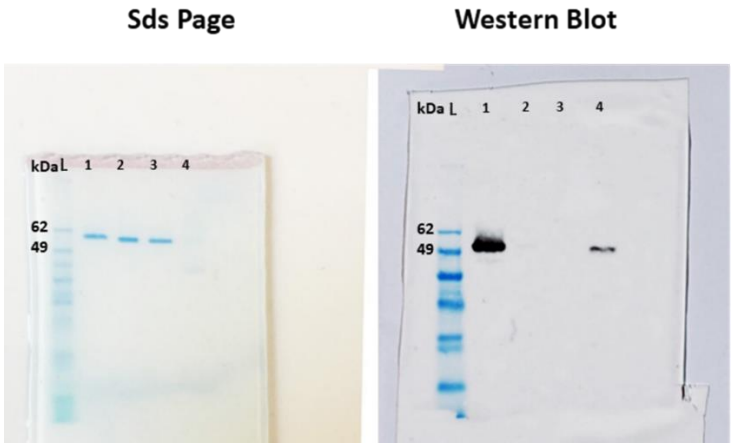
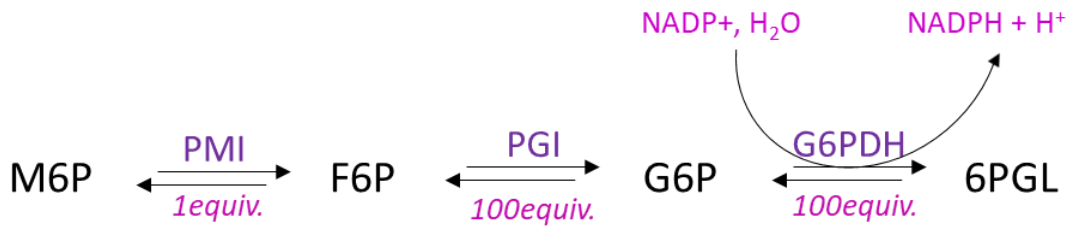


Figure 28: Left: Coomassie blue stained SDS-PAGE; Right: Western Blot image. In Both cases lanes: 1-PMI His; 2- PMI-His + AcTEV; 3- Flow through; 4- Eluted fractions. L: Protein Ladder

2.3.4. Assay of enzyme activity:

The enzymatic activity of PMI was assessed *via* the measurement of the rate of isomerization of M6P into F6P at 25°C using a method described by Gracy and Noltmann.[75] This method is based on three successive reactions: (1) the isomerization of M6P into F6P catalyzed by PMI followed by (2) the isomerization of F6P into G6P (glucose-6-phosphate) catalyzed by PGI (PhosphoGlucose Isomerase) and, finally, by (3) the oxidation of G6P to 6PGL (6-phosphogluconolactone) by G6PDH (Glucose-6-phosphate dehydrogenase) (Scheme 12). The later reaction requires $NADP^+$ that gets reduced into NADPH which can be detected by following its characteristic absorption at 340 nm (molar extinction coefficient = 6220 $M^{-1}.cm^{-1}$). Both auxiliary enzymes (PGI and G6PDH) are added in large excess and the activity is measured using both pure tag free PMI and His_6-PMI.



Scheme 12: The three successive reactions used to assay PMI activity by uv-visible spectroscopy. Λmax NADPH = 340 nm

By following this method, the initial velocity (Vi in mM/min = U) of NADPH+ formation increased upon the addition of increasing volume of His_6-PMI and PMI tag free. The recalculated slope of initial velocity increased with added volume of enzymes provided a specific activity of 10^{-4} U/µL and 7. 10^{-5} U/µL for His_6-PMI and PMI tag free, respectively (Figure 29). These results show that the two enzymes had a similar specific activity of 10^{-4} U/µL.

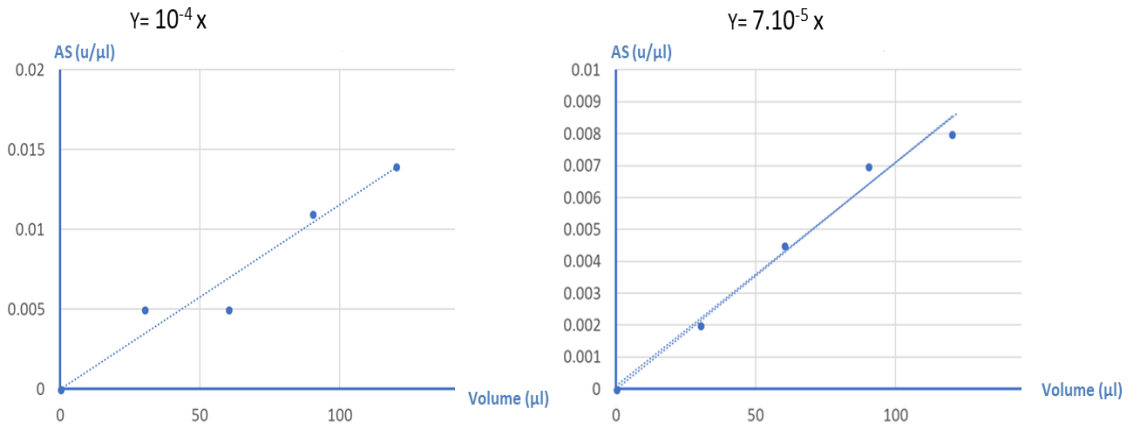


Figure 29: Pure His_6-PMI and tag free PMI specific activity (AS) as function of the volume of enzymes at 25°C in 50 mM MOPS, pH 7.4 in the presence of all substrates (M6P and $NADP^+$) and auxiliary enzymes (PGI and G6PDH) as indicated. The reported values are initial velocity (slope of the kinetic divided by the extinction coefficient of NADPH).

2.3.5. Zinc (II) removal:

It was previously reported by Gracy and Noltmann that up to 85 % of the PMI's activity could be lost by incubation with 10 000 equiv. of EDTA at 30°C for 8 h.[57] Metal concentration measurements also indicated a loss of Zn^{2+} upon incubation with EDTA.[57] This confirms that the removal of the metal cofactor, Zn^{2+} from the active site of PMI leads to a loss of the activity.

The effect of the addition of increasing amounts of EDTA (0-10^6 equiv.) on PMI at 4°C in 25 mM MOPS was thus investigated. A significant decrease of PMI activity was then observed upon incubation with increasing amounts of EDTA, which could be corelated to the amount of Zn^{2+} removed.[57] Indeed, upon addition of 10^4 equiv. of EDTA and after dialysis for 2 days, the enzyme loss 80 % of its activity whereas the addition of 10^6 equivalent of EDTA to PMI was required to achieve 98% removal of Zn^{2+} (Figure 30).

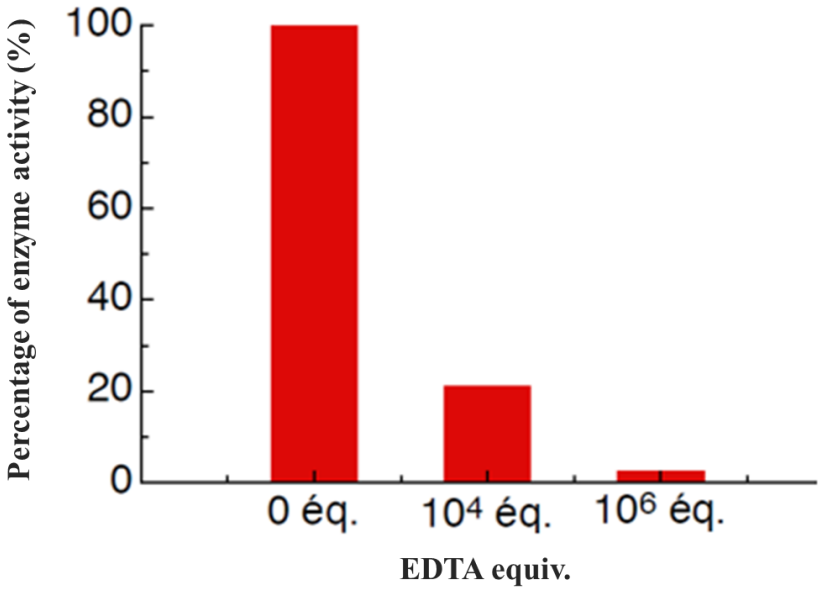


Figure 30: Percentage of the PMI activity as function of added EDTA equiv.

2.3.6. Incorporation of Cu(II) into zinc free PMI

The Incorporation of Cu(II) into zinc free His_6 tag free PMI was followed by fluorescence spectroscopy that is a very sensitive, rapid, and convenient method to study environmental changes in a protein, in particular those caused by the interaction of a protein with a ligand or a metal cation.

Upon addition of increasing amounts of copper nitrate (0.2-55 equiv.) in 50 mM MOPS, pH=7.4 to the zinc free PMI, 4.5 µM in 50 mM MOPS, pH = 7.4, the intensity of the protein intrinsic fluorescence signal increased (Figure 31). The fluorescence intensity value at 340 nm was then plotted against the concentration of added copper(II) nitrate, and the resulting curve plot (Figure 32) was fitted by using the following equation, which allowed the determination of the K_D (dissociation constant) value for the PMI-Cu(II) complex as well as the Cu/PMI stoichiometry:

$$\Delta F = \Delta F_{max} \frac{(n[P]_0 + [L]_0 + K_D) - \sqrt{(n[P]_0 + [L]_0 + K_D)^2 - 4n[P]_0[L]_0}}{2n[P]_0}$$

Where: F is the intensity of the fluorescence signal at 340 nm, $[P]_0$ and $[L]_0$ are respectively total protein and total copper(II) concentrations, n is the stoichiometry (number of ligand-binding sites per protein); $\Delta F_{max} = F_0 - F_{max}$; ($F_0$ and F_{max} are the

intensities resp. in absence of copper and when the protein is saturated with copper) represents the asymptotic value for ΔF. ΔF_{max}, K_D, and n were calculated from the fitting equation. A good fit was achieved with n = 1 and and K_D = 3 µM, indicating that one Cu^{2+} was bound tightly to one molecule of PMI to form a new Cu(II)-PMI protein the activity of which remain to be assayed.

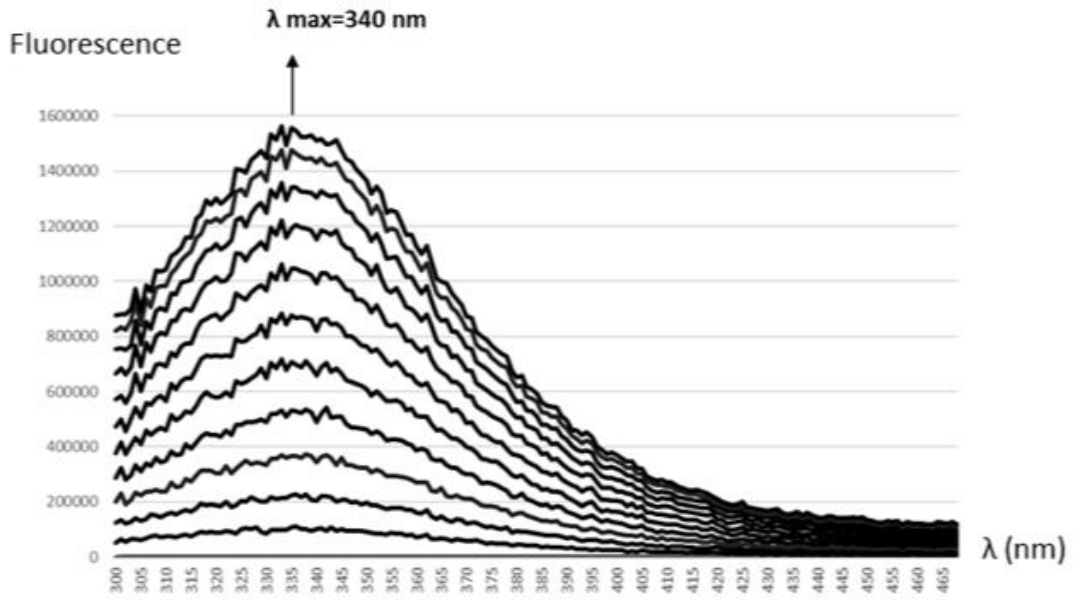


Figure 31: Fluorescence titration of the binding of $Cu(NO_3)_2$ to zinc free PMI upon excitation at 280 nm. Increase of the signal with increasing concentrations of added Cu^{2+} (0 to 55 equiv.). Measurements were performed in 50 mM MOPS, pH = 7.4 using 4.5 µM PMI.

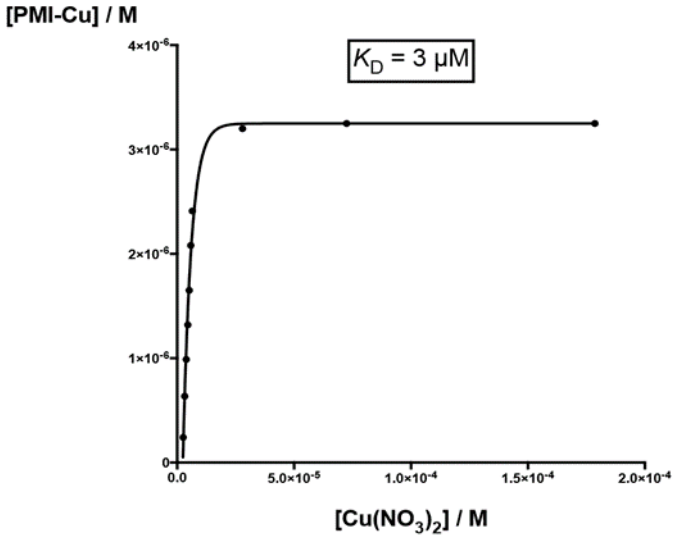


Figure 32: Curve fitting the concentration of Cu-PMI derived from the fluorescence values at 340 nm plotted against total added Cu(II) concentration.

2.3.7. Electron paramagnetic resonance EPR:

The new copper-PMI artificial metalloprotein was characterized by EPR spectroscopy. For this, the X-band EPR spectra of frozen solutions of zinc free PMI and both Cu(II)His_6-PMI and Cu(II)-PMI were recorded at 10 K. Zinc free PMI showed no EPR signal. Cu(II)-PMI (Figure 33 B) showed an anisotropic EPR signal with an axial symmetry that confirms the presence of a single Cu^{2+} binding site, that could be fitted with g// = 2.06, g⊥= 2.29 values, hyperfine coupling constants of A// = 16.7 mT.
In contrast, Cu(II)His_6-PMI (Figure 33 A) showed an isotropic EPR signal of at least two Cu^{2+} species, as seen in the g// part of the signal, and a total signal intensity equivalent to that of Cu-PMI. This suggests that the Cu^{2+} is binding both His_6 tag and the active center.

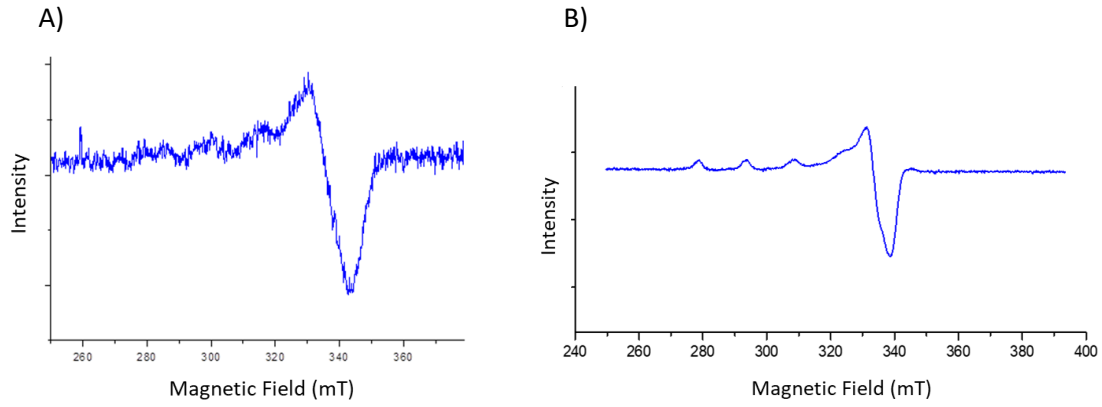


Figure 33: X-band EPR spectrum measured at 10 K of A) Cu(II)-His_6-PMI and B) Cu(II)- PMI.

2.3.8. TYCHO:

TYCHO measurements were done to verify the structural integrity and to determine the inflection temperature of the PMI. In Figure 34, TYCHO showed that the natural PMI is well folded and its thermal denaturation can occur at 60°C. Upon substitution of Zn^{2+} from its active site by Cu^{2+}, PMI retains its structure with little changes in the TYCHO signal.

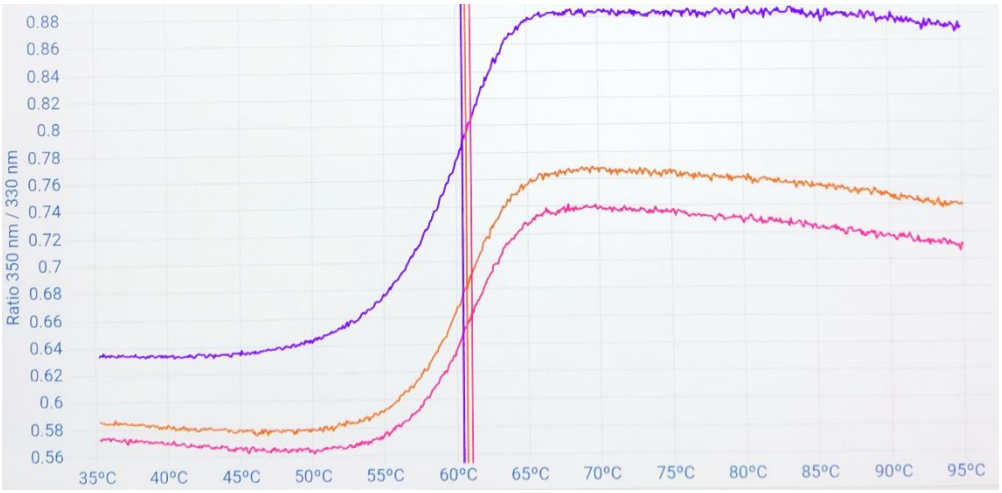
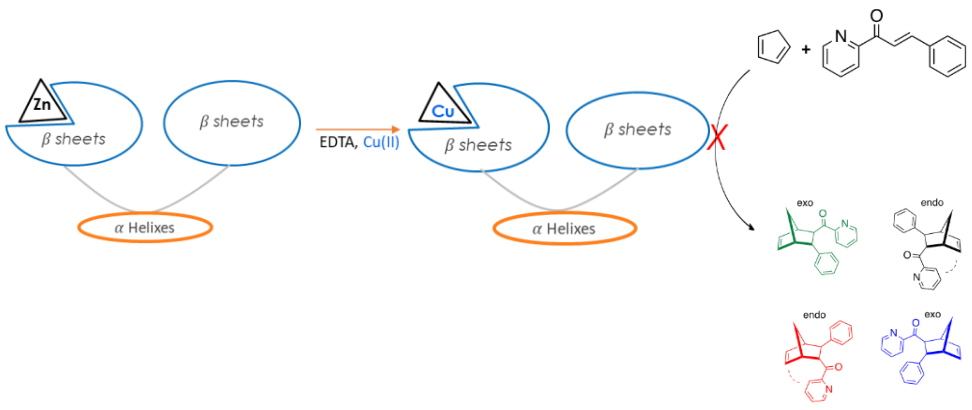


Figure 34: TYCHO measurements of natural PMI-Zn in purple; PMI zinc free in orange and Cu-PMI metalloprotein in pink.

2.3.9. Catalytic activities:

2.3.9.1. Diels-Alder cycloaddition

The Diels-Alder reaction is a cycloaddition of an alkene (dienophile) and a 1,3-diene with inter- and intramolecular forms being possible.[20] Because natural Diels-Alderases are rare, artificial Diels-Alderases are promising targets with a high potential for synthetic applications. it has been shown that a copper(II) is a good catalyst for this reaction.[5] First, we decided to asssay the catalysis of the Diels Alders reaction of hallmark reactants: Cyclopentadiene (CP) and 2-azachalcone (scheme 13) by the artificial Cu-PMI and natural PMI because Zn is also a good Lewis acid, thus may catalyze this reaction.



Scheme 13: Representation of Diels-Alder cycloaddition reaction of CP and Azachalcone by the artificial Cu-PMI enzyme

On the other hand, inspection of the crystal structure of PMI revealed that the active site is buried and only accessible through a narrow tunnel (Figure 35), which could prevent the access of Diels-Alder substrates.

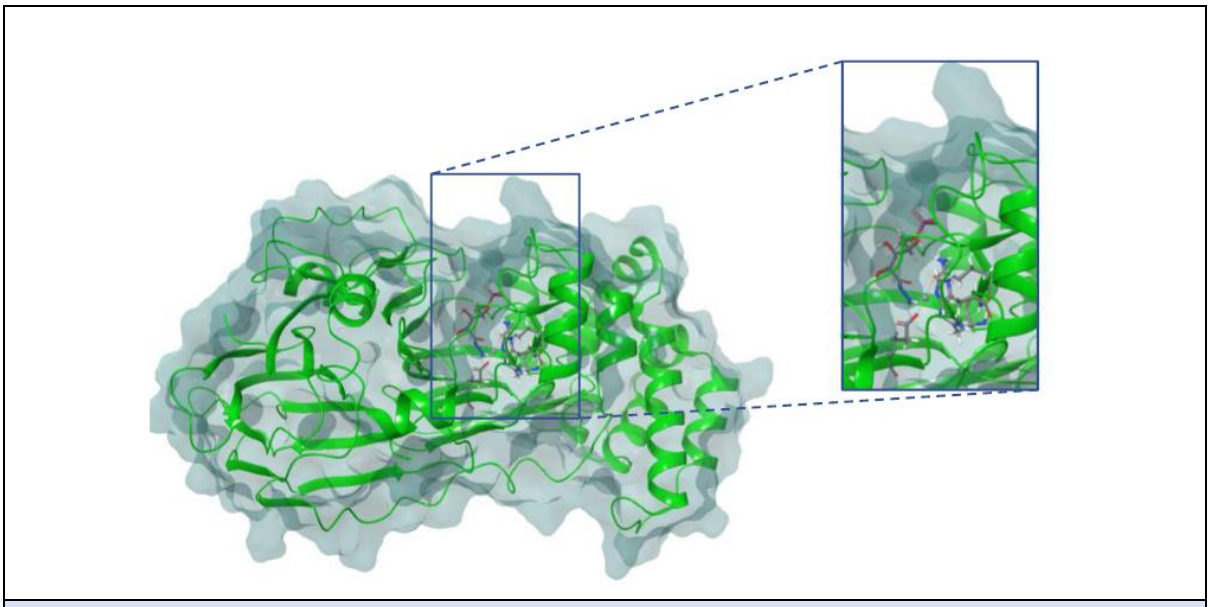
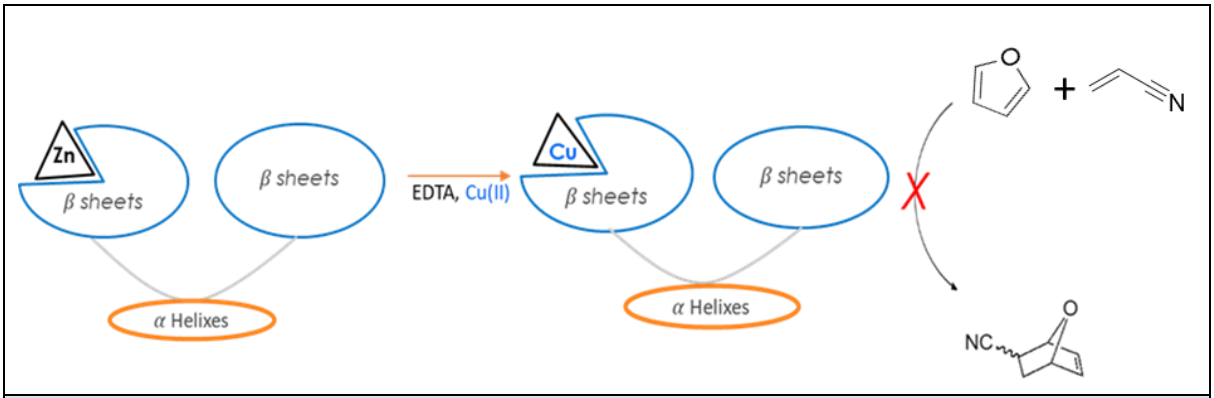


Figure 35: Left: Structure of PMI crystallized with 5PAHz; Inset: zoom on the active site.

Nevertheless, large chelators such as 8-hydroxyquinoleine were reported to inhibit PMI by accessing its active site and specifically binding Zinc(II).[57] Encouraged by a possible flexibility in the enzyme, the Diels-Alder cycloaddition of CP and 2-azachalcone was assayed but the reaction failed to proceed in the presence of either Zn-PMI or Cu-PMI.

Subsequently, the catalysis of the Diels-Alder cycloaddition of smaller commercially available substrates was aimed: Acrylonitrile and Furan. Here again, no Diels-Alder activity was observed (scheme 14) which could be explained by the fact that these are less reactive substrates.

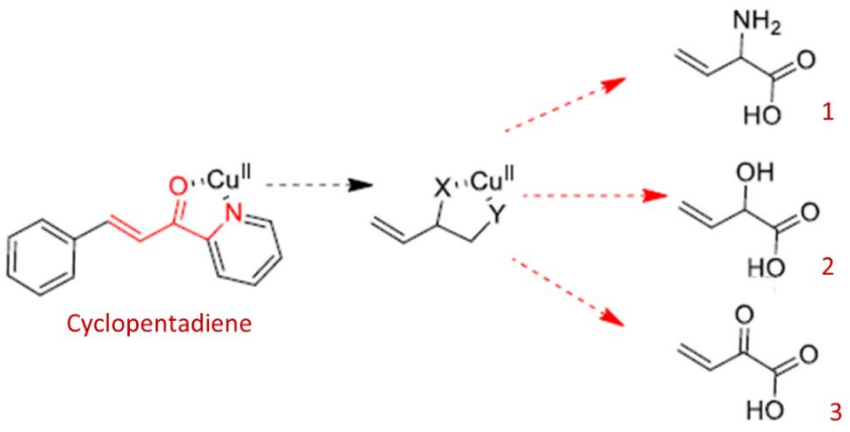


Scheme 14: representation of Diels-Alder cycloaddition reaction of Furan and Acrylonitrile by the artificial Cu-PMI enzyme.

Therefore, two strategies were followed. The first aimed at further i) reducing the size of the substrates, while the second aimed at the ii) catalysis of a reaction that does not require substrate-access to the active site.

i) Design and synthesis of smaller Diels-Alder substrates

Since PMI has a relatively small active site, three likely more suitable Diels-Alder substrates were designed. We designed structural analogs to the chelating motif of 2-azachalcone i.e. two heteroatoms arranged in a pinch able to chelate Cu(II), linked to a vinyl group by a simple bond. We set our sights on compound 1 keeping in mind the primary amine resulting in the lack of delocalization of the vinyl double bond thus a lesser effect of the coordination on the reactivity of this dienophile (Scheme 15). We therefore designed a second analog, compound 3 closely related to 2-azachalcone in terms of possible electron delocalization. Because 3 still bears the chelating motif, we expect that it will be able to bind Cu(II) in a bidentate way, and because it is significantly smaller than azachalcone it should be able to enter the active site of PMI. Alcohol 2, is an intermediate in the synthesis of 3 and although it is a weaker ligand of Cu(II) (vs. amines and carbonyls) and probably less reactive toward CP, it would be of interest in case the prospected artificial enzymes were highly efficient.

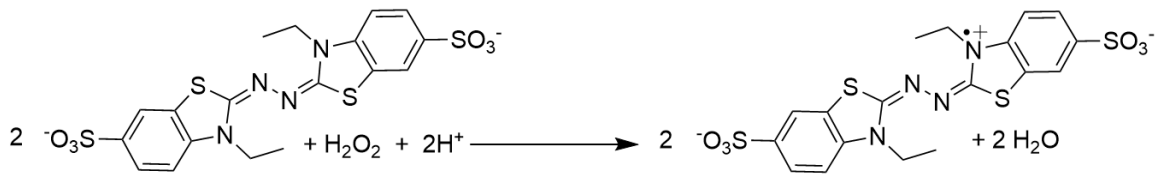


Scheme 15: Substrate design's strategy

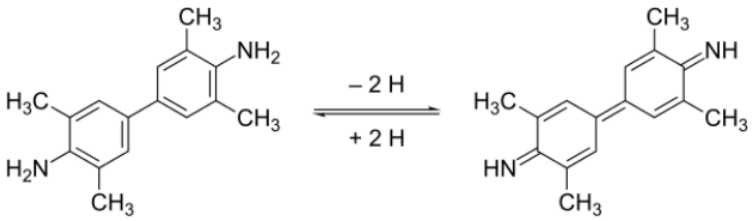
The synthesis of these compounds was the aim of the Master 2 thesis of Mr. Philippe Ibouanga currently a PhD student in the host laboratory. Two substrates over three were synthesized. The reactivity of these compounds as Diels-Alder cycloaddition substrates remains to be explored.

2.3.9.2. Oxidation reaction

In this second strategy, we proceeded to assay the possible peroxidase activity of Cu-PMI. Peroxidases are enzymes that catalyze the oxidation of a substrate by hydrogen peroxide or an organic peroxide as a catalyst.[78] Mostly, they are ferric heme proteins, but we can find also examples including superoxide dismutase containing copper[79] and even selenium-containing enzymes such as glutathione peroxidase.[78] Most peroxidases are devoid of substrate binding site. This catalytic reaction does not require substrate-access to the active site. The reaction proceeds by the oxidation of H_2O_2 at the metal cation in the active site. Then, the oxidized migrates outside the enzyme and react encountered substrates. The catalysis of the peroxidase activity of Cu-PMI was tested using H_2O_2 as oxidant and typical peroxidase co-substrates (co-S) such as 2,2'-azino-bis 3-ethylbenzothiazoline-6-sulfonic acid (ABTS) and 3,3',5,5'-tétraméthylbenzidine (TMB) (scheme 16 and scheme 17). Oxidized TMB takes on a blue-green color whose absorbance can be measured at 370 nm or between 620 and 655 nm, while the absorbance of oxidized ABTS can be measured at 420 nm. Preliminary results showed that the PMI-Cu artificial metalloenzyme does not catalyze the oxidation of TMB and ABTS. However, an important oxidation was obtained upon the addition of 60 equiv. of copper sulfate on the PMI-Cu enzyme. The binding of copper to the batch of apo-PMI was then evaluated by fluorescence titration. Fluctuation of fluorescence was observed which is not comparable with those we reported previously. This indicates that the protein batch was damaged/contaminated. The experiments will be repeated with a new batch that will be characterized by intrinsic fluorescence quenching and EPR beforehand.



Scheme 16: Oxidation reaction of ABTS by H_2O_2.



Scheme 17: Oxidation of TMB by H2O2.

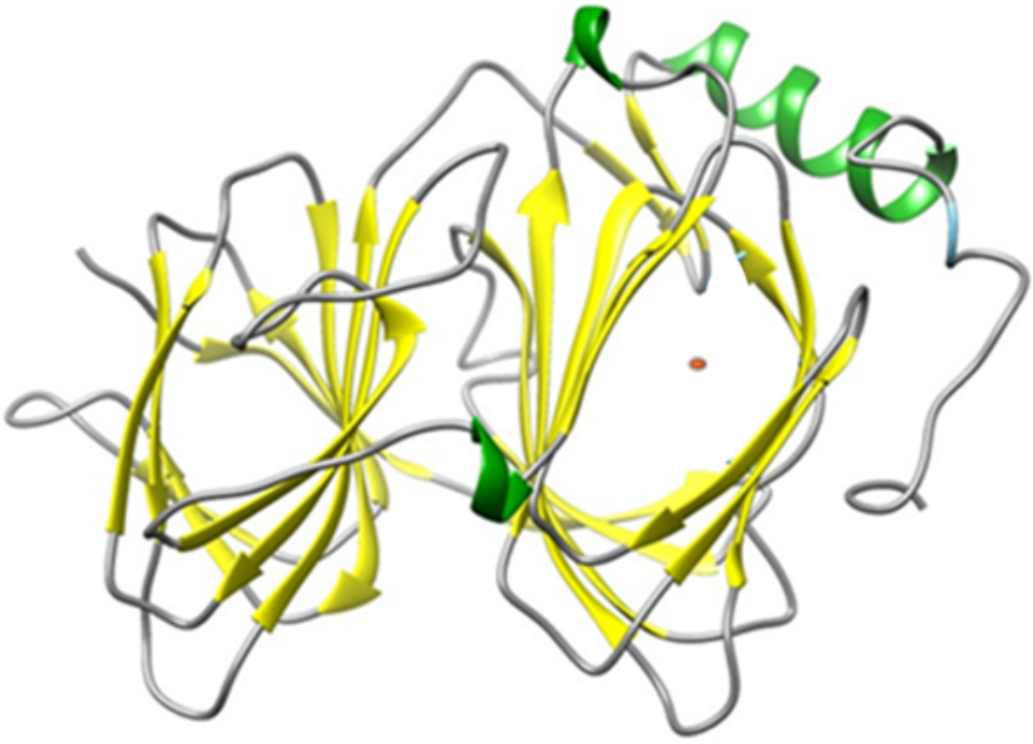
2.4. CONCLUSION:

PMIs are metalloenzymes that catalyze the reversible isomerization between M6P and F6P. PMIs belong to the cupin superfamily of enzymes and their active site is located in the characteristic barrel of β-sheets. Artificial enzymes formed by incorporating transition metals or their complexes in β-sheets are rare and are poorly studied. Aiming at the preparation of a chimeric PMI-PLP protein, we prepared and characterized an artificial metalloprotein based on PMI. The enzyme expression vector yielded PMI comprising a His_6 tag that we were able to cleave by introducing a $TEV_{c/site}$ using molecular biology techniques. Indeed, the expression plasmid was digested and a site for specific hydrolysis by acTEV protease was successfully introduced flanked by the sequence of the enzyme and that of the His_6 tag. After purification on Ni-NTA column, the enzyme was separated from the tag by digestion with TEV protease. TYCHO showed that the thermal denaturation of PMI can occur at 60°C. PMI contains one zinc(II) cation *per* enzyme molecule and the removal of zinc(II) from the active site was performed by incubating the protein with 10^6 equiv. EDTA. To prepare the artificial metalloenzyme, copper(II) was added and the successful transformation of PMI into an artificial copper(II)-containing enzyme was characterized by a relatively high affinity constant (K_D = 3 µM). The EPR spectrum of PMI-Cu protein confirmed the presence of a single copper in the binding site of the enzyme. PMI-Cu protein complex did not catalyze Diels-Alder reaction. However, a peroxidase activity can be obtained by the artificial PMI-Cu enzyme after future works.

CHAPTER 3

PIRIN LIKE PROTEIN

PLP



CHAPTER 3: PIRIN LIKE PROTEIN

3.1. INTRODUCTION

Following the studies that we carried out on PMI and described in chapter 2, we move on in this chapter to the second starting enzyme i.e. PLP. PLP is a bicupin containing an Fe^{2+} in its active site. The characteristics of this protein will be described in details in section 3.2. Then, we will prepare and characterize a new artificial metalloprotein starting from the PLP of *Pseudomonas aeruginosa* by substitution of its native Fe^{2+} by Cu^{2+}.

In the following sections, the construction of the expression vector of PLP, that will be later used to express the chimeric PMI-PLP protein (Chimera 1) will be presented. The plasmid encoding for the His_6-PLP that we obtained was revealed to cause co-expression problems thus, first we describe how we edited this plasmid to express only PLP. Second, the insertion of $TEV_{c/site}$ flanked by the His_6-tag and the PLP sequences is shown. This site allows for the His_6-tag to be cleaved after purification and before metal substitution. After further purification on a Ni-NTA column followed by the His_6-Tag cleavage by TEV protease, the obtained protein was characterized by SDS PAGE electrophoresis and Western blotting. Its transformation into an artificial metalloprotein by exchanging the native Fe^{2+} by Cu^{2+} will then be presented including its characterization by various spectroscopic techniques.

3.2. STATE OF THE ART

In 1997, Winnacker *et al.* discovered a protein that they named it Pirin.[80] This non-heme iron protein was produced in every human tissue[80] and was highly conserved in prokaryotes, fungi, plants, and mammals.[81] According to biochemical studies, Pirin is a nuclear protein of 290 amino acids and its molecular weight is about 32 kDa. [80]
Southern and Western Blots have shown that the Pirin genes' expression occurred in dot-like subnuclear cells structures for certain nuclear processes, such as the replication of DNA and/or the transcription of RNA polymerase II.[82] It has been shown that Pirin can be implicated in apoptosis and in cellular stress in eukaryotic organisms and in seed germination.[14,83]

Studies have shown that Pirins and quercetinases have a lot in common structurally.[14,84,85] Both of them belong to the cupin superfamily of proteins.[14,83] However, their metal ion needs are very different.[14] It has been established that metal ions such as Cu^{2+} Fe^{2+}, and Ni^{2+} were required for quercetinase activity.[14,86]

In 2013, it has been demonstrated that Pirin played a redox-sensing role in NF-kB regulation.[82] Only the ferric form of Pirin significantly enhanced the binding of NF-kB proteins to target kB genes. Though the ferrous and ferric Pirin differ only by one electron, their conformations was revealed to be different. [82] In particular, an R-shaped area forms the Pirin-NF-kB binding interface, which is responsible for the modulation of the NF-kB DNA binding 's properties. This area possesses 2 different conformations based on the identity and formal redox state of the metal. The Fe center is thus demonstrated to have an allosteric role in the binding of NF-kB proteins to the Pirin protein.[82] Pirin is thus proposed to act as a reversible functional switch that allows NF-kB to adapt to changes in the redox level at the cell nucleus.[82]

Other studies have shown that Pirin-encoding orthologs were involved in a variety of biological activities, and the cellular function of this family of proteins is frequently linked to stress conditions.[87]

In 2018, Pirin was discovered overexpressed in surgical specimens from hormone-sensitive prostate cancer (HSPC). This protein was therefore considered to be a potential novel therapeutic target for castration resistant prostate cancer.[88]

In 2015, Kataoka *et al.* searched for a protein classified into Pirin from *Pseudomonas stutzeri zobelli* which is a marine isolate strain used in agriculture, waste-water treatment and bioremediation. They planned to study the catalysis of oxidation reactions including quercetinase activity which has never been studied before in bioremediation. They found a protein that they called Pirin Like Protein (PLP). They

cloned it, expressed it and determined its 3D structure by based of the crystal structure of Pirin (Figure 36).[14] The structure shows that PLP is indeed also a bicupine.

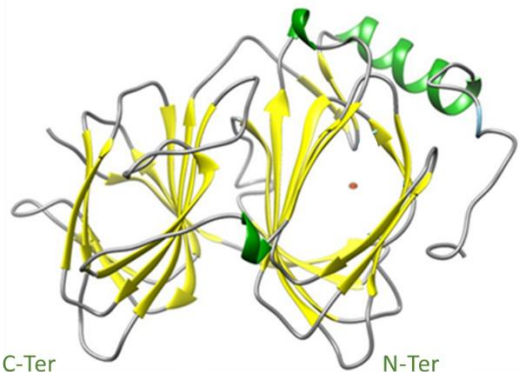


Figure 36: Swiss model of PLP showing its β barrel catalytic domain surrounded by an α-helixes domain on one side and a C-terminal β barrel domain on the other side.

PLP has a strong structural resemblance with Human Pirin protein as both proteins belong to the cupin superfamily which has two domains comprised of characteristic β-stranded motifs and intervening loops. They both have a high degree of amino acid sequence homology of 39% (Figure 37). Hence, it was suggested that the PLP contains one Fe^{2+} in the active site.[14] The structure of PLP was predicted accurately by analogy to human Pirin protein by using Swiss model as seen by the heatmap in Figure 38.

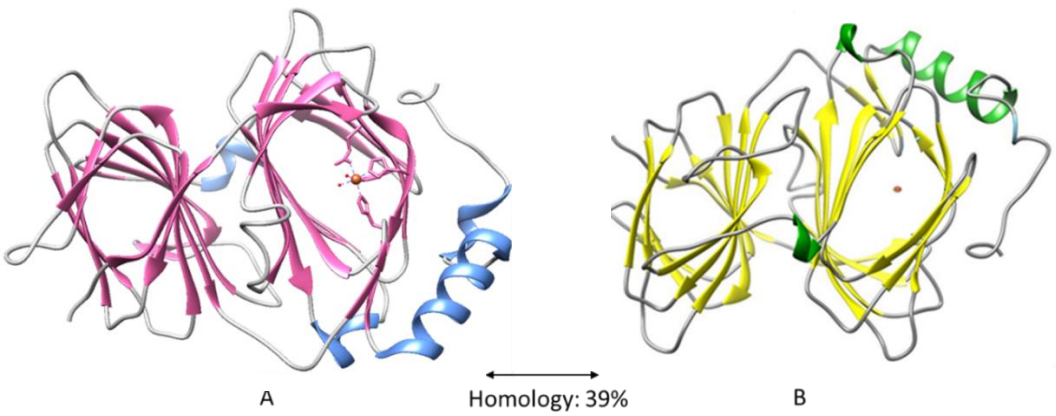


Figure 37: A) X-ray crystal structure of the human Pirin protein (pdb: 1J1L). B) Swiss model of the PLP metalloprotein.

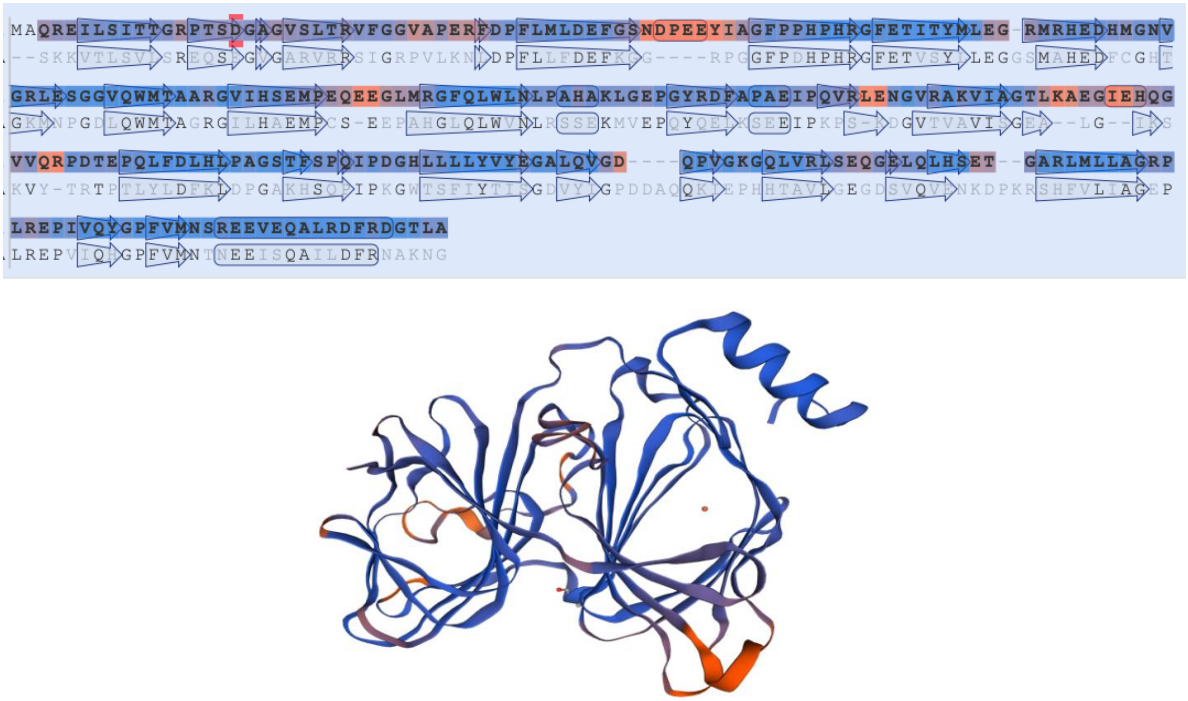


Figure 38: Heatmap of the goodness of the swiss model at every amino acid of the sequence of human Pirin protein. Blue = predicted with high accuracy; red = predicted with low accuracy.

PLP is formed of two domains each composed of a barrel of β-sheets which are connected by a single inter-domain composed of α-helixes and contains one cation of iron in its active site. PLP's active site lies in the β-sheet barrel domain located on the N-terminal side (Figure 39).[14]

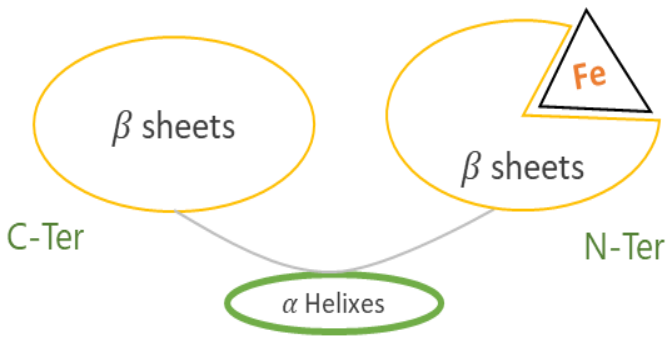


Figure 39: Representation of the structure of PLP showing the two conserved barrels of β-sheets linked by a single α-helixes inter-domain and containing an iron atom in the active site located in the barrel at the N-terminal side.

The biological role of PLP remains unclear as it is involved in a variety of biological processes.[87]

In 2018, Tala *et al* reported that the Pirin Like gene (PirA) was a unique redox-sensitive negative modulator of very long-chain acyl-CoA dehydrogenase, which catalyzes the beta-oxidation pathway in its first committed step.[87]

Furthermore, recent studies have shown that this protein exhibited quercetinase activities upon the incorporation of divalent metal cations such as Cu^{2+} and Zn^{2+} instead of iron.[14] In the case of Cu^{2+}, the holo-protein demonstrated the highest quercetinase activity as well as spectroscopic properties typical of type II Cu protein.[14]

Circular dichroism, EPR and absorption spectra of the Cu-PLP enzyme suggested that the coordination sphere contained His 59, His 61, His 103 and Glu 105 residues located in the consensus sequences of the cupin superfamily.[14]

In the present project, we have chosen this bicupin because its structure resembles that of the PMI. Indeed, the comparison of the proteins' structures has revealed a high structural resemblance between PMI and PLP: 45% of structure's identity and 68% of structure's similarity (Figure 40).

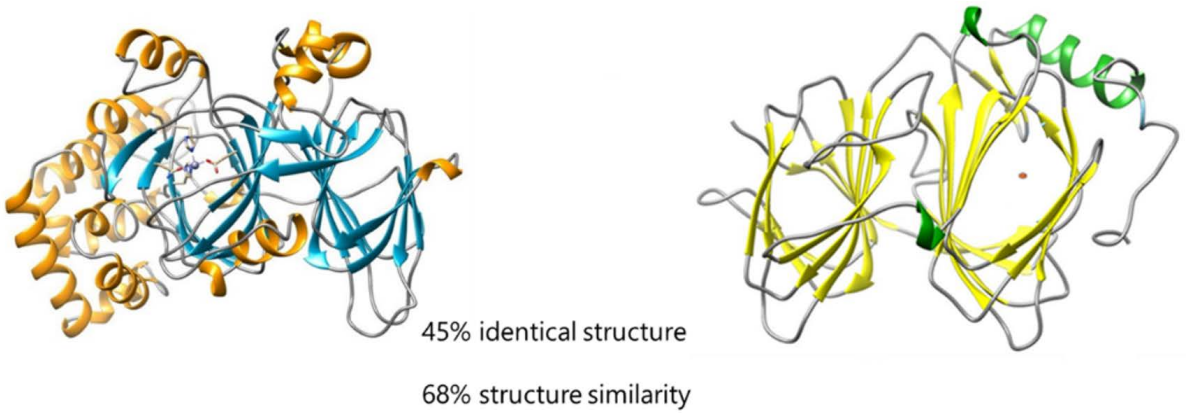


Figure 40: Structure's resemblance of PMI and PLP. Left: Crystalline structure of PMI; Right: Swiss model of the PLP.

3.3. RESULTS AND DISCUSSION

3.3.1. Editing of the expression vector for the PLP:

The pUC18 vector which provided ampicillin resistance and the expression of His_6-PLP was kindly provided to us by Professor Kataoka's team in Kanazawa University, Japan.[14]

The plasmid was used to transform E. *coli* BL21(DE3) strain. Transformed bacteria were selected on ampicillin containing agar plates. In cultured colonies, the protein expression was induced by 0.2 mM IPTG as described in literature[14] and the expression of the His_6-PLP was checked prior to purification. Anti His_6-tag antibody western blot was performed on the bacterial pellets. Figure 41 shows the presence of a double band on the western blot suggesting the expression of two His tagged proteins of similar MW. It was impossible to separate these two proteins and to proceed then with the purification of His_6- PLP on a Ni-NTA affinity column.

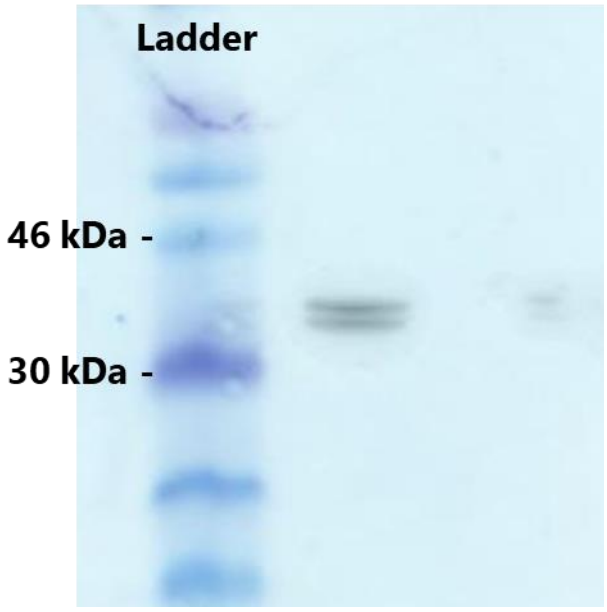


Figure 41 : Western Blot image of the coexpression of the His_6-PLP protein and the N-terminal 10 amino acids of LacZ protein fused His_6-PLP protein.

These results incited us to verify the DNA sequence of the provided plasmid by sequencing. The DNA sequence showed the presence of two Shine Dalgarno sequences and two start codons (Figure 42). This can result in the expression of the His_6-PLP protein as well as a His_6-PLP protein fused to a portion of LacZa (His_6-LcZ-PLP). This result can explain the presence of a double band on the western blot (Figure

41): one band refers to the His_6 PLP protein while the other band refers to the His_6-LcZ-PLP protein.

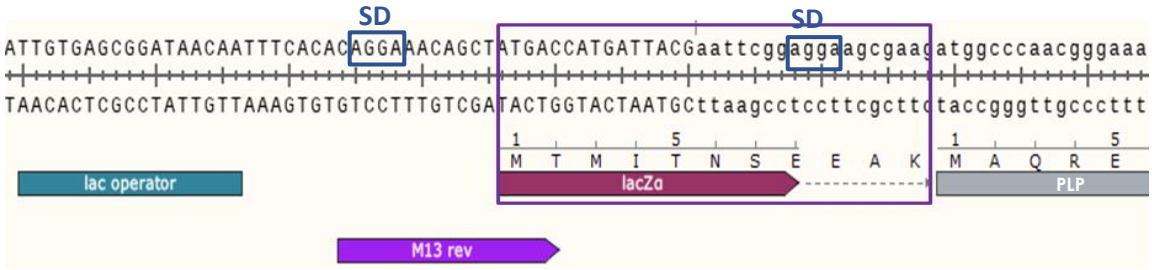


Figure 42: The vector linearized plasmid encoding for PLP showing two Shine Dalgarno (SD) sequences. LacZa 10 amino acids were fused to the 33 kDa PLP protein.

Therefore, to avoid the expression of LacZa fused His_6-PLP protein, we deleted the remaining portion of LacZa by PCR. The deletion was performed by designing forward and reverse primers that flank both sides of the region to be deleted (Figure 43).

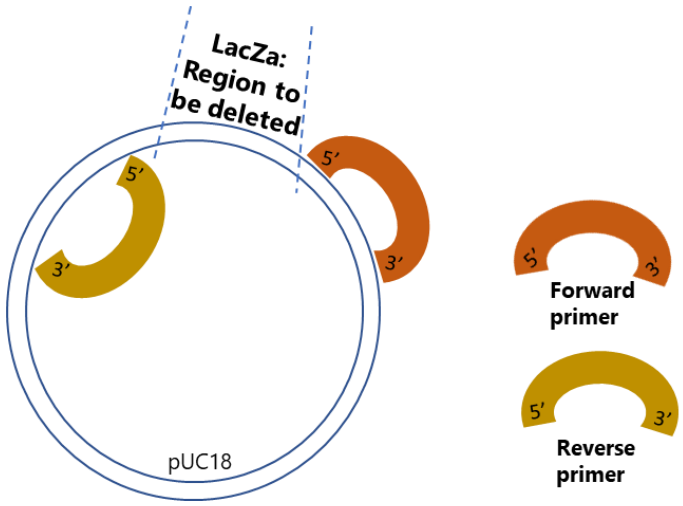


Figure 43: Primer design for deletion of a portion of LacZa. Forward and reverse primers flank the region to be deleted.

The PCR product was then purified and mixed with **K**inase, **L**igase and **D**pn1 (KLD). The KLD mix allows efficient phosphorylation, ligation and template removal in a single time. The mixture was then used to transform E.*coli* DH5a competent bacteria. Transformation resulted in bacteria that contained the desired plasmid. Then, a colony was grown, and plasmids were extracted and purified. DNA sequencing of the plasmid

allowed the verification of the sequence encoding for the His_6-PLP. The expected sequence was found (Figure 44).

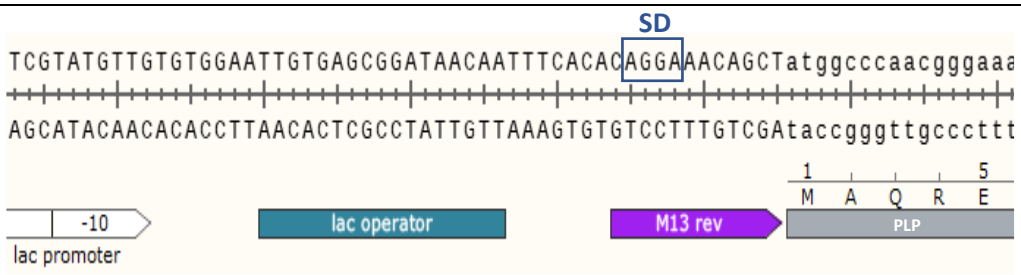


Figure 44: The vector linearized plasmid of PLP protein after deletion of a portion of LacZa to avoid the co-expression.

This deletion let to obtaining the desired recombinant protein that showed by western blot a single band at about 33 kDa corresponding to the expected MW of His_6-PLP (Figure 45).

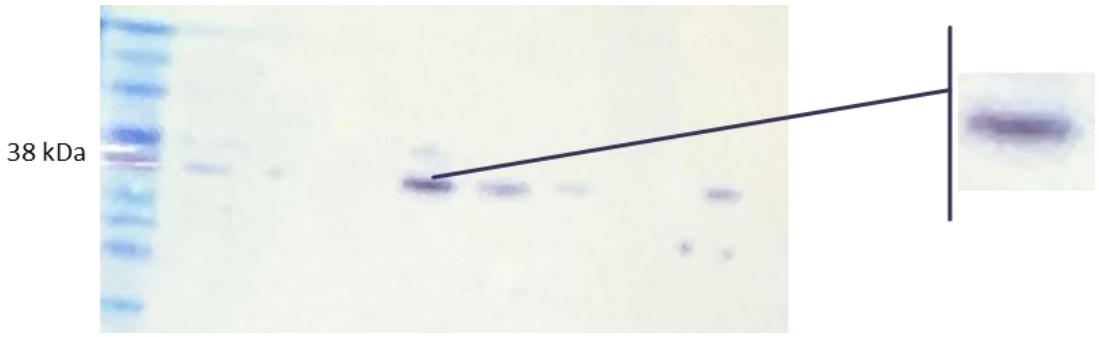


Figure 45: Westen Blot image of His_6-PLP after the deletion using Anti-His antibody.

The $TEV_{c/site}$ (5' GAAAACCTGTATTTTCAGGGC 3') was then inserted by PCR. The $TEV_{c/site}$ was flanked by the His_6-tag and the PLP sequence so that after purification, the tag may be cleaved before metal substitution (Figure 46).

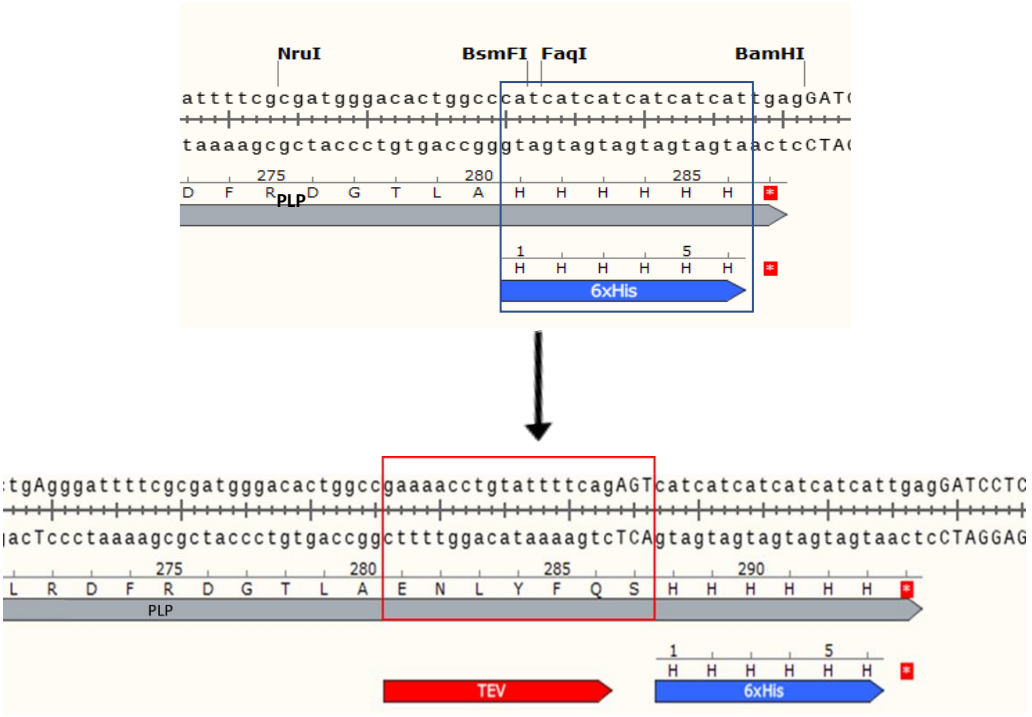


Figure 46: Insertion of the DNA encoding for the TEV protease cleavage site flanked by the His tag and the PLP linearized sequence.

First, we verified the presence or absence of DNA insert in plasmid constructs by colony PCR. The presence of the PCR amplicons and the size of the products were determined by electrophoresis alongside a DNA size marker on a precast 1.2% agarose gel followed by a western blot analysis. Figure 47 confirms the presence of PCR amplicons for colonies 1, 3 and 4, which proves that the $TEV_{c/site}$ was well inserted.

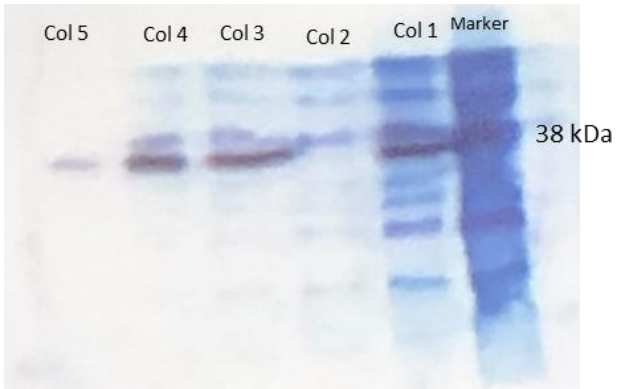


Figure 47: Western Blot image confirming the presence of PCR amplicons. Primary antibody: Anti-His_6tag antibody; Secondary antibody: Anti-mouse (H+L) antibody labelled with peroxidase.

Then, in order to confirm the presence of the DNA sequence encoding for the His_6-PLP, plasmids were extracted, purified and sequencing was performed. Sequencing showed the expected sequences by confirming mutations at the desired locations, which was enough to confirm the PCR results. To obtain large quantities of the protein, the modified plasmid was used to transform *E. coli* BL21(DE3) cells.

3.3.2. Expression and purification of the His_6-PLP

E.*coli* BL21 cells transformed with the modified plasmid encoding for the His_6-PLP-$TEV_{c/site}$ protein were cultured in LB medium supplemented with 75 µg/ml ampicillin at 37°C overnight with shaking (200 rpm). The production of the recombinant protein was induced with 0.2 mM IPTG as described in the literature.[14] Pellets of bacteria were resuspended in presence of lysozyme and benzonase. The mixture was then sonicated and centrifuged to eliminate all solids. The clear solution containing the His_6-PLP was purified on Ni-NTA agarose column. SDS-PAGE followed by Coomassie blue staining (see below) revealed a pure protein with a molecular weight of about 33 kDa (Lanes 6,7 and 8).

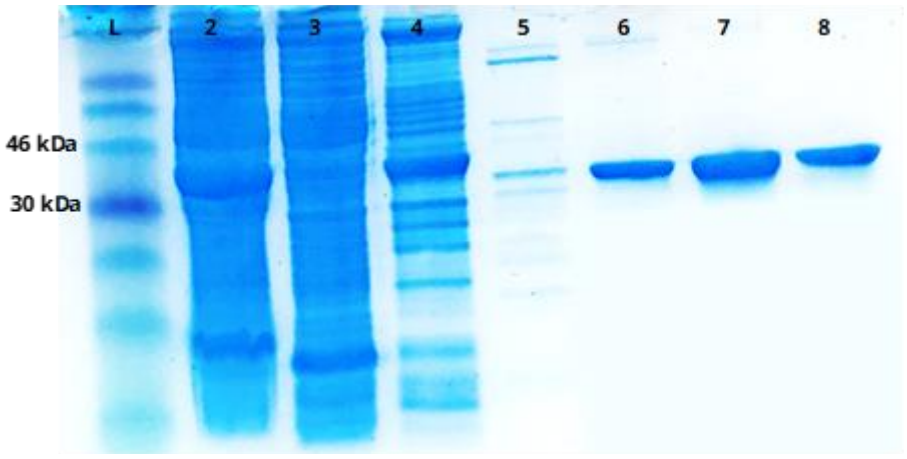


Figure 48: SDS-PAGE electrophoresis followed by Coomassie blue staining of the fractions collected after the purification of His_6-PLP on an Ni-NTA agarose column. Lane 2: supernatant; Lane 3: Flow through; Lane 4: washing fraction; Lanes 5,6,7 and 8 show bands corresponding to the expected molecular weight of pure His_6-PLP of about 33 kDa; L: Ladder

After the SDS-PAGE analysis, western blot was done and the protein bands were electro transferred horizontally onto a nitrocellulose membrane. The protein was detected using a specific primary antibody (Anti-His_6-tag) followed by a secondary antibody labelled with peroxidase. Western blotting confirmed that the purified starting protein

was indeed His-tagged (Figure 49). The eluted protein was dialyzed against a 50 mM MOPS buffer (50 mM MOPS and 300 mM NaCl), pH=7,6.

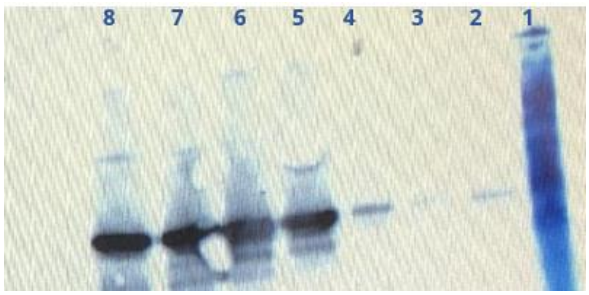


Figure 49: Western Blot image of His_6-PLP. Lanes: 1- Protein Ladder; 2- Flow through; 3- Washed fraction; 5-6-7-8- Eluted fractions.

3.3.3. Purification of His tag free PLP

To cleave the His_6-tag, the purified His_6-PLP protein fractions were incubated with acTEV protease. Following the digestion, the TEV protease with its polyhistidine tag was easily removed from the cleavage reaction by purification on a Ni-NTA affinity column. As confirmed by the western Blot below, pure tag free PLP could not bind on the Ni-NTA affinity column and was recovered in the flow through. The SDS gel revealed a pure protein with a molecular weight of about 30 kDa corresponding to the molecular weight of PLP tag free (Figure 50).

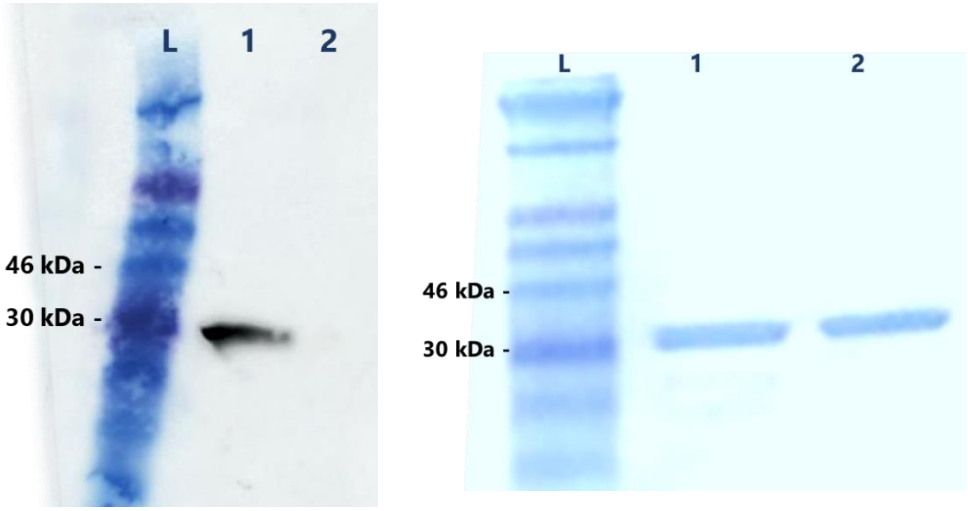


Figure 50: Right, Coomasie blue stained SDS-PAGE electrophoresis of PLP upon cleavage of the His_6 tag (on the right) followed by Western Blot image using Anti His primary antibody (on the left). L: Ladder; Lane 1: His_6-PLP + TEV; Lane 2: Tag free PLP

3.3.4. Transformation of PLP into an artificial metalloprotein:

3.3.4.1. Iron (II) removal

PLP was transformed into an artificial metalloprotein by exchanging its iron(II) cation with a copper(II) cation. For this, the native iron cation was removed by EDTA and copper nitrate was then added. It was reported that the iron(II) cation of Pirin could be removed from the active site by incubation with 10 equiv. of EDTA overnight.[82]

To completely strip the Fe^{2+} of PLP, we investigated the effect of the addition of various amounts of EDTA at 4°C in 50 mM MOPS containing 300 mM NaCl, pH=7.6. A significant increase of the fluorescence of the PLP tag free was observed upon increasing the EDTA concentration indicating that removal of iron causes a change in the protein folding which leads to the increase of the fluorescence.

Low increase in fluorescence was observed with 10 equiv. of EDTA indicating that 10 equiv. of EDTA were not enough to remove the native $Fe^{2+}$of PLP. Furthermore, adding Cu^{2+} to the PLP tag free incubated with 10 equiv. of EDTA did not affect the intensity of the fluorescence (Figure 51). This suggests that Cu^{2+} cannot replace the native Fe^{2+}.

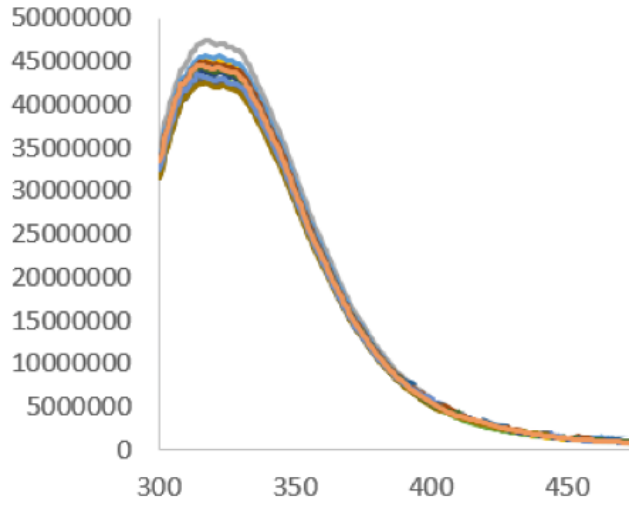


Figure 51: Fluorescence titration of the binding of Cu(II) to PLP with 10 equiv. EDTA upon excitation at 280 nm. No significant quenching of fluorescence was observed suggesting the remaining of iron cation in the active site of PLP.

A significant precipitation appeared upon the incubation with a concentration higher than 10^4 equiv. of EDTA. This precipitation suggested the denaturation of PLP that might be due either to the effect of EDTA on the PLP or to the stripped amount of Fe^{2+} from the PLP's active site. The denaturation of the PLP was verified by TYCHO which

verifies the structural integrity of the protein and determines the inflection temperature. Figure 52 showed that the native PLP folding is maintained at increasing temperature and its thermal denaturation can occur only at 75°C. Upon the addition of 10^4 equiv. of EDTA, PLP's signal changed, suggesting the change in its folding.

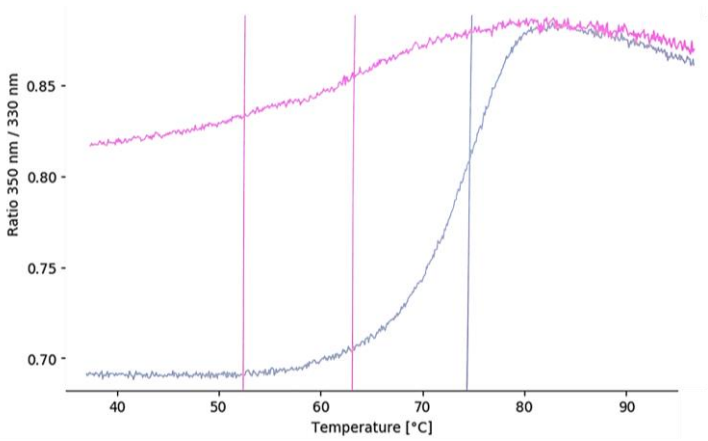


Figure 52: TYCHO measurements of 10 µM native PLP-Fe in black and 10 µM PLP iron free in pink.

Upon the addition of 10^3 equiv. of EDTA, the PLP retains its structure with little changes in the TYCHO signal and its thermal denaturation can occur at 75°C (Figure 53). Similarly, quenching of the fluorescence was also observe (see below 3.3.4.2). Hence, to achieve the complete removal of iron(II) we discovered that 10^3 equiv. of EDTA were required followed by dialysis for 2 days.

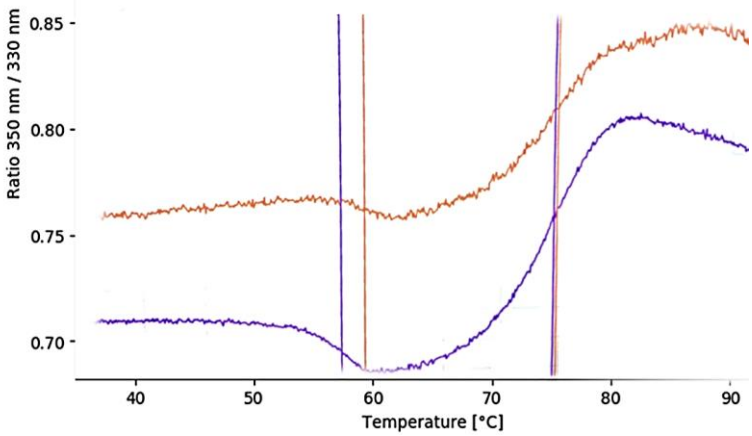


Figure 53: TYCHO measurements of 10 µM native PLP in orange and 10 µM PLP iron free upon addition of 10 mM EDTA in purple.

3.3.4.2. Evolution of fluorescence in the presence of metals:

To study the changes in the folding of the PLP, we investigated the interaction of the protein with different metal cations by fluorescence spectroscopy.

Since the His_6-tag is an excellent ligand for metal cations,[89] the added metals bind first the His_6-tag and then can bind to other amino acids. For this, the tag should be cleaved before the fluorescence titration for trustworthy results.

We studied the fluorescence titration of the PLP tag free by different metals before removing the native iron cation by EDTA. Figure 54 shows a non-significant variation of the fluorescence following the addition of both Copper sulfate and Zinc sulfate. This suggested that Copper and Zinc could not bind to the protein, which confirmed the presence of Fe^{2+} in the active site of the PLP preventing the binding of both metals.

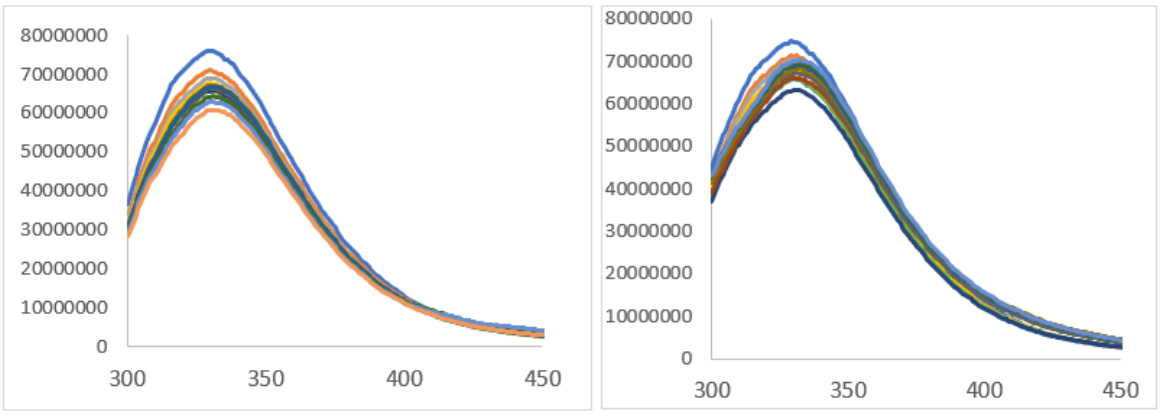


Figure 54: Titration of the fluorescence of native PLP upon the addition of increasing amounts (0.2-55 equiv.) of Cu(II) (left) and Zn(II) (right) (excitation at 280 nm). A non-significant quenching of the signal was seen in both cases. Measurements were performed in 50 mM MOPS, 300mM NaCl, pH = 7,6 using 2.5 µM PLP tag free.

As for Figure 55, a significant fluctuation of the fluorescence could be observed upon the addition of increasing amounts of nickel nitrate. This could be due to non-specific binding of nickel to PLP with low affinity.

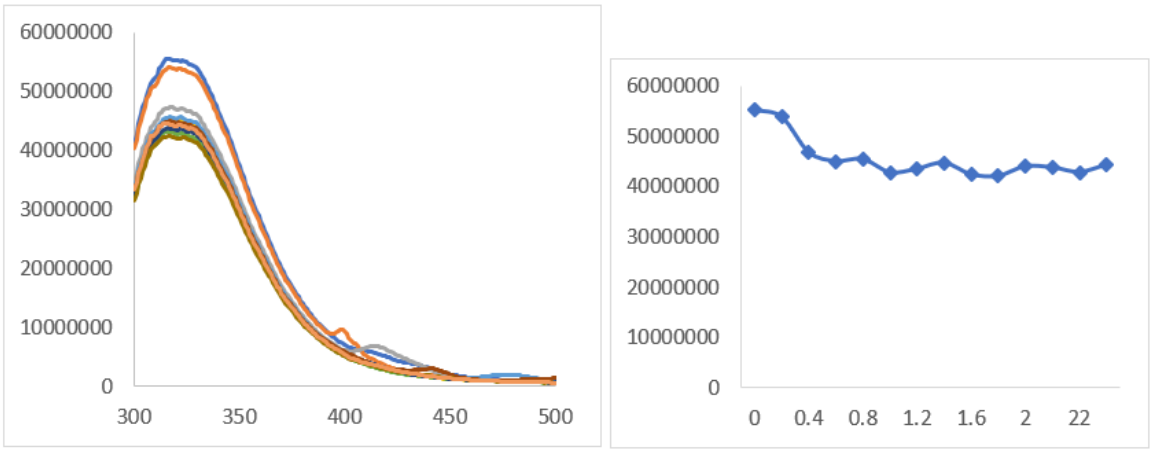


Figure 55: Left: Titration of the fluorescence of native PLP upon the addition of increasing amounts of Ni^{2+}(excitation at 280 nm). Right: Fluorescence of PLP at 320 nm as function of increasing equivalents of added Ni^{2+}. Measurements were performed in 50 mM MOPS, 300mM NaCl, pH = 7,6 using 2.5 µM PLP.

Since PLP is suspected to contain Fe^{2+} in its active site, it was necessary to remove this metal cation before transforming PLP into an artificial metalloenzyme. For this, Fe^{2+} was removed by incubating PLP with 1000 equiv. of EDTA for 8 hours at 4°C. Then, EDTA was removed by dialysis in 50 mM MOPS, 300 mM NaCl, pH=7.6 for two days.

The iron free His_6-tag free protein could then be reconstructed by binding copper(II). Upon addition of increasing amounts of Cu(II) (0.2-55 equivalents) to the iron free PLP, the intensity of the protein intrinsic fluorescence signal decreased drastically (Figure 56).

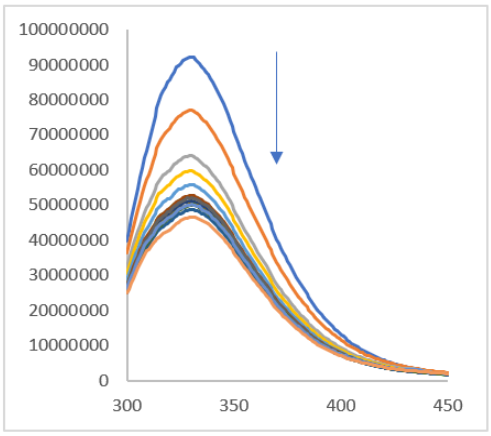


Figure 56: Quenching of the fluorescence of iron free PLP upon the addition of increasing amounts (0.2-55 equiv.) of Cu(II) (excitation at 280 nm). Measurements were performed in 50 mM MOPS, 300mM NaCl, pH = 7,6 using 2.5 µM PLP.

As for PMI, fluorescence values were plotted against the total added copper(II) concentration, and the resulting curve (Figure 57) was fitted by using the following equation, which allowed the determination of the K_D value for the PLP-Cu(II) complex as well as the Cu/PLP stoichiometry:

$$\Delta F = \Delta F_{max} \frac{(n[P]_0 + [L]_0 + K_D) - \sqrt{(n[P]_0 + [L]_0 + K_D)^2 - 4n[P]_0[L]_0}}{2n[P]_0}$$

Where: F is the intensity of the fluorescence signal at 340 nm, $[P]_0$ and $[L]_0$ are respectively total protein and total copper(II) concentrations, n is the stoichiometry (number of ligand-binding sites per protein); $\Delta F_{max} = F_0 - F_{max}$; ($F_0$ and F_{max} are the intensities resp. in absence of copper and when the protein is saturated with copper) represents the asymptotic value for ΔF. ΔF_{max}, K_D, and n were calculated from the fitting equation. A good fit was achieved with n=1 and yielded a K_D value of 0.5 µM. This indicates that one molecule of Cu(II) was bound per molecule of PLP to form a new metalloprotein of which the activity remains to be tested.

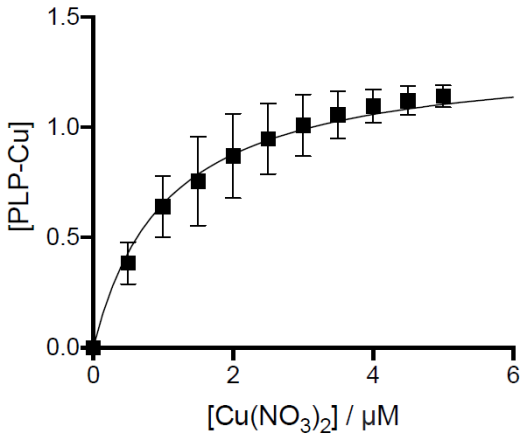


Figure 57: Curve showing the increase of Cu-PLP concentration, derived from the fluorescence values at 340 nm, as a function of the total concentration of added Cu(II).

The insertion of other metal ions such as Zn^{2+}(Figure 58) and Ni^{2+} (Figure 59) has also been studied by fluorescence spectroscopy. Similarly to the titration using copper(II), a quenching of the fluorescence of PLP was also observed.

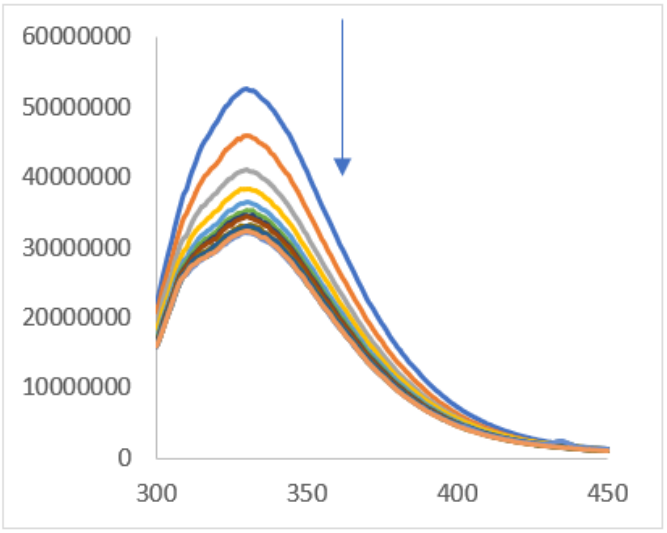


Figure 58: Quenching of the fluorescence of native iron free PLP upon the addition of increasing amounts (0.2-55 equiv.) of Zn(II) (excitation at 280 nm). Measurements were performed in 50 mM MOPS, 300mM NaCl, pH = 7,6 using 1.25 µM PLP.

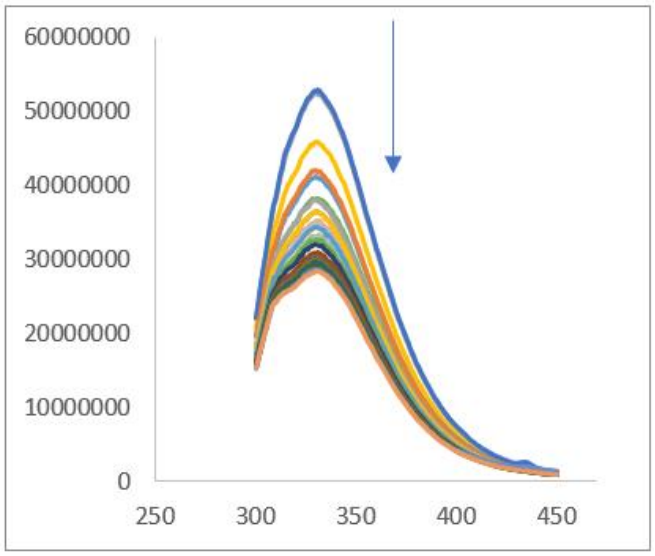


Figure 59: Quenching of the fluorescence of native iron free PLP upon the addition of increasing amounts (0.2-55 equiv.) of Ni(II) (excitation at 280 nm). Measurements were performed in 50 mM MOPS, 300mM NaCl, pH = 7,6 using 1.25 µM PLP.

To compare the affinity of each metal for PLP, we present in Figure 60 the variations of the fluorescence of PLP at 320 nm upon addition of increasing concentrations of the various metal cations. The lowest fluorescence quenching was observed in the case of Zn^{2+} which indicated that there was a weak interaction between Zn^{2+} and PLP (Figure 60). A slightly higher variation in fluorescence was observed in the case of Ni^{2+} which indicated that there was a better interaction between Ni^{2+} and PLP (Figure 60).

Nevertheless, we noticed the presence of two inflexion points that could be due to non-specific binding of nickel to the protein. Finally, the higher variation in fluorescence was observed in the case of Cu(II) which indicated that this metal cation had the best interaction with PLP (Figure 60).

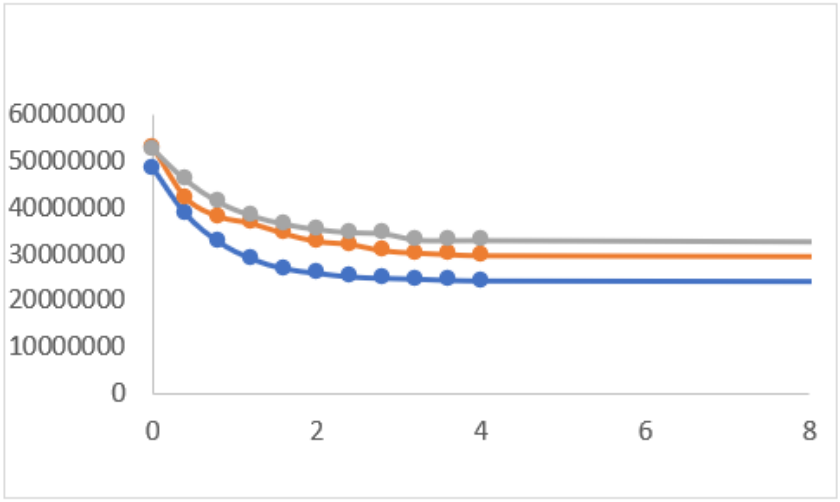


Figure 60: Fluorescence of PLP at 320 nm as function of increasing equivalents of metal cations added. PLP-Cu(II) complex in blue, PLP-Ni(II) complex in orange and PLP-Zn(II) complex in gray.

Similarly to the PLP-Cu complex, the PLP-Ni concentration was plotted against the total added Nickel concentration and the resulting curve (Figure 61) was fitted by using the same equation as above, which allowed the determination of the K_D value for the PLP-Ni(II) complex. A good fit was achieved with n=1 and yielded K_D = 0.7 µM. However, a better fit was obtained with n= 2 and yielded K_{D1} = 0.5 and K_{D2} = 2.1 µM. This indicates that the PLP can possess more than one binding site for Ni^{2+}.

The titration thus confirmed the highest affinity of PLP for copper. Furthermore, it confirmed the formation of 1/1 copper/protein complex characterized by a K_D of 0,5 µM.

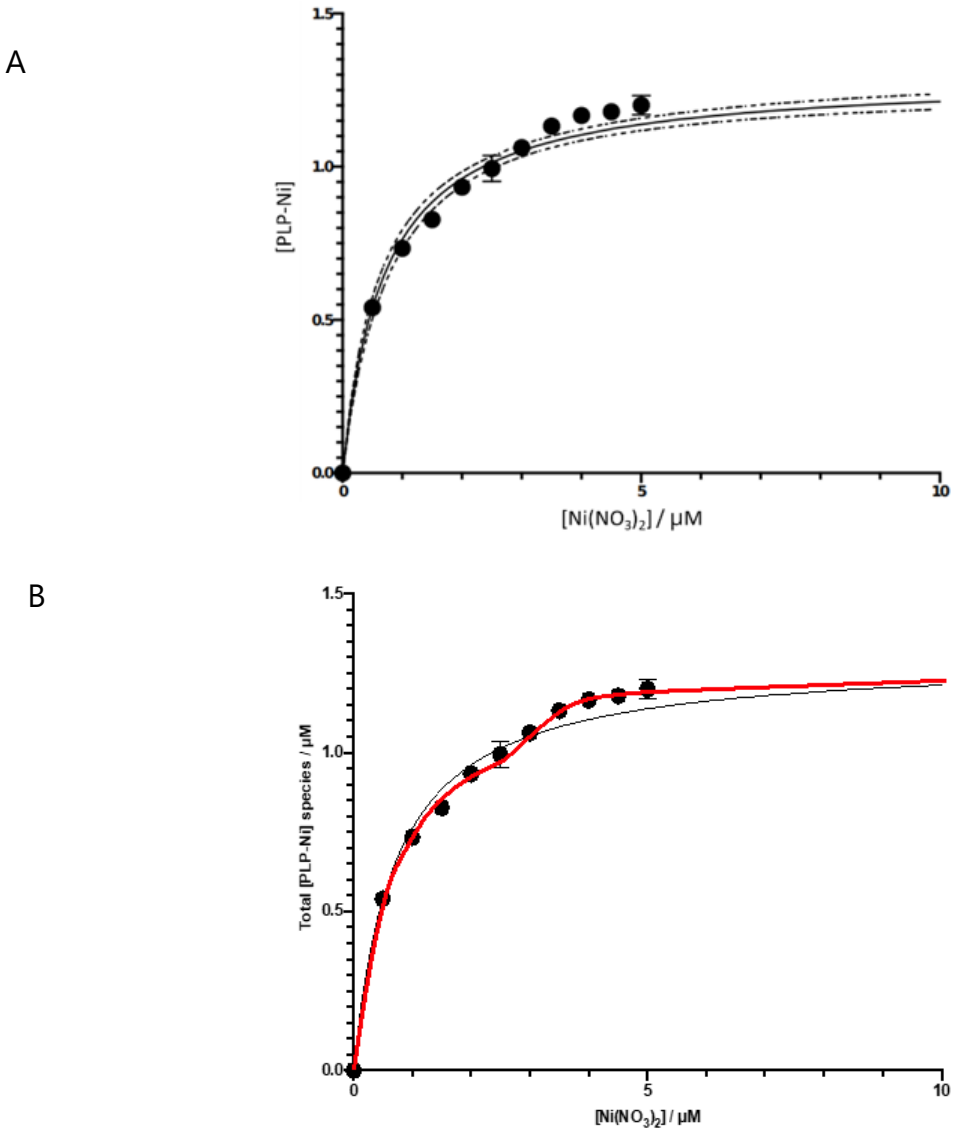


Figure 61: A) concentration of PLP-Ni (derived from the fluorescence values at 340 nm) plotted against the total added Ni(II) concentration. The fit was achieved with n=1. B) total concentration of different PLP-Ni species (derived from the fluorescence values at 340 nm) plotted against the total added Ni(II) concentration. The fit was achieved with n=2.

Since the PMI showed an increase in fluorescence following the addition of copper, unlike the PLP which showed a decrease, it was necessary to find an alternative method to study the copper affinity for the protein that could be further used to characterize the chimera. Thus, another technique, isothermal titration calorimetry ITC, was used to confirm the fluorescence results as described in the paragraph below.

3.3.4.3. Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) is one of the best methods used to determine the thermodynamic parameters of interactions in solution.[90] It is most often used to study the binding of small molecules (such as medicinal compounds) to larger macromolecules (proteins, DNA etc.). It is based on the concept that heat can be either absorbed or generated during the process. At constant temperature, a solution containing a ligand is added to a cell containing a solution of the macromolecule. The interaction between the two partners then results in heat release or absorption.[90]

The binding of copper(II), zinc(II) and Ni(II) to metal-free PLP was first studied without adding EDTA. As shown in Figure 62, Figure 63 and Figure 64, when metal cations were added into the native PLP, no heat exchange was found during these three experiments. As for fluorescence titration, this result proved that there is no interaction between PLP and any metal before stripping the iron from its active site.

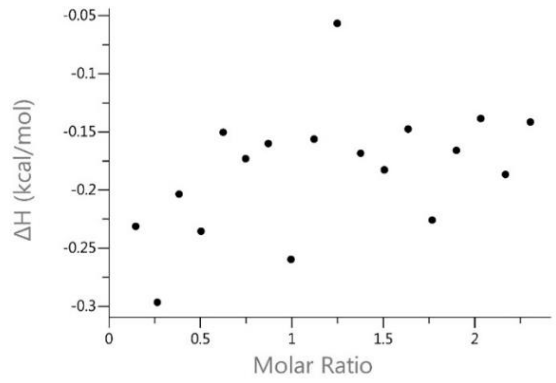


Figure 62: Microcalorimetric titration of PLP by copper sulfate, heat flow against Molar Ratio. Successive injections at 20°C, of 2 µl of a 600 µM solution of Copper sulfate in 50mM MOPS, 300mM NaCl were made on 50 µM PLP in 50mM MOPS, 300mM NaCl .

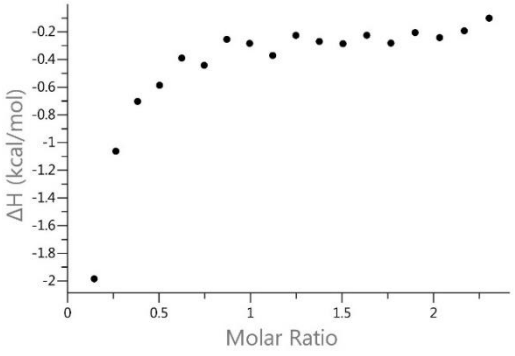


Figure 63: Microcalorimetric titration of PLP by nickel sulfate, heat flow against Molar ratio. Successive injections at 20°C, of 2 µl of a 600 µM solution of nickel sulfate in 50mM MOPS, 300mM NaCl were made on 50 µM PLP in 50mM MOPS, 300mM NaCl.

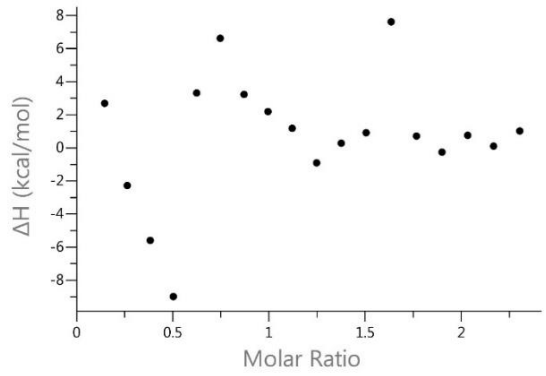


Figure 64: Microcalorimetric titration of PLP by zinc sulfate, heat flow against Molar ratio. Successive injections at 20°C, of 2 µl of a 600 µM solution of zinc sulfate in 50mM MOPS, 300mM NaCl were made on 50 µM PLP, 50mM MOPS, 300mM NaCl.

To completely remove the Fe(II) cation of the PLP, 10^3 equiv. EDTA were added. As shown in Figure 65, no heat exchange was observed when Zinc sulfate was added into the iron free PLP. This result shows that there is no interaction between PLP iron free and Zn(II) which corroborates the results obtained by fluorescence titration.

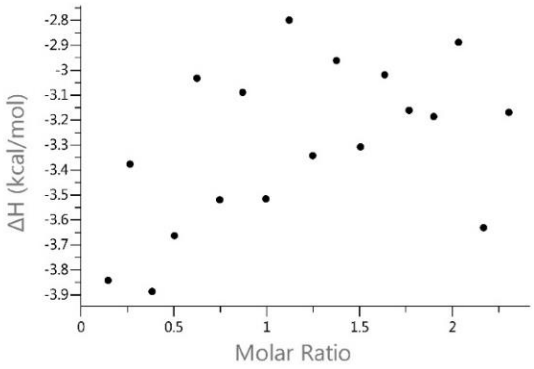


Figure 65: Microcalorimetric titration of iron free PLP by zinc sulfate, heat flow versus Molar Ratio. Successive injections at 20°C, of 2 µl of a 300 µM solution of zinc sulfate in 50mM MOPS, 300mM NaCl were made on 24 µM PLP, 50mM MOPS, 300mM NaCl.

However, when copper(II) was injected into a solution of iron free PLP, no exothermic reaction was observed. As for Figure 66, it is not possible to reliably determine any thermodynamic parameter. This may be due to the fact that this method was not suitable to determine the affinity of copper into PLP or the protein could bind to the alloy of the ITC cell preventing the binding of the added metal cations.

Furthermore, we can suggest that only a part of the protein can bind the metal which can be due to the EDTA effect of the PLP's folding.

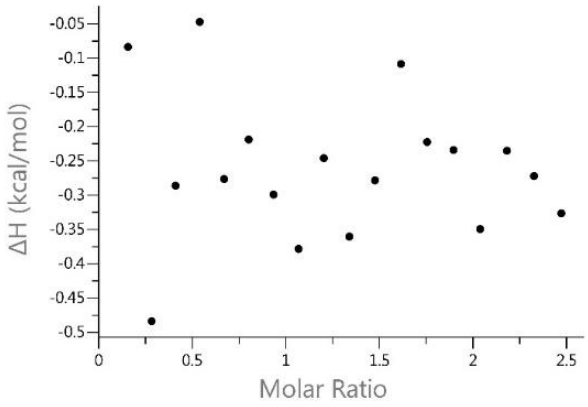


Figure 66: Microcalorimetric titration of iron free PLP by copper sulfate, heat flow against time. Successive injections at 20°C, of 2 µl of a 300 µM solution of copper sulfate in 50 mM MOPS and 300 mM NaCl were made on 24 µM PLP in 50 mM MOPS and 300 mM NaCl.

However, when nickel(II) was injected into a solution of iron free PLP, an exothermic reaction was observed after the first few injections. Figure 67 shows a typical example of the curve that was obtained after integration of heat signal. However, for the PLP-Ni complex, a 0,3 Ni(II)/PLP stoichiometry could be determined which could be the sign of the presence of multiple binding sites for Nickel in PLP (Figure 67) or denatured protein. These results confirm the existence of interactions between Ni^{2+} and PLP. Consequently, the K_d PLP-Ni were determined from ITC measurements and were equal to 0,6 µM.

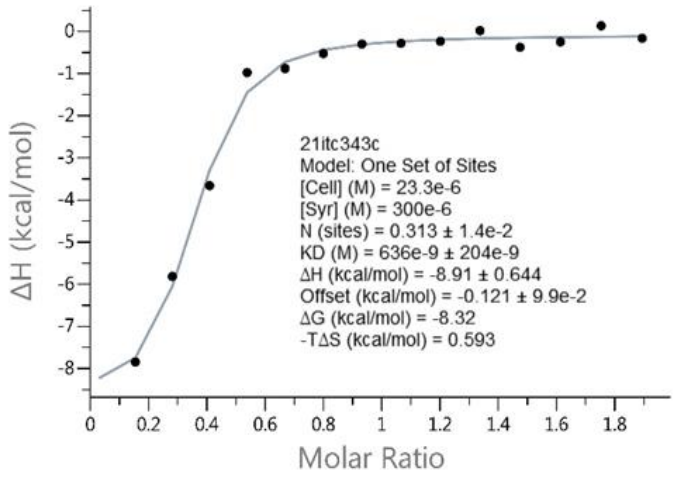


Figure 67: Microcalorimetric titration of iron free PLP by Nickel nitrate, heat flow against molar ratio. Successive injections at 20°C, of 2 µl of a 300 µM solution of nickel nitrate were made on 24 µM PLP in 50 mM MOPS and 300 mM NaCl.

After ITC experiments, the samples coming out of the titrations were analyzed by TYCHO to check the integrity of the protein fold. Figure 68 shows a little difference in the signal. This confirms that the protein retains its folding upon the EDTA addition followed by the addition of different metals. The thermal denaturation of these proteins occurs around 75°C.

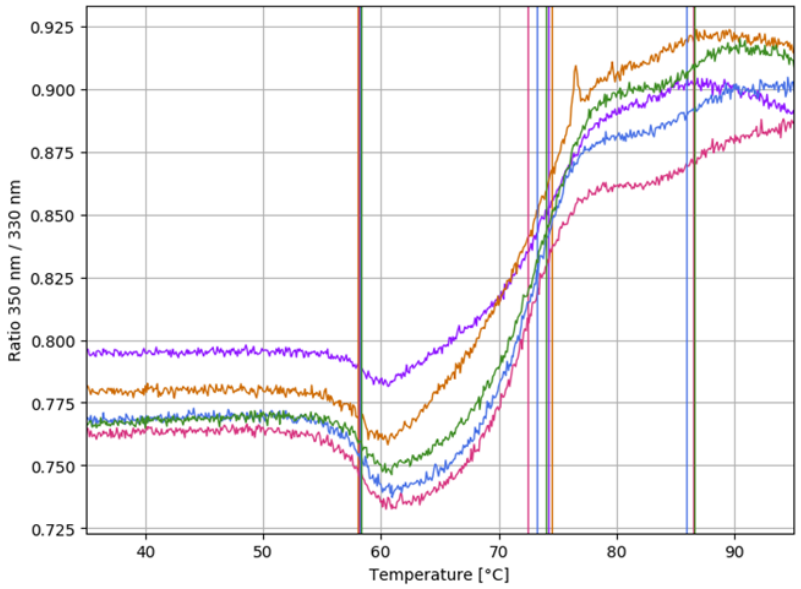


Figure 68: TYCHO measurements of PLP-metal complexes upon addition of 10^3 equiv. of EDTA. Orange: PLP-$CuNO_3$; Purple: PLP; Pink: PLP-$NiNO_3$; Blue: PLP-$ZnSO_4$; Green: PLP-$CuSO_4$.

However, the different ITC results were not reproducible when the PLP remained longer time without any metal and the stoichiometry was changing. This confirmed that the metal of PLP conferred a structural stability to the protein.

Hence, based on the last ITC results:

- ITC is not a suitable method to analyze the binding of Cu^{2+} in PLP.
- The ITC cells are made of metal alloys thus PLP can bind to the cell preventing the binding of the added metal cations.
- The folding problems of PLP can be due to two reasons:
  - i) EDTA's can affect the protein's folding when added in high concentration. It is noteworthy that this chelator is known to have high affinity for binding metalloproteins, destabilizing their structure and forming aggregates in solution.[91,92]
  - ii) The absence of the iron cation presented in the PLP's active center conferring a structural role which can alter the PLP's folding.

3.3.4.4. Electron paramagnetic resonance EPR:

The EPR spectrum of the metal free PLP and PLP-Cu complex were measured at 10 K at a frequency of 9.3 GHz. As expected, the iron free PLP did not show any EPR signal, which confirmed that the active site was free of iron(III) after incubation with EDTA. However, the EPR spectrum of the PLP-Cu complex showed a quadruplet signal centered at g// = 2.06, g⊥= 2.32 values, hyperfine coupling constants of A// = 18.6 mT which confirmed the presence of a single type of copper(II) in the binding site (Figure 69).

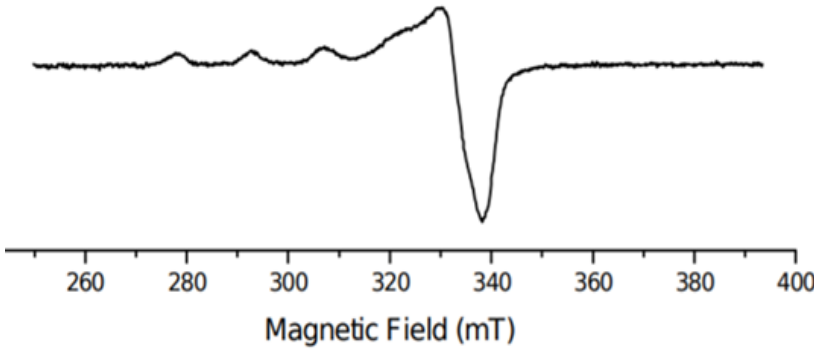


Figure 69: EPR spectrum of PLP-Cu protein measured at 10 K at a 9.3 GHz frequency; 80 µM PLP-Cu in 50mM MOPS buffer, 300 mM NaCl, 15% Glycerol.

3.3.4.5. Effect of high concentrations of copper on the PLP:

To investigate the effect of the addition of copper to iron free PLP, we studied the protein's structural folding of the protein as a function of the increasing amounts of copper sulfate by TYCHO.

Upon the addition of 2 equiv. of Cu(II), the PLP retains its folding (Figure 70). However, upon the addition of 10 equiv. of Cu(II), the folding profile of PLP-Cu protein complex changed (Figure 71). This result indicates that high concentrations of copper can alter PLP's folding.

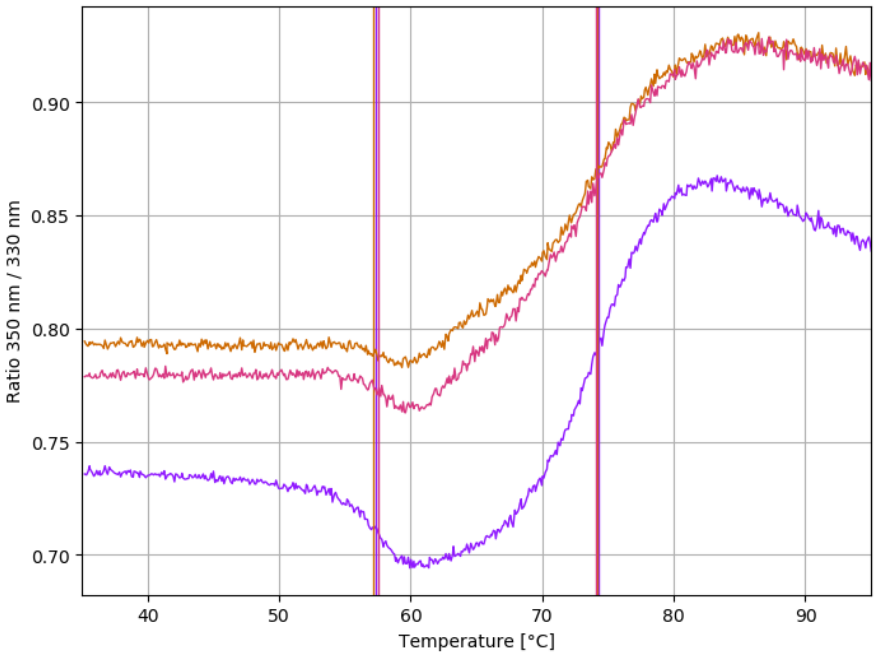


Figure 70: TYCHO measurements of PLP before and after the copper addition. Purple: 10 µM PLP; orange: 30 µM iron free PLP (30 mM EDTA); Pink: 30 µM iron free PLP (30 mM EDTA) +2 equiv of Copper sulfate.

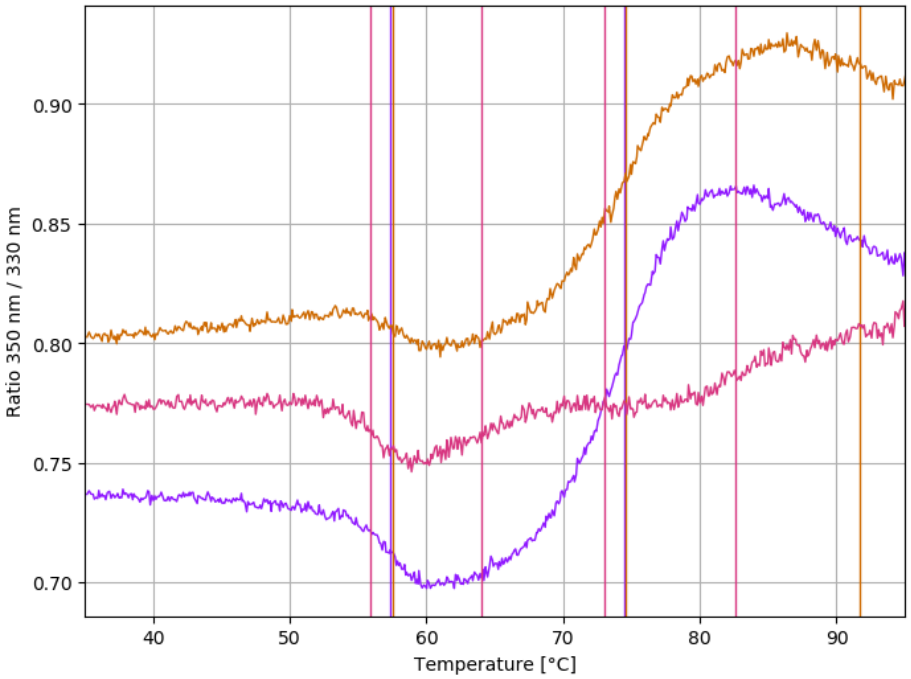


Figure 71: TYCHO measurements of PLP before and after the addition of 10 equiv. of copper. Purple: 10 µM PLP; orange: 30 µM iron free PLP with 30 mM EDTA; Pink: 30 µM iron free PLP with 30 mM EDTA +10 equiv. Copper.

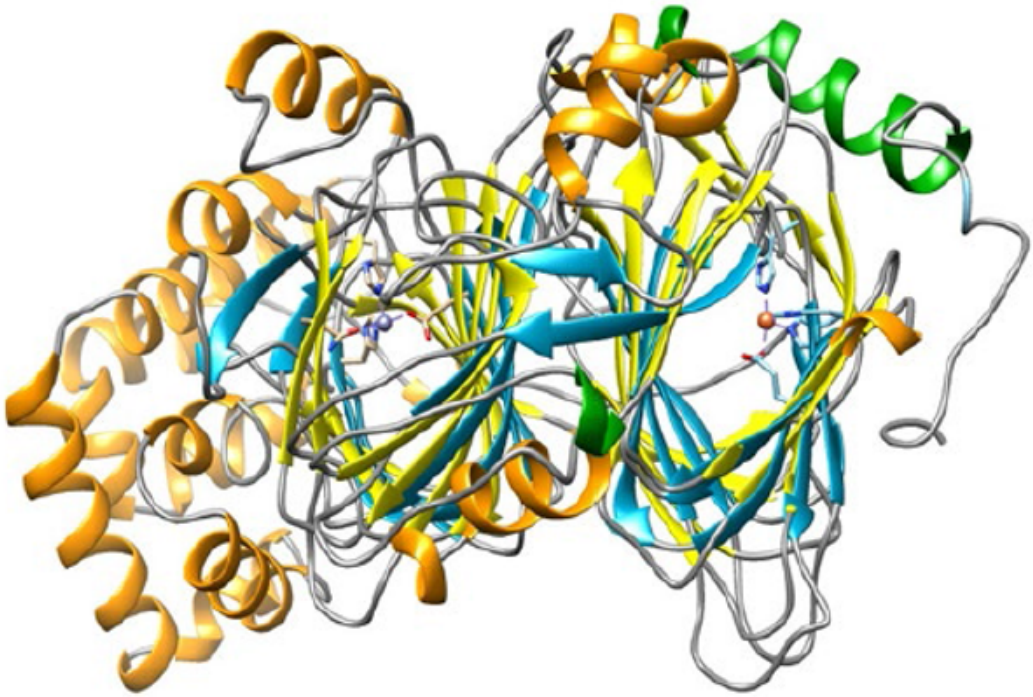
3.4. CONCLUSION:

Pirin Like protein (PLP) is a metalloenzyme that resembles PMI. It belongs to the cupin superfamily of enzymes, and their active sites are located in the characteristic barrel of β-sheets. The biological role of PLP remains unclear. A previous study has shown that PLP could be transformed into an artificial metalloenzyme by substituting its iron cation by a divalent metal cation such as Cu^{2+}. Aiming at the preparation of a chimeric PMI-PLP protein, we first prepared and characterized an artificial metalloprotein based on PLP. The enzyme expression vector yielded PLP comprising a His_6-tag co-expressed with second protein formed of a portion of LacZ fused to His_6-PLP. The coexpression was confirmed by an anti His_6 western blot which revealed two bands instead of a single band. To avoid this coexpression, the expression plasmid was edited by deleting the remaining portion of LacZ. The expression plasmid was digested and a site for specific hydrolysis by TEV protease was successfully introduced flanked by the sequences of the enzyme and that of the His_6-tag. Purified and sequenced plasmid DNA containing the insert sequence was used to transform BL21(DE3) for expressing the protein. After expression and purification on an Ni-NTA column, the enzyme was separated from the His_6-tag by digestion with TEV protease. Following the digestion, the TEV protease with its polyhistidine tag was easily removed from the cleavage reaction by purification on a Ni-NTA affinity column. The produced PLP contains one Fe(II) cation *per* enzyme molecule, therefore the removal of Fe(II) from the active site was performed by incubating the protein with 10^3 equiv. EDTA. To prepare the artificial metalloprotein, Ni^{2+}, Zn^{2+} and Cu^{2+} were added to the iron free protein. A relatively high affinity constant (K_D = 0.5 µM) was measured by fluorescence quenching upon addition of increasing amounts of copper sulfate. The EPR spectrum of the PLP-Cu complex confirmed the presence of a single copper species. TYCHO results showed that the thermal denaturation occurs at 75°C and high concentrations of EDTA and copper sulfate can alter the folding of PLP.

Hence, PLP was successfully transformed into an artificial copper(II)-containing enzyme. Finally, the Diels-Alder activity of the Cu-PLP enzyme can now be examined.

CHAPTER 4

CHIMERIC ENZYMES



CHAPTER 4: CHIMERIC ENZYMES

4.1. INTRODUCTION:

As mentioned in chapter 1 section 1.2, artificial fragments of DNA can now be easily synthesized and subsequently used to express synthetic proteins such as chimeric enzymes.

Chimeric enzymes have been found useful for a wide range of applications including in medicine, chemistry, and agriculture.[28] Therefore, we designed and followed a new strategy to develop a chimeric PMI-PLP enzyme. Such an enzyme contains two neighboring active sites and should be able to catalyze cascades of reactions.

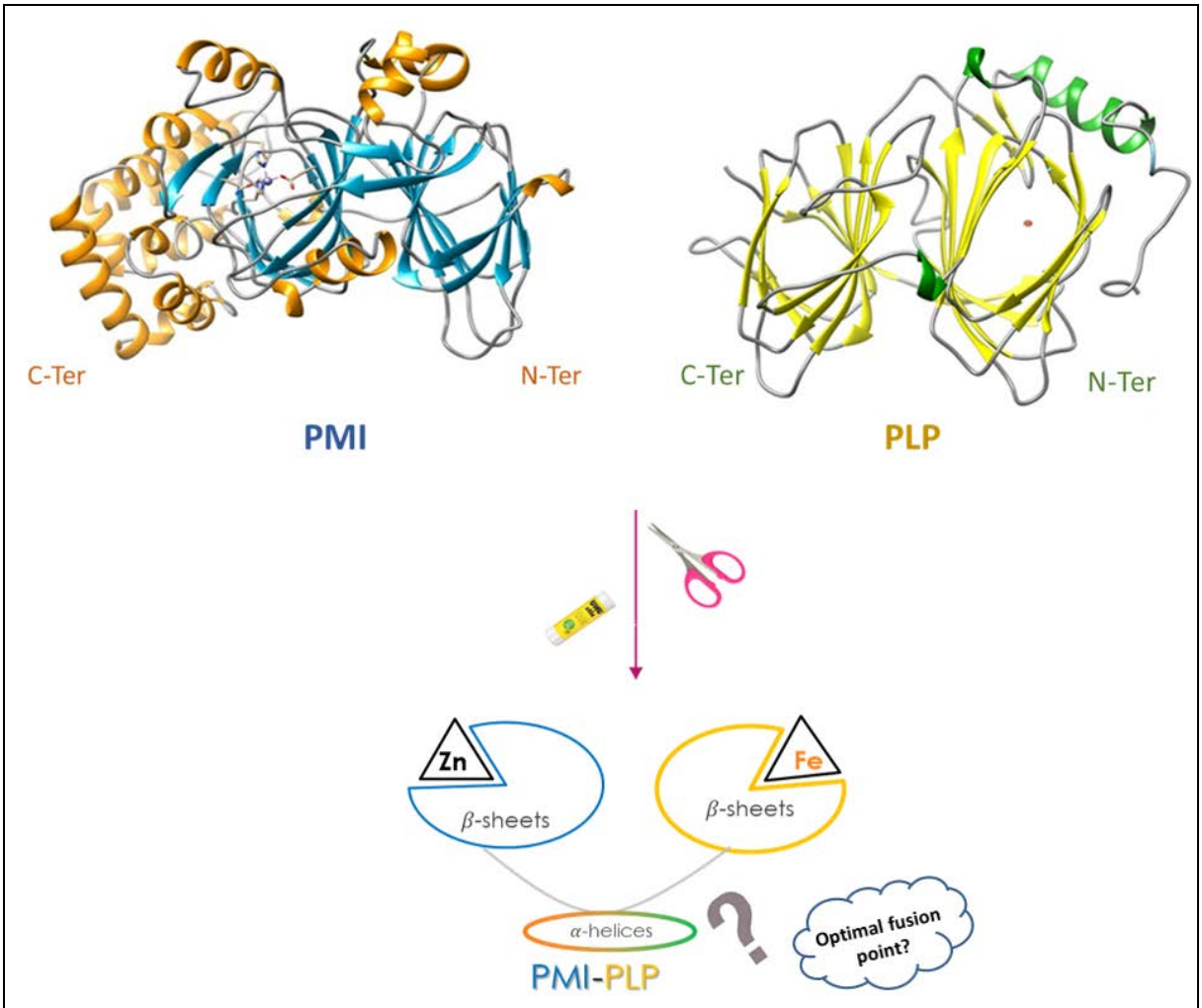
To design fusion enzymes such as chimeric enzymes, it is especially important to start from partner proteins that are active under similar conditions such as temperatures and pH profiles, as mismatched parameters could affect the overall performance.[28] In this project, we therefore chose two enzymes of the Cupin family i.e. PMI and PLP which have high structural similarities. They are both bicupins characterized by a structure formed of two beta barrel domains, each possessing one active site. To construct the PMI-PLP bicupine chimera, the inter-domain of the two metal-containing cupin subunits of the two original bicupins will be fused.

In this chapter, we will discuss the followed strategy to design chimeric enzymes by combining catalytic domains of PMI and PLP. Structural modeling studies, determination of the suitable linking point and molecular biology techniques used for ensuring the successful expression of the enzymes will all be discussed. Furthermore, all the difficulties that we have faced in the construction as well in the purification of the chimeric proteins will be presented.

4.2. RESULTS AND DISCUSSION

4.2.1. Structural modeling

As stated earlier, the chimeric enzyme will be designed by combining the catalytic domains of PMI and PLP (i.e. the C-terminal domain of the PMI and the N-terminal domain of PLP respectively). For this, we searched for a suitable linking point that is featured in PMI and PLP but is neither involved the active sites nor in the overall folding of the proteins. (Scheme 18). Hence, we started searching where there are little interactions between residues in both proteins.



Scheme 18: Strategy for the design of the chimeric PMI-PLP enzyme. Searching for the suitable linking point.

For this, we used the web interface *elNémo* to simulate the low frequency normal modes of vibration of PMI and of PLP. Note that, *elNémo* can perform calculations for all-atom systems and provides simple visualized results.[93,94]

Figure 72 shows the sequence of frames of the video providing calculations resulting by *elNémo*. Upon providing *elNémo* with the crystal structure of PMI (pdb number 1PMI) the visual inspection of the first, second and third normal modes of vibration show that the two cupin domains vibrate independently (Figure 72). This suggests that it is reasonable to associate different cupin domains in a chimeric enzyme. The third normal mode of vibration originates at the center of the linker of the cupin domains. At high intensity this vibration leads to the dissociation of the linker at its center. It is thus reasonable to assume that center of the linker is a suitable association point for the different cupin domains. For PMI, the center of the vibration in the interdomain linker is Met357.

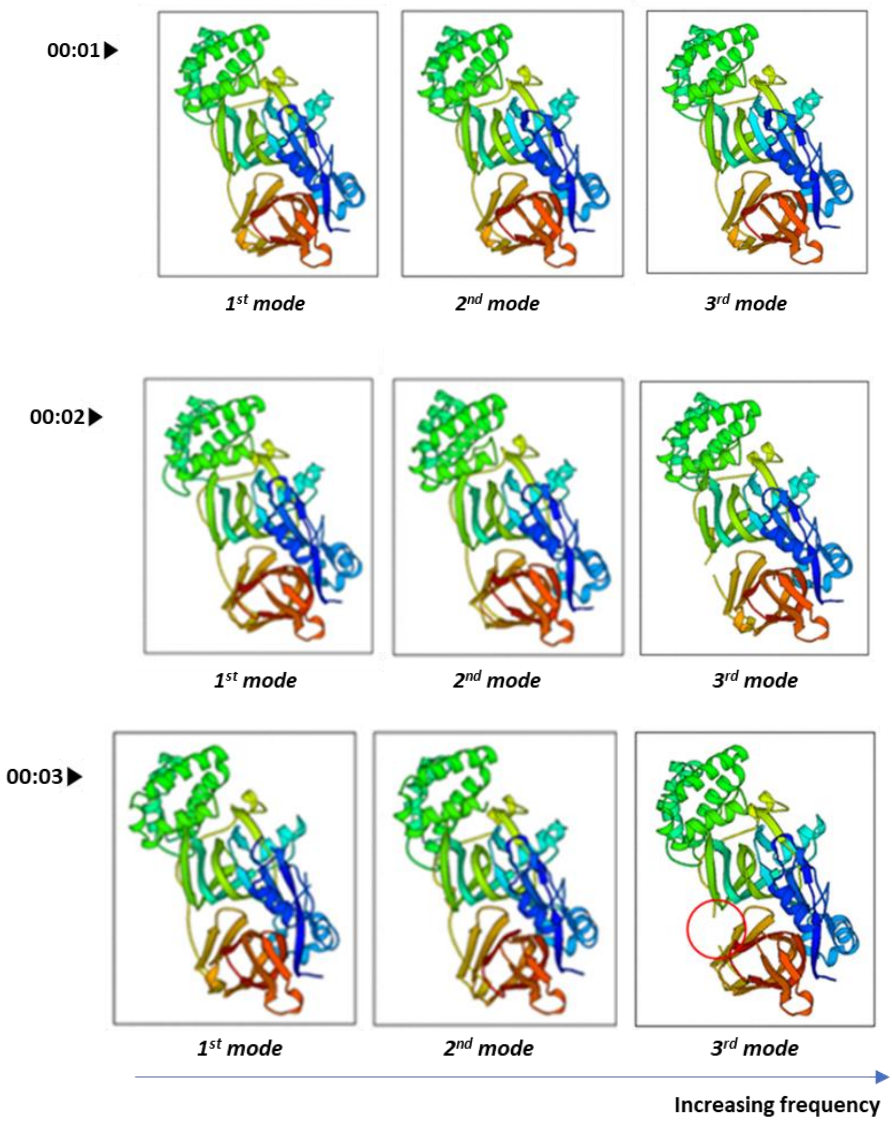


Figure 72: Visualizing and analyzing normal modes of vibration of the PMI protein

For PLP, same procedure was followed to simulate the low frequency normal modes of vibration. Similar studies showed that in the case of PLP, the center of the third normal mode of vibration is Ala138.

Based on these results, the chimera was envisioned to be formed of PMI amino acids Met1 to Met335 fused to PLP amino acids Gly12-Ala138.

Then, we proceeded to the structural modeling of the envisioned chimera using two different molecular modeling servers dedicated for protein structures prediction based on known structures of proteins of similarities sequences: MODELLER and I-TASSER.

As we mentioned before in section 1.2.3.1, MODELLER automatically calculates a model by aligning a sequence with known related structures[95,96] whereas I-TASSER predicts a 3D structural model based on aa sequences.[97]

For this, the structural models of the chimeric PMI-PLP protein were predicted by I-TASSER according to its amino acid sequences (Figure 73). Figure 73 shows that the overall chimeric PMI-PLP protein superposes well with each of the two starting enzymes.

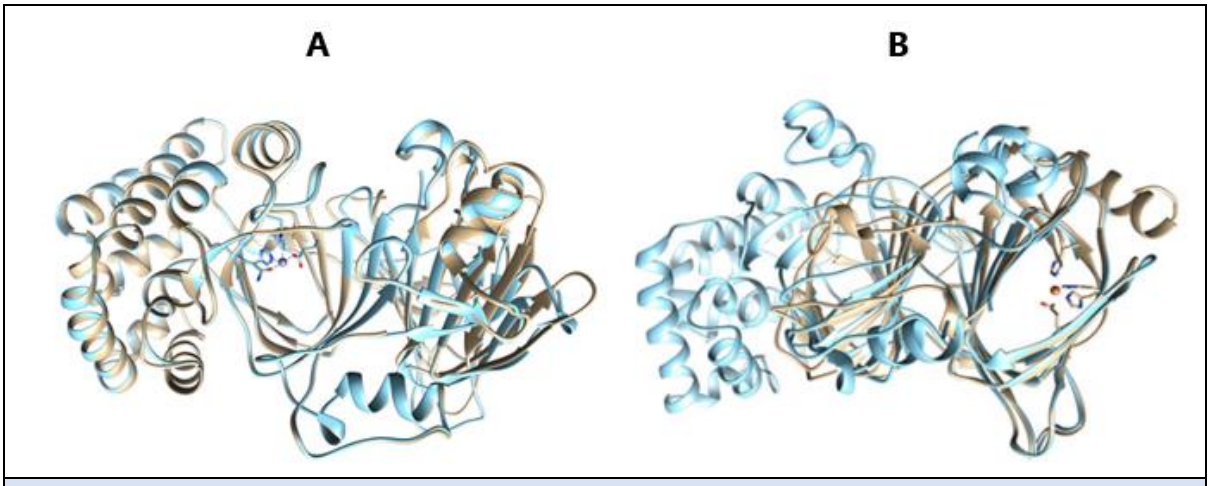


Figure 73: A) Overlay of the crystal structure of **PMI** (in gold) and modeled structure of the chimeric enzyme (in blue); B) Overlay of the structure of **PLP** (in gold-published model[7]) and modeled structure of the chimeric enzyme (in blue).

More careful inspections revealed that molecular modeling using I-TASSER allowed to predict a 3D structure of the chimeric enzyme presenting a correct folding (Figure 74). On the contrary, molecular modeling using MODELLER led to a structure with a helix from the PMI fragment (in red) clashing with the cupin domain of the PLP fragment (in yellow).

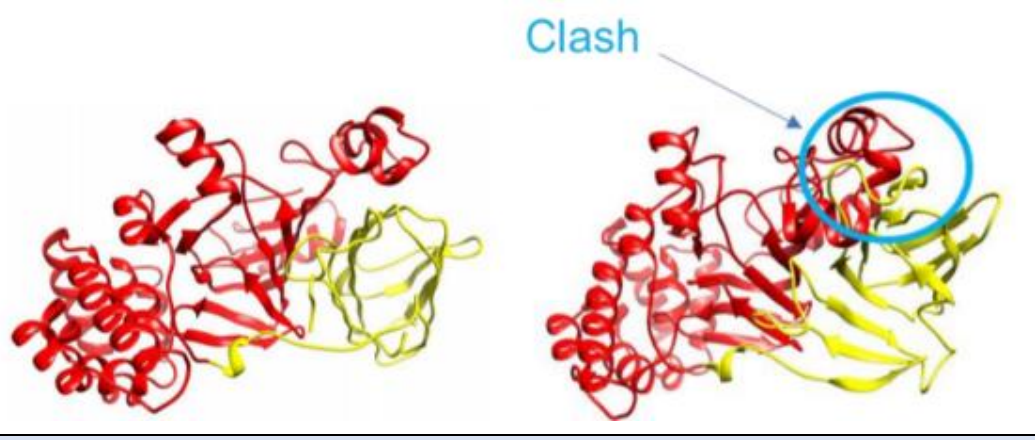


Figure 74: Modeled structure of the chimeric PMI-PLP enzyme as predicted by I-TASSER (left), and by MODELLER server (right). Fragment of PMI in red; fragment of PLP in yellow.

To avoid this possible problem, a second chimera called thereafter ''Chimera 2'' was designed. In this case, the PMI loop responsible for the clash was replaced by the analogous loop from PLP (loop from the second cupin domain of PLP) which is much smaller. Fortunately, the modeling of this new structure by both I-TASSER and MODELLER indicated that Chimera 2 should fold correctly (Figure 75).

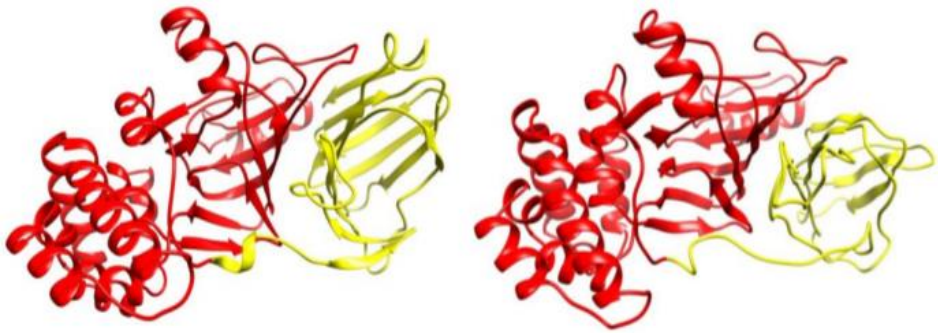
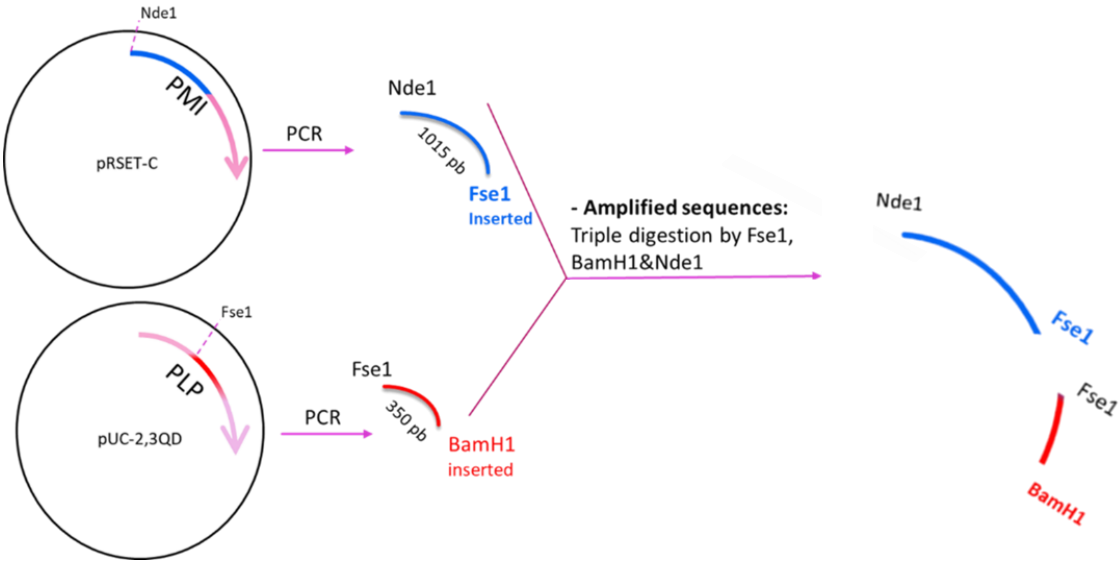


Figure 75: Predicted structures of the chimera 2. Left: prediction with MODELLER. Right: prediction with I-TASSER server. In red: PMI; in yellow: PLP

It was nevertheless decided to start to construct the expression vector of the first chimeric PMI-PLP protein (Chimera 1) as this construct will be made of two protein fragments i.e. one PMI fragment and one PLP fragment, while the construct of Chimera 2 will be made of three protein fragments i.e. one PMI fragment and two PLP fragments. In case of failure in expression or/and purification of Chimera 1, it was envisaged to move on to the design and purification of Chimera 2 according to a plan B.

4.2.2. Design and construction of the expression vector of the first chimeric PMI-PLP protein:

We examined the DNA sequence of the starting proteins that we wanted to amplify and we designed the primers that include restriction sites facilitating the fusion and are suitable PCR amplification of the fragments. The steps followed to construct the expression vector encoding for the Chimera 1 are illustrated below (scheme 19 and scheme 20). Then, the sequence encoding for the PMI fragment was amplified and Nde1 and Fse1 restriction sites were inserted. The former was planned for inserting the sequence in the expression vector. The latter was introduced to be used to fuse the two amplified fragments. The PLP fragment was also amplified and BamH1 and Fse1 restriction sites were inserted for the same reason as earlier.



Scheme 19: Construction of the expression vector of the chimeric enzyme. In blue: the sequence encoding for the PMI's active site. In red: the sequence encoding for the PLP's active site.

The presence of the PCR amplicon and its size were determined by electrophoresis alongside a DNA size marker on a precast 1.2 % agarose gel (Figure 76). Figure 76 shows bands around 1300 bp and 400 bp corresponding respectively to the molecular weight of PMI and PLP fragments. The two PCR products were then purified.

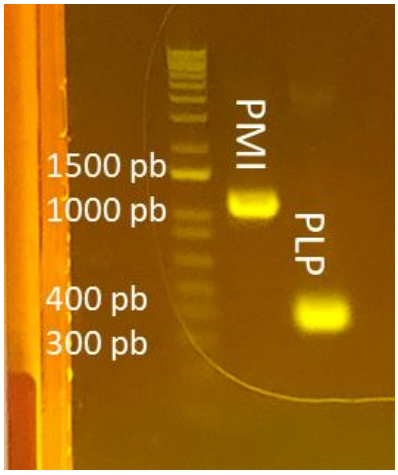
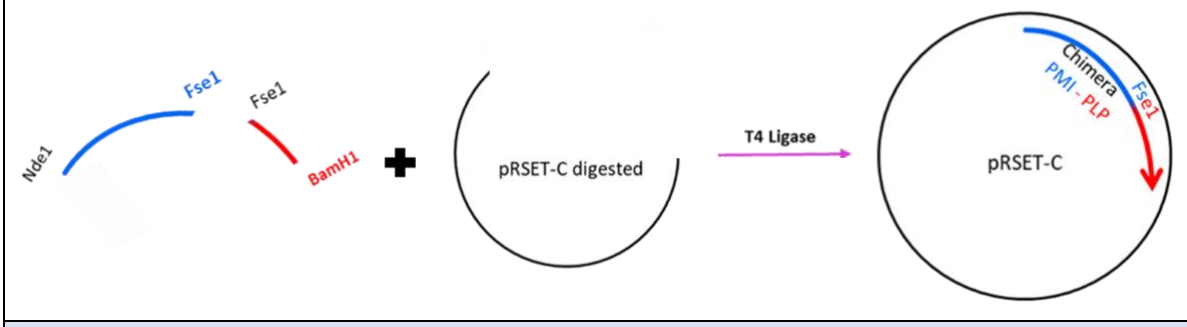


Figure 76: Agarose gel electrophoresis analysis stained with SYBR Green of PCR products obtained using the primers which were complementary to the sequence coding for the chimeric protein.

The PMI and PLP purified fragments were then triple digested by using the restriction enzymes Fse1, Nde1 and BamH1.

In parallel, the plasmid for we previously used for expressing PMI (pRSET-C vector containing PMI sequence flanked by BamH1 and Nde1 restriction sites) was also double digested by using the restriction enzymes Nde1 and BamH1. The digestion reaction of the plasmid was analyzed by electrophoresis alongside a DNA on a 1.2 % agarose gel. After 45 minutes of migration at 70 Volts, the colorant was at the edge of the gel. The electric current was then stopped, and the gel was visualized with an ultraviolet (UV) transilluminator (Biorad). The superior DNA band, that had the desired MW of the vector (pRSET-C), was cutout and purified while the inferior DNA band corresponding to the PMI's sequence was discarded. DNA extraction from agarose gel followed. The purified vector was mixed with an excess of DNA inserts (3 molar equivalents of DNA sequences encoding for PMI and PLP amplified and digested) and a ligase mix, following the providers procedure (scheme 20). The ligation mixture was used to transform E. *coli* (DH5alfa) strain. This ligation was repeated several times under various conditions, but transformations all failed.



Scheme 20: Assembly of the purified amplified sequences encoding for the PMI and PLP's active sites in the pRSET-C vector.

To solve this problem, we proceeded with the **Gibson assembly method** that is specifically designed for gene fragment assembly. New primers were thus designed and new PCRs were run.

The desired fragments of the PMI and PLP vectors were linearized and amplified by PCR using primers containing 20 bp overlap between both fragments. PCR products were purified and added to the Gibson Assembly mix. The reaction was carried out under isothermal conditions by three enzymes provided in the mix: a 5' exonuclease that generates long overhangs, a polymerase that fills in the gaps of the annealed single strand regions, and a DNA ligase that seals the nicks of the annealed and filled-in gaps (Figure 77).

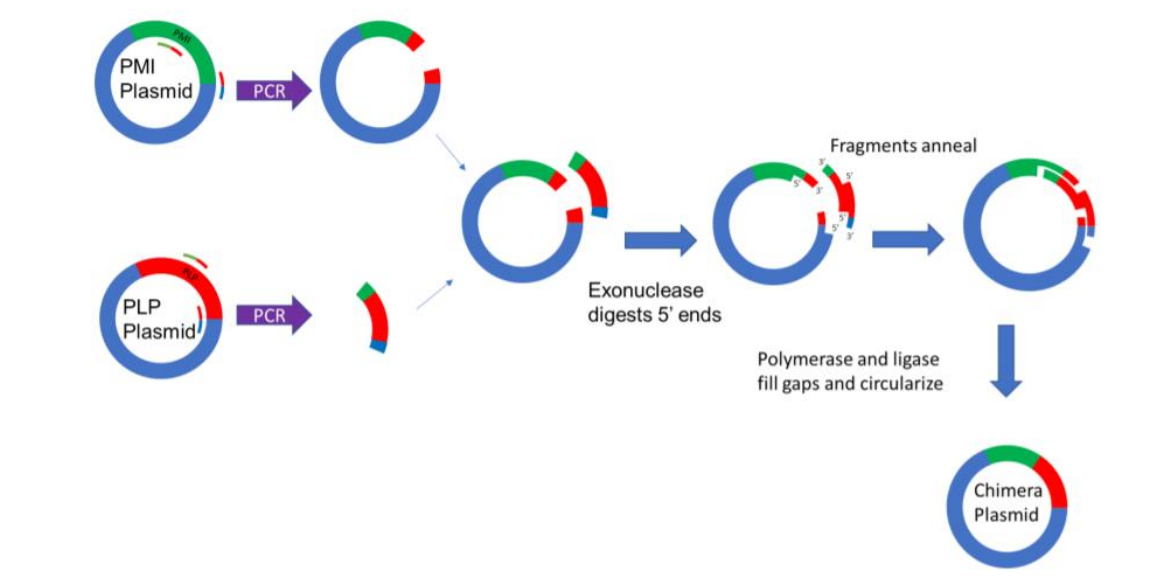


Figure 77: Principle of Gibson assembly method.

Subsequently, DpnI i.e. a restriction enzyme that possesses a recognition sites that is methylated, was added to the mix to catalyze the hydrolysis of the original PMI and PLP expression vectors. Transformation resulted in bacteria that contained the desired plasmid as demonstrated by specific PCR amplifications.

First, we verified the presence or absence of DNA insert in plasmid constructs by Colony PCR. Then, in order to confirm the presence of the DNA sequence encoding for the Chimera 1, sequencing was done at Eurofins using the Sanger's method. It showed the expected sequence. E.*coli* BL21 (DE3) strain was used for expressing the protein on agarose/ampicillin plates.

4.2.3. Expression and purification of the chimeric PMI-PLP enzyme:

To obtain substantial quantities of protein, the modified plasmid was used to transform E. *coli* BL21(DE3) strain. A pre-inoculum from a single colony was grown in a starter LB-ampicillin (100 µg/mL) culture, at 37 °C for 18 hours. Terrific Broth (TB) supplemented with ampicillin (50 µg/mL) was inoculated from the preinocculum at 1/100th and grown at 37 °C until the OD_{600} reached a value between 0.6 and 0.8. At this stage, either 0, 0.1, 0.2, or 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added and incubation was continued at 20 °C for 18 h.

The expression level of the induced protein was checked by SDS PAGE electrophoresis followed by Coomassie blue staining. For this, bacterial cells were pelleted and resuspended with sterile water. As shown in Figure 78, no band was observed without IPTG addition. On the contrary, a band of about 55 kDa corresponding to the molecular weight of the His_6 chimeric enzyme was observed upon addition of 0.1 mM IPTG or more. Hence, this 0.1 mM IPTG is enough to induce the production of the first chimeric PMI-PLP protein.

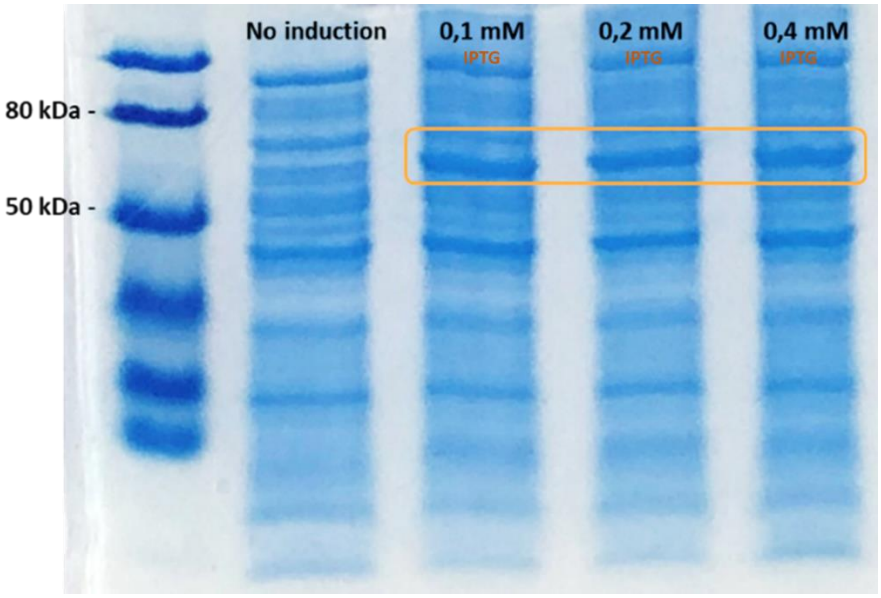


Figure 78: Coomassie blue stained SDS page electrophoresis of pellet fractions after the expression of the His_6 tagged Chimera 1 by addition of different concentrations of IPTG.

In parallel, anti His_6-tag antibody western blot was performed on the bacterial pellets. It confirms that the produced induced protein was indeed His tagged (Figure 79).

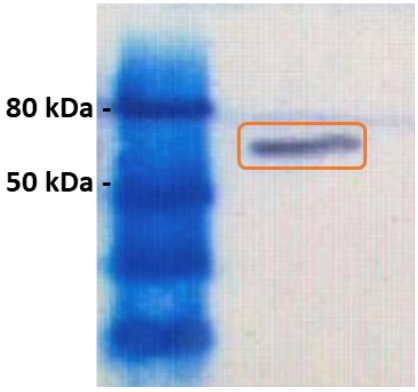


Figure 79: Western blot image of the His_6 – Chimera 1.

Cultures in which protein production was induced by addition of IPTG were pooled and centrifuged. The obtained pellets of bacteria were resuspended in lysis buffer containing lysozyme and benzonase. The slurry was sonicated and then centrifuged to eliminate all insoluble material. The clear solution was passed through a Ni-NTA agarose column and retained proteins were eluted by increasing the concentration of imidazole in the elution buffer.

The purification was monitored by SDS-PAGE electrophoresis followed by Coomassie blue staining that revealed a band at a molecular weight of about 55 kDa (Lanes 7, 8 and 9 corresponding to this protein). Unfortunately, other bands were also present thus the tagged protein was not pure (Figure 80A). In parallel, western blotting using anti His-tag antibody confirmed its identity (lanes 7,8 and 9, Figure 80B).

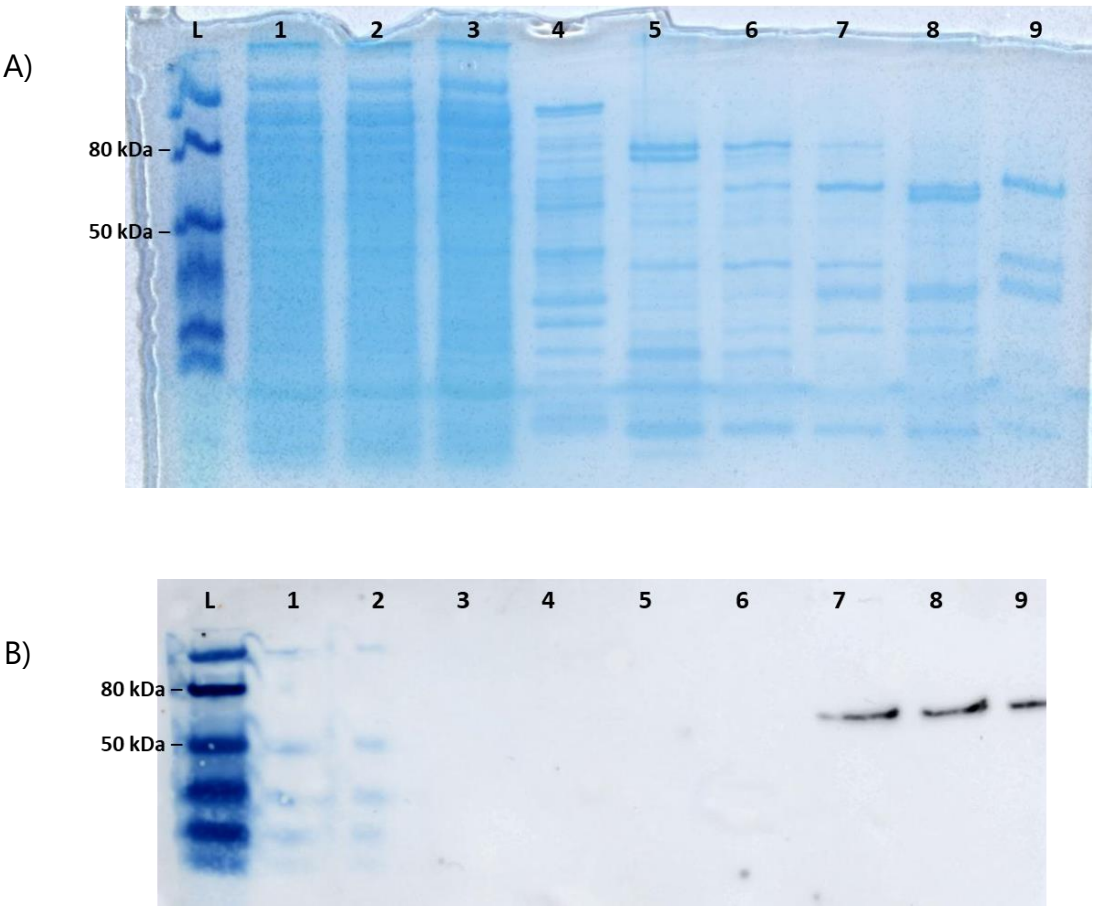


Figure 80: A) Coomassie blue stained SDS-PAGE electrophoresis of collected fractions of the first purification on an Ni-NTA column of the chimeric enzyme. B) Western Blot image. In Both cases, lanes: 1, 2 and 3: Flow through; 4, 5 and 6: Washed fractions; 7, 8 and 9: Eluted fractions; L: Protein Ladder

The fraction containing the tagged protein were combined and the protein concentration measured. Only 1 g per liter of culture were measured. This result indicates that only a small quantity of the over expressed protein observed after induction with 0.1 mM IPTG (Figure 78) was not found in the soluble protein fraction

that was used for purification. The low purification quality was thus due to low quantity of soluble chimera.

In addition, we assayed to optimize the separation conditions by changing the imidazole concentration in order to reduce the binding of contaminant proteins, and thus increase the final purity. Nevertheless, a TEV protease cleavage followed by a purification on His Trap Column were done but the purification remained unsuccessful.

Several strategies were followed to attempt to favor the solubilization of the protein such as:

- Changing the bacterial strain and using *E. coli* BL21(DE3) pLysS cells which allow high efficiency protein expression, express T7 RNA polymerase which can improve the expression and also encode T7 lysozyme that suppresses basal expression of toxic target proteins prior to induction. This can prevent toxicity problems when adding IPTG.
- Variation of the induction temperature and culture time to optimize the chimera's expression (37°C for 4 hours, 6 hours and 8 hours, 30°C for 6 hours). Nevertheless, it has been known that usually low temperatures can improve the protein solubility.[94] Hence, 37 °C gives faster growth and usually higher levels of production for proteins that don't tend to form insoluble inclusion bodies.
- Cool down the culture by placing it on ice before induction with IPTG aiming at decreasing the rate of protein synthesis, especially if the chimera is toxic for E.*coli* proteins.

Unfortunately, despite all these attempts, the soluble fraction remained low and the purification not successful. Because folding problems were suspected for chimera 1 (see section 4.2.1), we did not explore further options for its solubilization. Instead, we proceeded by using chimera 2.

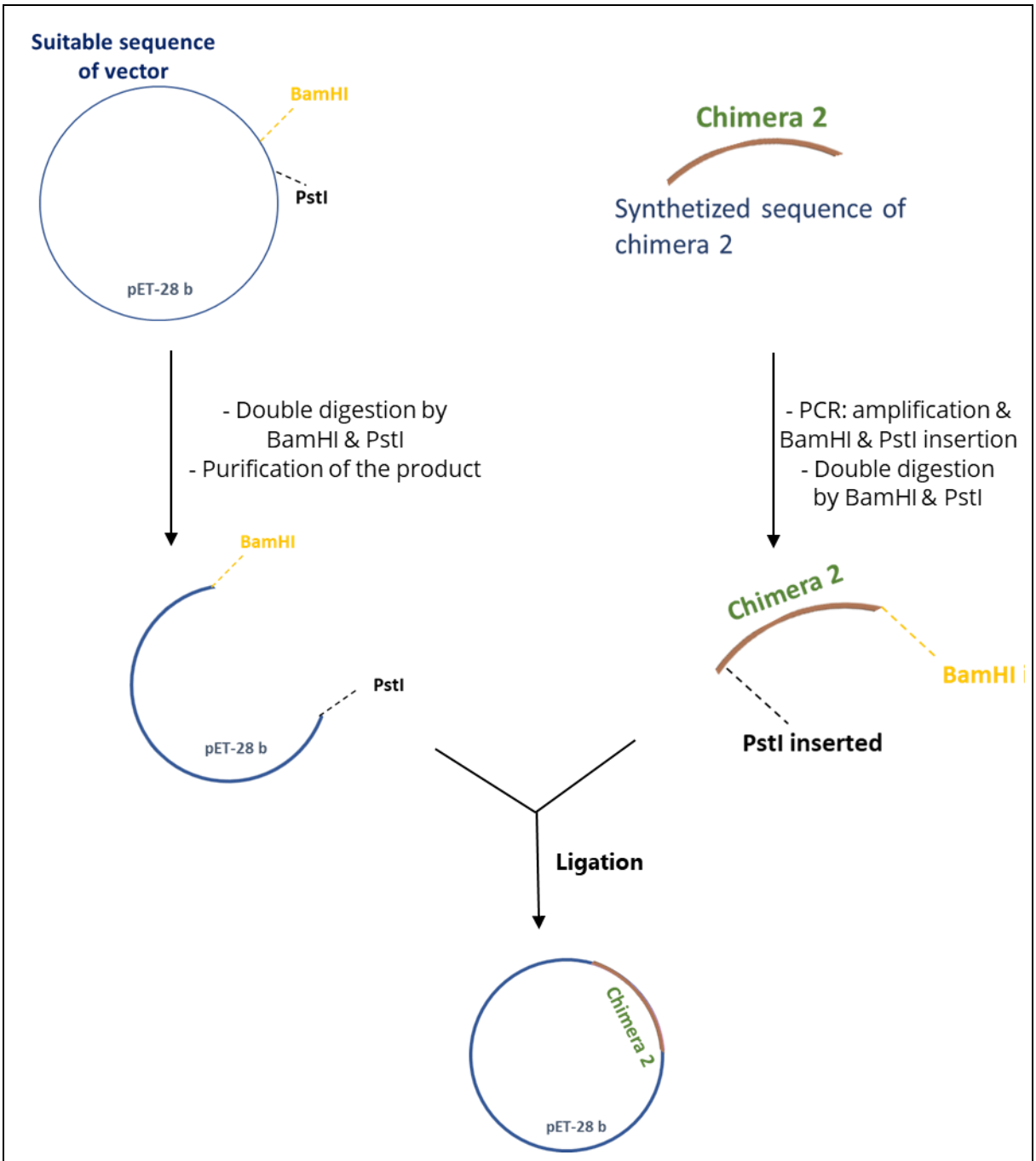
4.2.4. Design and Construction of the expression vector of the Chimera 2:

The sequence encoding for chimera 2 was purchased. As we mentioned before, Chimera 2 is similar to Chimera 1 with the difference that a PMI loop suspected to clash with a cupin barrel in chimera 1 was replaced by a much smaller loop from PLP in chimera 2 (Figure 81).

Figure 81: Amino acid (aa) sequence alignment of Chimera 1 with the synthetic Chimera 2. Identical aa are in black, similar aa are in blue and different aa are in red.

The steps followed to construct the expression vector encoding for the Chimera 2 are illustrated below (scheme 21).

The pET-28b plasmid known for its high efficiency for the expression of proteins was chosen as a vector. This vector is able to reduce leaky expression of proteins which is under the control of T7 lac promoter to promote tightly and strong gene expression.[93,98] This vector encoding for the **S**mall **U**biquitin-like **Mo**difier (SUMO) tag protein was double digested by BamHI and PstI and then purified. The result of the digestion of the vector was analyzed by agarose gel electrophoresis and ethidium bromide staining (Figure 82). A single band around 6 Kb was visible after the digestion which confirmed the digestion of the plasmid. The undigested plasmid bands appeared at three different levels in the gel which can be assigned to the different tertiary structures of circular plasmid DNA (Supercoiled, nicked and relaxed).



Scheme 21: Construction of the expression vector of the Chimera 2.

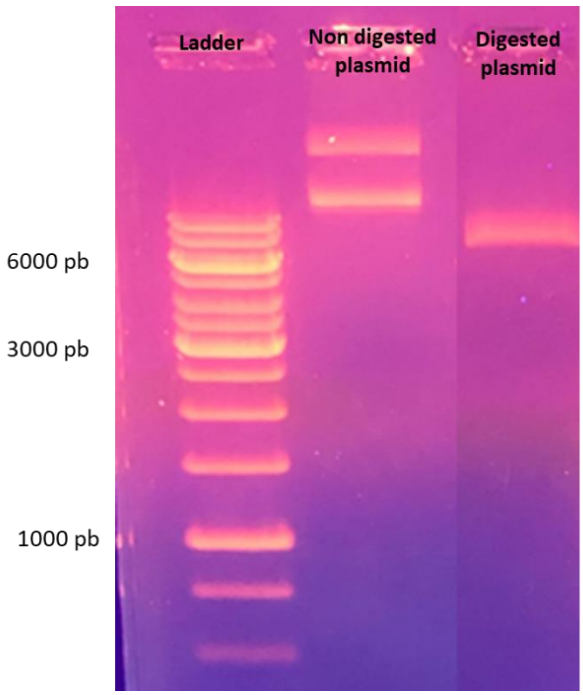


Figure 82: Analysis by agarose gel electrophoresis followed by ETBr staining of the result of the digestion of the pET-28b plasmid by BamHI and PstI. From left to right: First lane: Molecular adder; Second lane: Intact plasmid; Third lane: band at about 6 kB corresponding to the digested plasmid.

The digested plasmid was extracted from the gel as described in materials and methods. In parallel, the synthetized sequence of chimera 2 was amplified by PCR, which also allowed the insertion of restriction sites BamhI and PstI. The amplification was followed by a double digestion by BamhI and PstI. The result of this digestion was purified and analyzed by agarose gel electrophoresis and ETBr. The two purified products extracted from the gel could anneal. Excess of the digested double stranded DNA insert encoding for the chimera 2 was added to the digested pET-28b plasmid and the mixture ligated using instant DNA ligase. The ligation mixture was used to transform DH5alfa competent bacteria. A pre-inoculum from a single colony was grown in a starter culture LB supplemented with 75 µg/mL Kanamycin and cells were grown overnight. Then, the plasmid was extracted and its sequence was verified by sequencing. It showed the expected sequence.

4.2.5. Expression of Chimera 2:

To obtain large quantities of protein, the sequenced purified plasmid DNA containing the insert sequence was used to transform BL21(DE3) that were then selected on agar/kanamycin. A pre-inoculum from a single colony was grown in a starter culture YT-Kanamycin. YT supplemented with Kanamycin was inoculated from the preinocculum at 1/100th and grown at 37 °C until the OD600 reached a value between 0.6 and 0.8. Different concentrations of IPTG were then added and incubation was continued at 20 °C for 18 h. Bacterial cell were pelleted and kept at -20 °C. Analysis of supernatant fractions and pellets was done by SDS-PAGE followed by Coomassie blue staining (Figure 83).

Upon five hours of induction, a band with the same intensity (Figure 83 left) was observed in all pellet lanes at a MW of about 60 kDa corresponding to different IPTG concentrations. However, no band was observed in the first lane where the protein was not induced. This first assay shows that 0,1 mM IPTG is enough to induce the production of the Chimera 2 in BL21 E.*coli*.

Upon 18 hours of induction, a band with high intensity than the one before was observed in all pellet lanes (lanes 2 and 3, Figure 83 right) at a MW of about 61 kDa, which reflected the expression of Chimera 2. In contrast, no protein band could be detected in the supernatant fractions (lane 1, Figure 83 right). This means that the chimera 2 is better expressed at low temperature for overnight incubation.

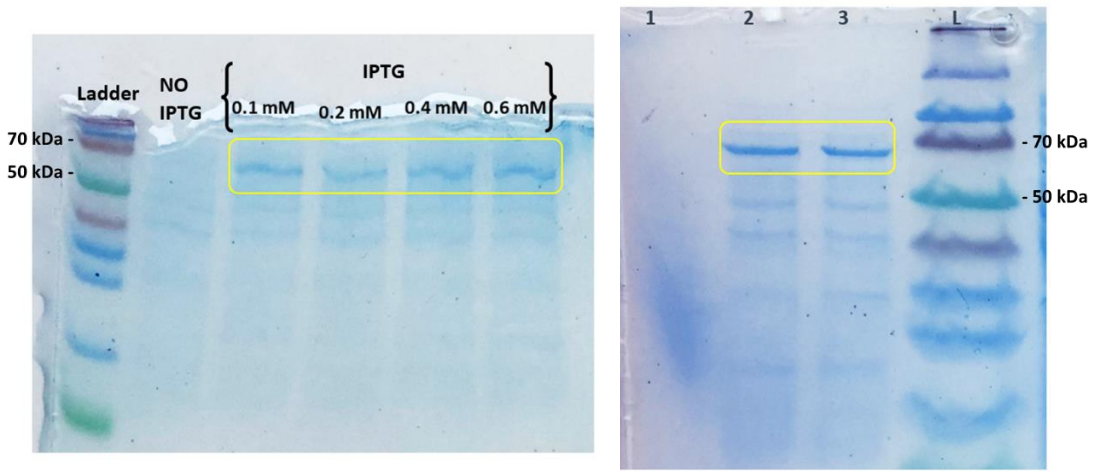


Figure 83: SDS-PAGE electrophoresis followed by Coomassie blue staining of supernatant fractions and pellets after the expression of the Chimera 2 by addition of different concentrations of IPTG. Left: After 5 hours of induction; Right: Overnight expression; **Lane 1:** Supernatant fraction; **Lanes 2 and 3:** Pellet fractions **L**: Protein Ladder.

Before proceeding to the first purification, we tested the solubility of this protein. For this, BL21 cells, transformed with the plasmid containing the DNA encoding for chimera 2, were grown either in LB or in TB medium. The produced protein was then analyzed on SDS PAGE gel before and after the cell lysis in order to verify its solubility in the cells.

As shown in Figure 84, before lysis, Chimera 2 is well expressed in cells. First, in absence of induction, in LB medium, no band characteristic of this protein was observed after Coomassie blue staining (lane 1). On the contrary, after IPTG induction, a band of high intensity at MW of about 60 kDa was observed in both media (lanes 2, 3 and 4 for the LB medium, lanes 6, 7, 8 for the TB medium). In the case of TB, induction was observed without IPTG addition (lane 5), as it is often the case with this auto-inducing medium.

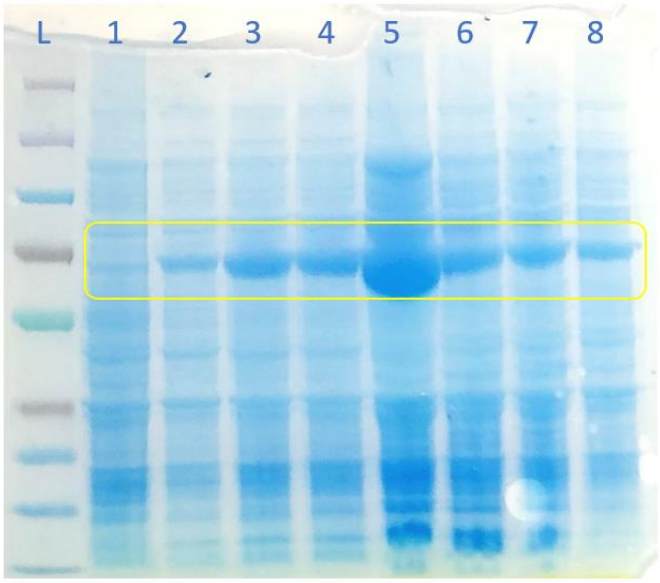


Figure 84: Characterization of the expression of the chimeric 2 before cell lysis. Cultures were made in LB and TB media supplemented with Kanamycin. L: Ladder; Lanes 1 to 4: LB medium and Lanes 5 to 8: TB medium. Lanes: 1&5: No induction – Lanes 2&6: 0.1 mM IPTG – Lanes 3 & 7: 0.2 mM IPTG – Lanes 4&8: 0.4 mM IPTG

After cell lysis, SDS-PAGE analysis followed by Coomassie blue staining of the various insoluble fractions (inf) and soluble supernatant fractions (sf) was performed (Figure 85). Without IPTG induction in LB medium, the chimeric protein was not found in the soluble fraction (Lanes 1 and 2).

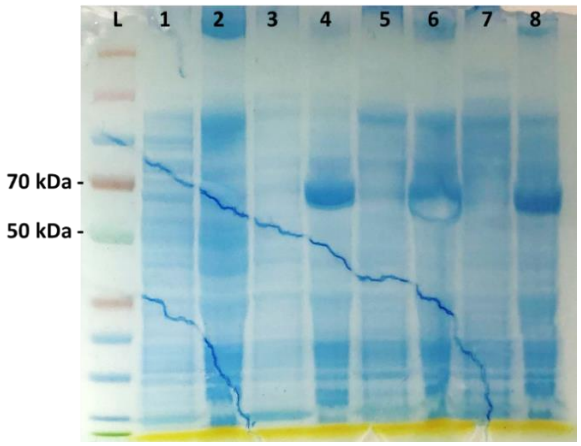


Figure 85: SDS-PAGE electrophoresis followed by Coomassie blue staining of supernatant fractions and pellets after the cell lysis of the expressed Chimera 2. Lanes 1 to 4: in LB medium Lanes 5 to 8: in TB medium. Lanes 1, 3, 5 and 7: Soluble supernatant fractions; Lanes 2, 4, 6 and 8: insoluble fractions. L: Protein Ladder.

However, upon induction by 0,1 mM IPTG a band with a high intensity, at a molecular weight of about 60 kDa corresponding to that of chimera 2 could be observed in lanes 4, 6 and 8 (induction with 100 µM, 200 µM and 400 µM IPTG respectively), thus chimera 2 was found only in the lysate most probably in inclusion bodies. Such a band could not be seen in lanes 3,5 and 7 corresponding to the soluble fractions. Hence, this problem may be due to different reasons such as:

- Chimera 2 can be a toxic protein.

- Chimera 2 does not fold correctly.

This was unexpected since, as mentioned before in section 4.1, the structural modeling using both MODELLER and I-TASSER indicated that Chimera 2 should fold correctly. A recent new prediction of the structure of chimera 2 using Rosetta[95] confirmed that it should fold correctly.

- Solubility problems deriving from those of the starting enzymes:

To check this point, we moved to verify the solubility of the two PMI and PLP starting enzymes. The same procedure was followed as for the Chimera 2 and SDS-PAGE electrophoresis followed by Coomassie blue staining (Figure 86) as well as Western Blot (Figure 87) of supernatant fractions and pellets after the lysis of the expressed proteins were done.

A comparison of the SDS-PAGE analysis with the Western blot using anti His_6-tag antibody shows that the PMI is indeed expressed in the supernatant, as a band at a

MW of about 50 kDa was observed in lane S_{PMI}, which corresponds to the molecular weight of His_6-PMI.

On the other hand, it appears that PLP is expressed in both supernatant and inclusion bodies. Indeed, a band at MW of about 33 kDa was observed in both P_{PLP} and S_{PLP} lanes.

Concerning the chimeras, bands were observed in lanes P_{Ch1} and P_{Ch2} at MW corresponding to the expected MW respectively for Chimera 1 (50 kDa) and Chimera 2 (60 kDa) which suggests that the two proteins are mainly expressed in the inclusion bodies rather than the supernatant. Thus, this can be due to the expression of PLP in both supernatant and inclusion bodies. Therefore, Figure 87 suggests that the Chimera 2 appears to be better expressed (P_{Ch2}) than the first chimeric PMI-PLP protein (P_{Ch1}).

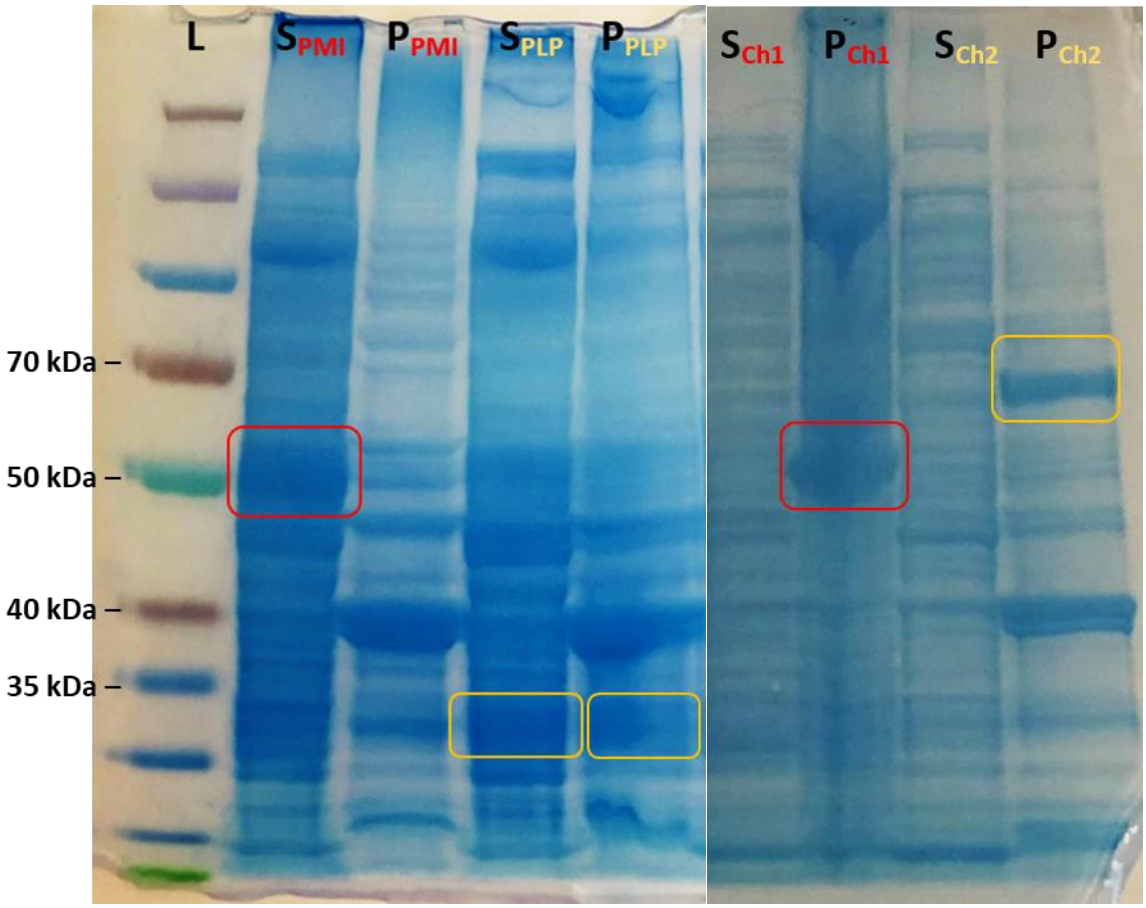


Figure 86: SDS-PAGE electrophoresis followed by Coomassie blue staining of supernatant fractions and pellets after the cell lysis of the expressed proteins. S: Supernatant – P: Pellet – Ch1: Chimera 1 – Ch2: Chimera 2

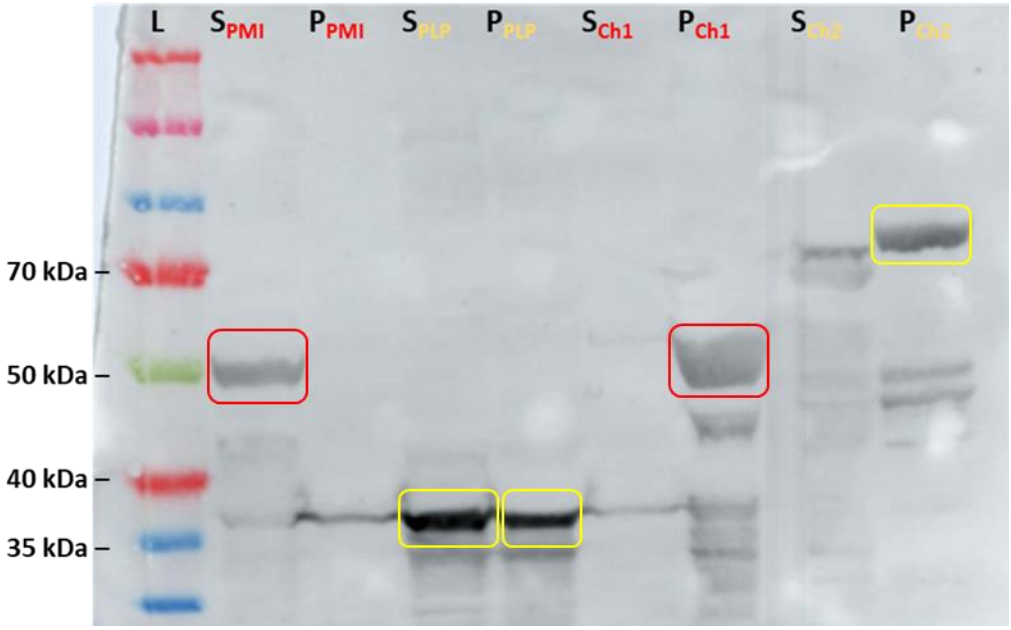


Figure 87: Western Blot image. L: Ladder – S: Supernatant – P: Pellet - Ch1: Chimera 1 – Ch2: Chimera 2

Several means can be envisioned to enhance the expression of a protein under a soluble form including:

i) The change of the expression vector by using weak promoters.[96,97,99] It has been know that regulating the gene's expression is one of the most important processes in protein production.[99] Therefore, promoters are one of the most important regulatory components in a host cell since they may directly influence the degree of recombinant gene expression.[99]

ii) The change of the host strain to avoid possible missing post translational modifications or incorrect disulfide formation.[96,100] It has been shown that protein's solubility can be improved upon osmotic induction with high salt as NaCl.[100]

iii) The coexpression of the protein with chaperones that aid its correct folding[96,101] and in some cases, leads to enhance the production of active proteins.[102,103]

The solubilization of the protein out of inclusion bodies (denaturation of the protein followed by renaturation) can also be envisioned especially in this case as the structure of the refolded protein is expected via molecular modeling and the starting proteins PMI and PLP.[101]

Hence, to generate a soluble chimeric protein, we proceeded with the coexpression of chaperones. The latter help in folding other proteins and maintaining the proteostasis in cells, which prevents the formation of aggregates and thus inclusion bodies.[102,103] For this, competent cells containing a pG-KJE6 plasmid[102] were employed as it encodes for DnaK-DnaJ-GrpE and GroEL-GroES chaperone teams[102] (Figure 88). These two chaperone teams are known to have different and cooperative functions:[102,104,105] On the first hand, the lack of one chaperone team can increase the requirement for the other and vice versa. On the other hand, a protein can be stabilized in substantially aggregated forms by DnaK-DnaJ mutations while it can be destabilized by GroELand GroES mutations.[102]

In addition, these proteins can enhance the stability, the folding and the disaggregation of different individual proteins.[102,104,105]

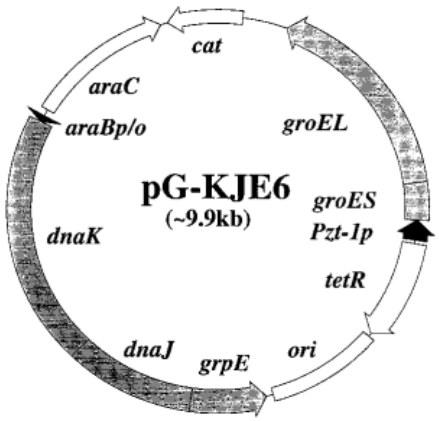


Figure 88: Structure of chaperone coexpression plasmid pG-KJE6. Ori: replication origin of pACYC184; cat: chloramphenicol acetyltransferase gene; Pzt-1p: promoter; tetR: tetR repressor gene; araBp/o: araB promoter-operator; araC: araC repressor gene.[102]

For this, plasmid DNA encoding for chimera 2 was used to transform BL21-GOLD *E.coli* competent cells already containing the chaperone pG-KJE6 plasmid that were then selected on agar/kanamycin/chloramphenicol. A pre-inoculum from a single colony was grown in a starter culture in LB or TB supplemented with Kanamycin/Chloramphenicol. LB or TB supplemented with Kanamycin and Chloramphenicol was inoculated from the pre-inoculum at 1/100th and grown at 37 °C until the OD600 reached a value of 0.6. Arabinose was then added. When the OD600 reached a value of 0.75, IPTG was added and incubation was continued at 20 °C for 18 h.

SDS-PAGE analysis of supernatant fractions and pellets reflects the expression of the SUMO tag-Chimera 2 protein as a band with high intensity was observed at a MW of about 60 kDa in all inclusion bodies after Coomassie blue staining (Lanes 2, 4, 6 and 8). In contrast, no bands were observed in lanes 1, 3, 5 and 7 corresponding to the supernatant fractions (Figure 89).

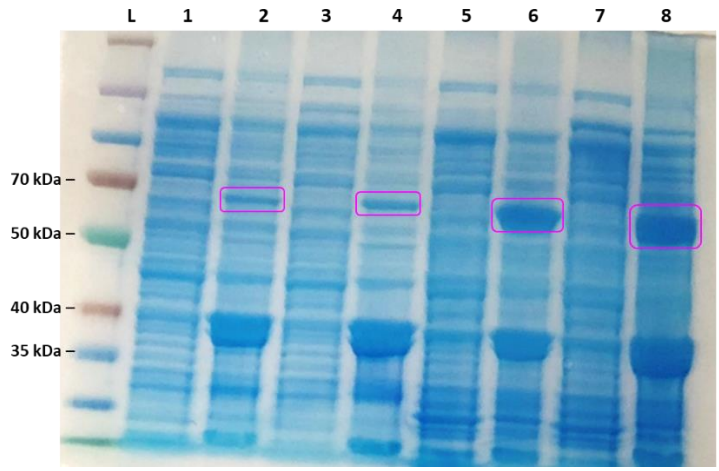


Figure 89: Analysis by SDS PAGE to characterize the expression of the chimera 2 co-expressed with chaperone proteins in LB and TB medium. Lanes: 1, 2, 3 and 4: LB medium - 5, 6, 7 and 8: TB medium. 1 and 5: = supernatant, no induction – 2 and 6: inclusion body, no induction – 3 and 7: supernatatnt, induction – 4 and 8: inclusion body, induction – L: Ladder

As shown in Figure 90, a band with high intensity corresponding to a protein with a molecular weight of about 61 kDa was observed for fractions 5 to 8, which suggested that those fractions contained the SUMO tagged Chimera 2 (SUMO-Chimera 2). Hence, western blot confirms that Chimera 2 remains highly expressed in the inclusion bodies and its coexpression with chaperones failed to solubilize it.

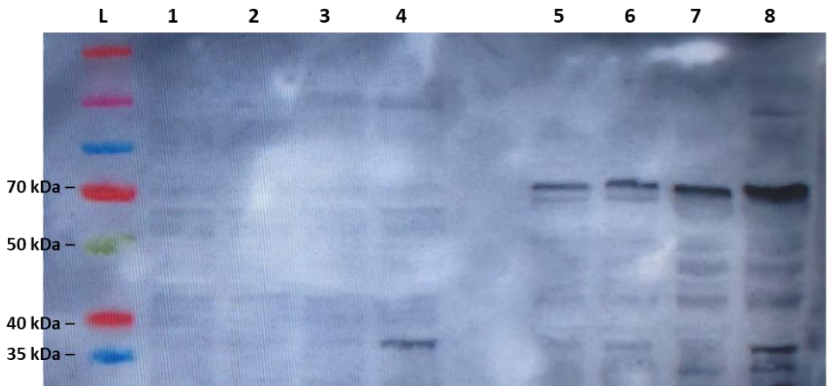


Figure 90: Western blot image using anti his tag antibody blot to characterize the expression of the chimera 2 co-expressed with chaperone proteins in LB and TB medium. Lanes 1, 2, 3 and 4: supernatant fractions. Lanes 5, 6, 7 and 8: inclusion bodies fractions.

4.3. CONCLUSION:

We have studied here an interesting new strategy that aims to develop an artificial chimeric metalloenzyme that possesses multiple active sites by the fusion of catalytic domains of 2 related enzymes: PMI and PLP.

First, the strategy we followed to prepare the PMI-PLP bicupin chimera was based on structural modeling studies and on the determination of the linking point between the two catalytic domains of PMI and PLP. *elNémo* was used *to* simulate to simulate the low frequency normal modes of vibration of PMI and of PLP. The movements of all atoms/domains in PMI and PLP were thus studied and visualized. The 1st and the 2nd mode of vibration of PMI show that the 2 cupin domains vibrate independently. The 3rd normal mode of vibration originates at the center of the linker of the cupin domains. So, the association of the two different cupin domains can occur by fusion at the center of their interdomain linkers.

The sequences encoding for the PMI active site and the PLP active site were cloned by PCR. PCR products were then purified, annealed and ligated via the Gibson Assembly method. Hence, the expression vector containing the DNA encoding for the PMI-PLP chimera was constructed, and the chimera was successfully expressed in *E.coli* BL21 but unfortunately the first purification was not satisfying.

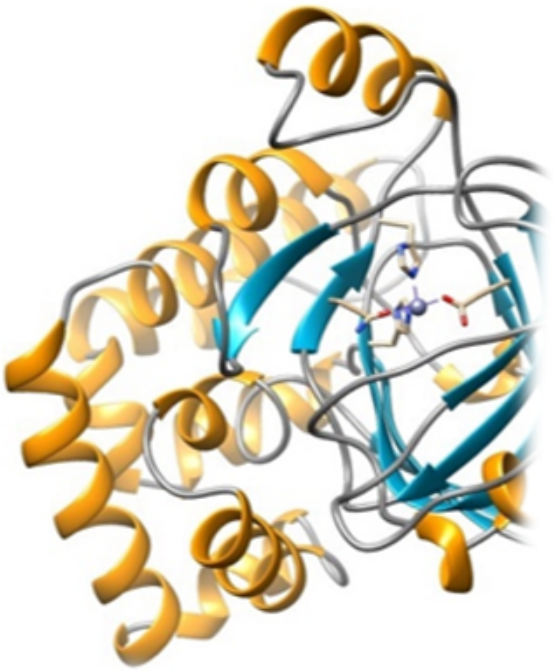
Since the obtained protein was not pure, we tried to optimize the purification via different methods. Since molecular modelling using MODELLER revealed a possible clash between the α helixes in the variable region deriving from PMI and the barrel deriving from PLP, a second chimeric enzyme, Chimera 2, was constructed by replacing the PMI loop responsible for the clash by the analog PLP loop which is much smaller. Chimera 2 was expressed better than the first chimera but unfortunately, the expressed protein was found in the insoluble fractions, most probably in inclusion bodies. The co-expression of chaperones proteins with chimera 2 failed in solubilization the protein.

As perspectives, alternative methods such as the use of denaturing agents will be examined to solubilize the inclusion bodies. Hence, the protein will be denatured which will require its refolding.[106,107] The correct folding of the chimera could be verified by comparison with that of the parental proteins.

In the future, this chimeric enzyme could be used to catalyze different reactions and different cascade reactions.

CHAPTER 5

HALF ENZYMES



CHAPTER 5: HALF PROTEINS

5.1. INTRODUCTION

As we mentioned before, PMI and PLP are so-called bicupines because they are formed of two cupin domains composed each of one barrel of four β sheets. These cupin domains are connected by an inter-domain consisting of an α helixes.

We decided to study the effect of the loss of the first cupin domain on the second one, which contains the active site. For this, each half of the 2 starting enzymes (Half PMI and Half PLP) made of one cupin domain which contains the active site were express and purified. The aim was then to evaluate the maintenance of the fold and the initial activity of the protein in the absence of the second domain.

In this chapter, we will describe the cloning by molecular biology techniques of the two half enzymes, half-PMI and half PLP and their characterization by SDS Page electrophoresis and Western blot.

The construction of the expression vector encoding for the His_6-tagged half PMI as well as for the His_6-tagged half PLP will first be described, followed by the cloning and expression of these proteins in *E. coli*.

After purifications on a Ni-NTA column, the characterization of the obtained proteins by SDS PAGE electrophoresis and Western blot will then be shown.

5.2. RESULTS AND DISCUSSION

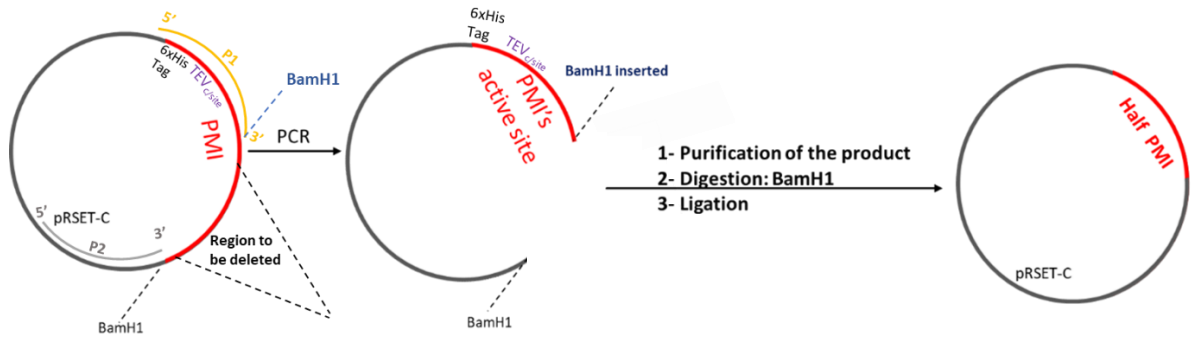
5.2.1. Half PMI protein

5.2.1.1. Construction of the expression vector for Half PMI

To construct the expression vector for Half PMI, we used the enzyme expression vector (pRSET-C vector) encoding for the PMI comprising a His_6 tag as well as a TEV protease cleavage site. This vector was described in Chapter 3 and was used to express both PMI and Chimera 1.

The target DNA sequence encoding for the His_6 tagged PMI's active site and His_6 tag flanking $TEV_{c/site}$ was examined and the suitable primers to be used to bind and allow its replication by PCR were designed. The designed primers facilitated the amplification of the first cupin domain of PMI, $TEV_{c/site}$, His_6 tag and the pRSET-C vector including BamH1 restriction site. Additionally, the primers allow the insertion of BamH1 restriction site at the other end (Scheme 22). This amplification resulted in the deletion of the DNA sequence encoding for the cupin domain that does not contain the active site. The amplified samples were analyzed by electrophoresis/Syber safe staining and purified.

The purified product was then digested by BamH1 restriction enzyme generating two sticky ends with unpaired bases at the end of the fragments. Subsequently, DpnI was added to the mix to catalyze the hydrolysis of the original PMI expression vector. Digestion was followed by ligation using instant sticky ligase master mix kit. The ligation mixture was used to transform E. coli (DH5alfa) strain. Transformed cells were suspended into LB medium supplemented with 100 ug/ml Ampicillin.



Scheme 22: Construction of the expression vector of Half PMI. P1: Forward primer; P2: Reverse primer.

In order to confirm the DNA sequence of the construct, a colony of the transformed and selected bacteria was grown and plasmids were extracted, purified and sequenced. They showed the expected sequence.

5.2.1.2. Expression and purification of the His_6-Half PMI

E. coli BL21(DE3) cells were used for expressing the protein. The purified and sequenced plasmid DNA was used to transform BL21(DE3) cells that were then selected on agarose/ampicillin plates.

A pre-inoculum from a single colony was grown in a starter culture in LB-ampicillin medium. TB supplemented with ampicillin was inoculated from the preinoculum at 1/100 and grown at 37 °C until the OD at 600 nm reached a value between 0.6 and 0.8. 0.1 mM IPTG was then added and incubation was continued at 20 °C for 18 h. Bacterial cells were pelleted and kept at -20 °C. To lyse the cells prior to the purification, the bacterial pellets were resuspended in the lysis buffer in the presence of benzonase and lysozyme. The mixture was then sonicated and centrifuged to eliminate all solids. The clear solution containing the His_6 tagged protein was purified on an Ni-NTA agarose column. The various eluted fractions were analyzed by SDS-PAGE followed by Coomassie blue staining. As shown in Figure 91, the elution fractions show different bands revealing a non-pure protein. However, the corresponding molecular weight of the Half PMI was expected to be about 40 kDa. No band with high intensity was observed in the elution fractions corresponding to this molecular weight. This result indicates that the protein is not well expressed.

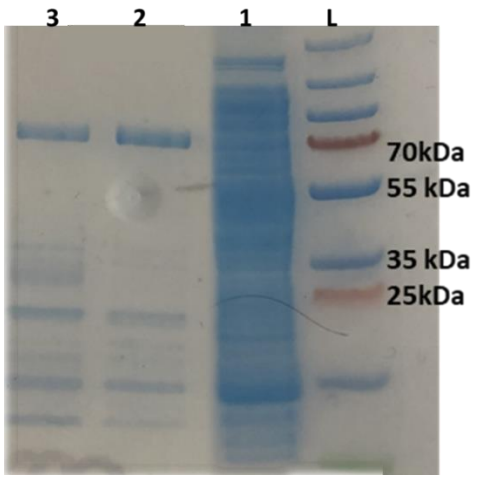


Figure 91: Coomassie blue stained SDS-PAGE electrophoresis of the collected fractions after the purification on an Ni-NTA agarose column of His_6 tagged Half PMI. L: Ladder; Lane 1: Washed fraction, Lanes 2 and 3: Elution fractions

To optimize the expression of the protein, another bacterial strain was used. Purified and sequenced plasmid DNA encoding for Half PMI was used to transform BL21(DE3) pLysS (see Chapter 2 PMI) that were then selected on agarose/ampicillin. The same procedure as above was followed to express and purify the protein on a Ni-NTA agarose column. As shown in Figure 92, the elution fraction shows different bands than those observed above: A band at a molecular weight of about 40 kDa corresponding to this protein was observed, other bands were also present thus the protein was not pure. Despite this attempt, the purification was not successful.

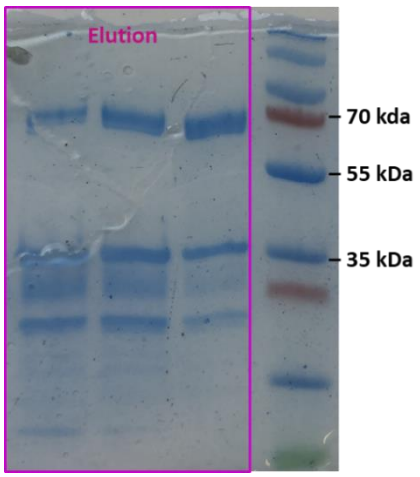


Figure 92: SDS-PAGE electrophoresis followed by Coomassie blue staining of the elution fractions collected after the purification on an Ni-NTA agarose column of His_6-Half PMI using the BL21 E.*coli* pLysS strain.

A western blot using anti His_6-tag antibody was done. Figure 93 shows three bands corresponding to the fractions eluted with buffer containing 300mM imidazole. This suggests that the elution fractions contained different His_6 tagged proteins: a band with high intensity corresponding to a protein with a molecular weight of about 40 kDa corresponds to the expected MW of His_6-Half PMI. This confirms that the protein was indeed His_6 tagged. Two other bands may correspond to other histidine rich proteins that could bind the Ni-NTA column and weakly bind to the His_6-tag antibody. Hence, folding problems were suspected and the low purification quality can be due to low quantity of soluble His_6-Half PMI that can be found in inclusion bodies.

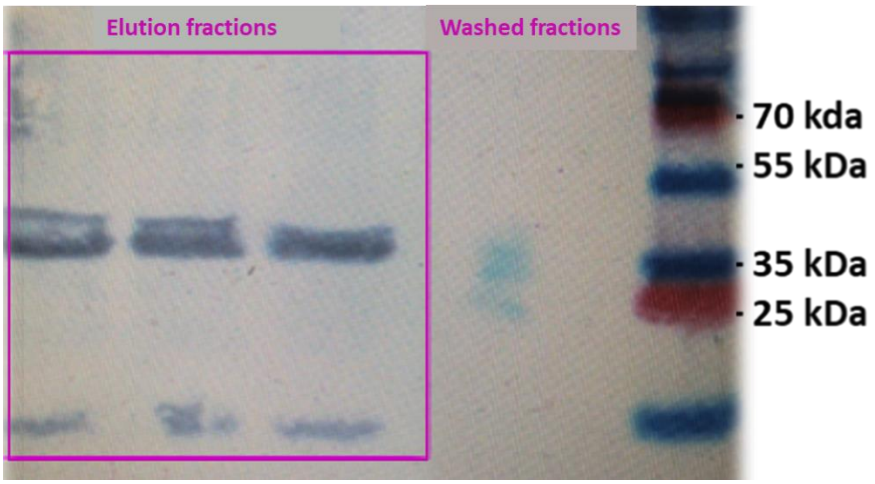


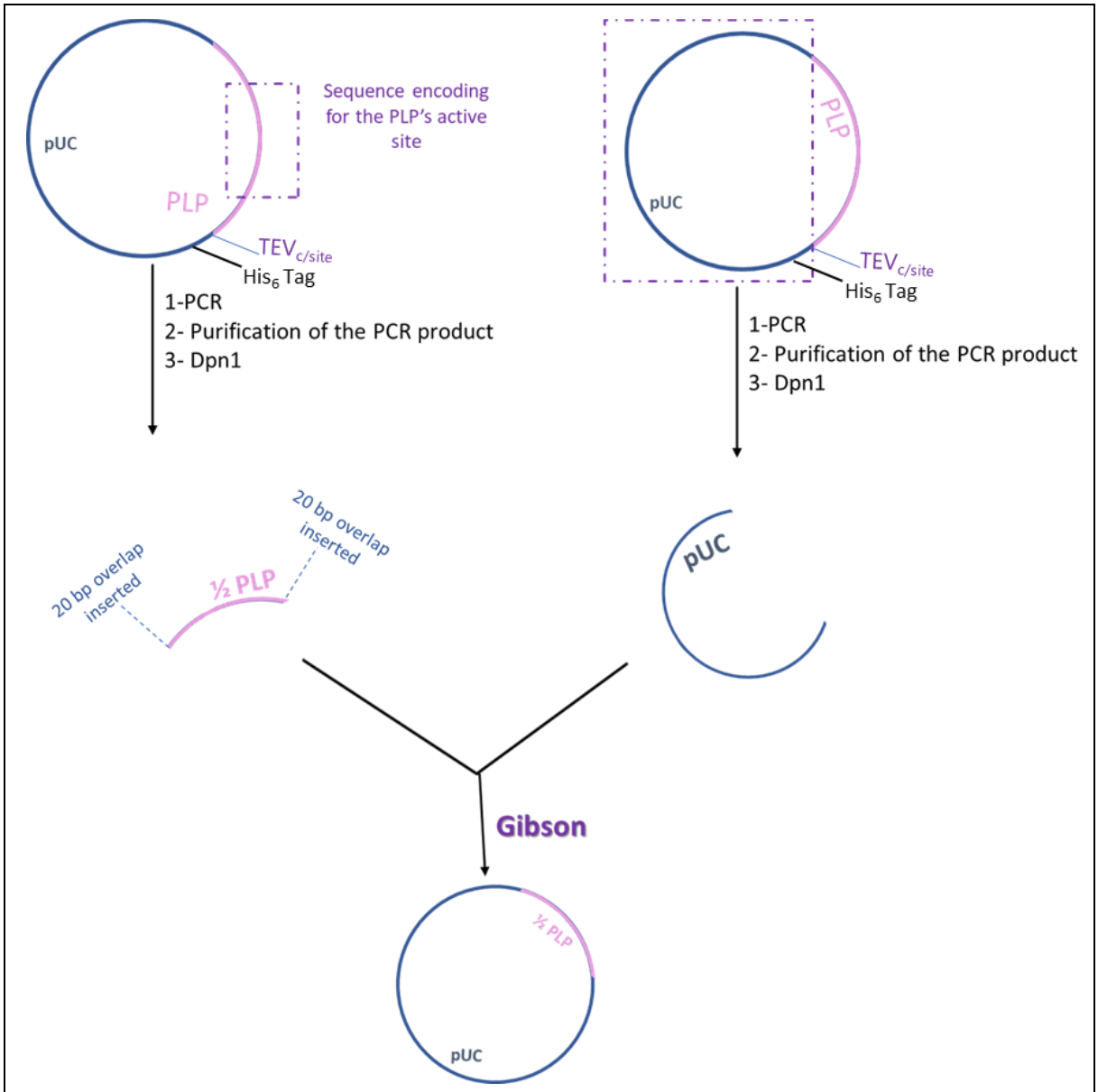
Figure 93: Western Blot image using anti His_6 antibody of the fraction eluted during the purification of the Half His_6 PMI protein.

5.2.2. Half PLP protein

5.2.2.1. Construction of the expression vector for Half PLP

Starting from the full PLP expression vector, the objective was to extract the sequence encoding for the active site-containing cupin domain of PLP and to incorporate the extracted sequence into PUC18 vector since this vector is the one used to express full PLP (scheme 23). Also, it was important to maintain the His_6 tag and the $TEV_{c/site}$ for purification on Ni-NTA column. For this, we proceeded by amplifying two fragments of the full PLP expression vector: One fragment contained the PLP's active site and the second fragment contained the pUC18 vector, the His_6 tag and the $TEV_{c/site}$. The two PCR products were then purified and digested by Dpn1. The two fragments were then

assembled using the Gibson assembly method as described in chapter 4. Transformation and selection resulted in colonies of bacteria that contained a plasmid formed from both fragments as demonstrated by specific PCR amplifications. To confirm the DNA sequence of the plasmid, and to examine the junctions between the fragments, sequencing was done. One selected colony was grown then plasmids were extracted, purified and sequenced. The expected sequence was obtained.



Scheme 23: Construction of the expression vector of the His_6 - Half PLP by the Gibson assembly method.

5.2.2.2. Expression and purification of the His_6 - Half PLP

To obtain substantial quantities of protein, the resulting plasmid was used to transform E. *coli* BL21(DE3) cells. As for the half PMI protein, a pre-inoculum from a single colony was grown in a starter LB-ampicillin culture. TB supplemented with ampicillin was inoculated from the pre-inoculum at 1/100 and grown at 37 °C until the OD at 600 nm reached a value between 0.6 and 0.8. 0.1 mM IPTG was then added and incubation was continued at 20 °C for 18 h. Bacterial cells were pelleted and kept at -20 °C. The bacterial pellets were resuspended in the presence of benzonase and lysozyme. The mixture was then sonicated and centrifuged (RCF=13000) to eliminate all solids. The clear solution containing the His_6 tagged protein was purified on an Ni-NTA agarose column. The various fractions eluted from the column were analyzed by SDS-PAGE followed by Coomassie blue staining. No bands were observed in the elution fractions (Data not shown). This can be due either to the low expression or to its expression in the inclusion bodies as it was observed in the case of full His_6-PLP (see chapter 4).

However, western blot using an anti His_6-tag antibody (Figure 94) showed a band with a good intensity corresponding to the protein with a molecular weight of about 11 kDa for the elution fractions. This confirmed that the protein was indeed His_6 tagged.

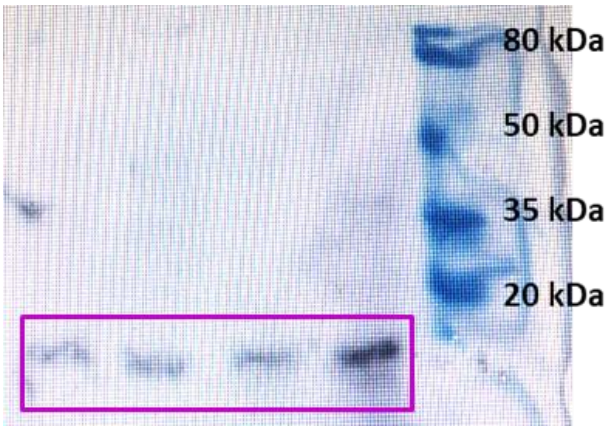


Figure 94: Western Blot image using anti His_6 antibody of the collected elution fractions of the first purification on an Ni-NTA column of the Half PLP protein.

5.3. CONCLUSION:

As we mentioned before, PLP and PMI are similar bicupins characterized by a structure formed of two beta barrel domains, each possessing one active site in one of their beta barrel domains.

Aiming at evaluating the maintenance of the initial activity of the protein in the single barrel of β sheets containing the active site without the presence of the second barrel of β sheets, we proceeded to the preparation of each half of the 2 starting enzymes (Half PMI and Half PLP) made of one barrel of β sheets which contains the metal in their active site.

The expression vectors containing the DNA encoding for the Half proteins were successfully constructed and the proteins were expressed in E.*coli* BL21. Western blot confirmed that both half proteins were indeed expressed but initial purifications were not satisfying. This can be due to folding problems for both Half proteins. As for Half PLP, it can be due to solubility problems deriving from the staring full PLP which was expressed in both supernatant and inclusion bodies (see chapter 5). However, even if the Half proteins are misfolded, the chimeric PMI-PLP could be well folded as it contains two cupin domains which might be necessary for the fold. This result highlights the advantage of developing chimeric proteins.

The optimization of the purifications should thereafter proceed by first checking the solubility of the half enzymes. For this, the produced half proteins will be analyzed on SDS PAGE gel electrophoresis before and after the cell lysis to verify their expression quantity and location.

CHAPTER 6

GENERAL CONCLUSION AND FUTURE WORKS

CHAPTER 6: GENERAL CONCLUSION AND FUTURE WORK

Biocatalysis using natural enzymes is mostly limited to natural substrates of the employed enzyme and to biotic reactions. Starting from synthetic substrates, artificial enzymes that are developed to catalyze chemist-invented reactions are successfully leading to pure enantiomers of industrially relevant products.[5] Metal complexes are the most efficient man-made catalysts and are also routinely used to form the catalytic cofactors of artificial metalloenzymes. The latter combine both the advantages of natural metalloenzymes such as stereoselective catalysis in aqueous solution, under mild conditions and the advantages of chemically synthesized metal complexes catalysts such as chemist-invented catalysis and compatibility with artificial substrates.[16]

Different strategies were described over the years to construct artificial metalloenzymes. But in this project, a new strategy seemed most appealing to us: **Development of chimeric metalloenzymes.** We prepared an artificial chimeric protein that possesses multiple active sites by the fusion of catalytic domains of two related enzymes. This chimeric enzyme would be used to catalyze reactions in cascade under ecofriendly conditions.

To prepare the chimeric enzyme, we relied on the cupin superfamily of proteins. Therefore, the chimeric enzyme combines the catalytic domains of two similar bicupin proteins: PMI and PLP, which should avoid denaturation and lead to a correct folding.

Prior to the preparation of the chimeric PMI-PLP metalloproteins we proceeded to express and purify each of the two starting enzymes i.e. PMI and PLP and to test their activities before and after exchanging the metal cations of their respective active site. We plan to compare these activities with those of the chimeric PMI-PLP metalloproteins.

PMI are metalloenzymes that catalyze the reversible isomerization of *D*-mannose-6-phosphate into *D*-fructose-6-phosphate their active site is located in a characteristic barrel of β-sheets.[13] Artificial enzymes formed by incorporating transition metals or their complexes in β-sheets are rare. Aiming at elucidating the role of such active site, we prepared and characterized an artificial metalloprotein based on PMI. The enzyme expression vector yielded PMI comprising a His_6 tag that we were able to cleave by introducing a TEV protease site using molecular biology techniques. The expression plasmid was digested and a site for specific hydrolysis by TEV protease was successfully introduced flanked by the sequence of the enzyme and that of the His_6 tag. After

purification on a Ni-NTA column, the enzyme was separated from the His_6 tag by digestion with TEV protease. TYCHO showed that the thermal denaturation of PMI occurred at 60°C. PMI contains one zinc(II) *per* enzyme molecule and the removal of zinc(II) from the active site was performed by incubation with 10^6 equiv. EDTA. To prepare the artificial metalloenzyme, copper(II) was added and PMI was successfully transformed into an artificial copper(II)-containing protein that was characterized by a relatively good dissociation constant (K_D = 3 µM). The EPR spectrum of the PMI-Cu complex confirmed the presence of a single copper in the binding site of the enzyme. The PMI-Cu complex did not catalyze Diels-Alder reaction most probably because the active site was buried within the enzyme and not accessible for the hallmark substrates of artificial metallo-Diels-Alderases. As perspectives, smaller Diels-Alder substrates will be used with this protein as catalyst. A peroxidase activity is also envisioned for the artificial PMI-Cu protein.

PLP is a metalloenzyme that resembles PMI. It belongs to the cupin superfamily of enzymes, and its active site is located in the characteristic barrel of β-sheets. The biological role of PLP remains unclear. A previous study has shown that PLP could be transformed into an artificial metalloenzyme by substituting its iron cation by a divalent metal cations such as Cu^{2+}.[14] Aiming at the preparation of a chimeric PMI-PLP protein, we first prepared and characterized an artificial metalloprotein based on PLP. The enzyme expression vector yielded PLP comprising a His_6 tag co-expressed with a second protein formed of a portion of LacZ fused to His_6-PLP. The coexpression was confirmed by an anti His_6 western blot which revealed two bands instead of a single band. To avoid this co-expression, the expression plasmid was edited by deleting the remaining portion of LacZ. The expression plasmid was digested and a site for specific hydrolysis by TEV protease was successfully introduced flanked by the sequences of the enzyme and that of the His_6-tag. Purified and sequenced plasmid DNA containing the insert sequence was used to transform BL21(DE3) for expressing the protein. After expression and purification on an Ni-NTA column, the enzyme was separated from the His_6 tag by digestion with TEV protease. Following the digestion, the TEV protease with its His_6 tag was easily removed from the cleavage reaction by purification on a Ni-NTA affinity column. The produced PLP contains one Fe(II) cation *per* enzyme molecule, therefore the removal of Fe(II) from the active site was performed by incubating the protein with 10^3 equiv. EDTA. To prepare the artificial metalloprotein, Ni^{2+}, Zn^{2+} and Cu^{2+} were added to the iron free protein. A relatively good dissociation constant (K_D = 0.5 µM) was measured by fluorescence quenching upon addition of increasing amounts of copper sulfate. The EPR spectrum of the PLP-Cu complex confirmed the presence of a single copper species. TYCHO results showed that the thermal

denaturation occured at 75°C and that high concentrations of EDTA and of copper sulfate could alter the folding of PLP. Hence, PLP was successfully transformed into an artificial copper(II)-containing protein. Quercetin is relatively bigger than the azachalcone and cyclopentadiene and since copper-containing PLP catalyzed quercetin oxidation we expect that it will also catalyze the Diels-Alder cycloaddition after future works.

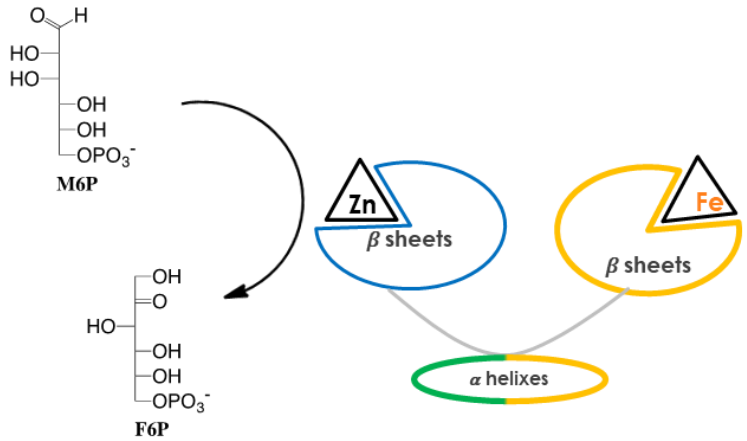
Since the two starting enzymes were expressed, purified, transformed into artificial metalloproteins and characterized by different techniques, we proceeded to develop an artificial chimeric metalloenzyme that possesses multiple active sites by the fusion of catalytic domains of these two related enzymes. First, the strategy we followed to prepare the PMI-PLP bicupin chimera was based on structural modeling studies and on the determination of the linking point between the two catalytic domains of PMI and PLP. *elNémo* was used to simulate the low frequency normal modes of vibration of PMI and of PLP. The movements of all atoms/domains in PMI and PLP were thus studied and visualized. The 1st and the 2nd mode of vibration of PMI show that the two cupin domains vibrate independently. The 3rd normal mode of vibration originates at the center of the linker of the cupin domains thus the association of the two different cupin domains can occur by fusion at the center of their interdomain linkers. The sequences encoding for the PMI active site and the PLP active site were cloned by PCR. PCR products were then purified, annealed and ligated via the Gibson Assembly method. The constructed expression vector containing the DNA encoding for the PMI-PLP chimera was used for expression in *E.coli* BL21. Initial purification was not successful as the chimera remained contaminated. Since molecular modelling using MODELLER revealed a possible clash between an α helix in the variable region deriving from PMI and the cupin barrel deriving from PLP, a second chimeric enzyme, Chimera 2, was constructed by replacing the PMI loop responsible for the clash by the analog PLP loop which is much smaller. Chimera 2 was also expressed but similarly to chimera 1 the purification here failed again. Indeed, the proteins were found mainly in the insoluble fractions, most probably in inclusion bodies. The co-expression of chaperones proteins with chimera 2 failed aiding solubilization.

Aiming at evaluating the conservation of the initial activity of a bicupin by a single cupin domain containing the active site thus in the absence of the second cupin domain, we proceeded to the preparation of each half of the two starting enzymes (Half PMI and Half PLP) made of one barrel of β sheets which contain the native metal in their active site. Hence, the expression vectors containing the DNA encoding for the Half proteins were constructed and the proteins were expressed in E.*coli* BL21. Western blot confirmed that both half proteins were indeed expressed but initial purifications

were also not satisfactory. We suspect that these proteins were also produced in inclusion bodies, which remains to be verified by analysis on SDS PAGE gel electrophoresis before and after the cell lysis. This can be due to folding problems for both Half proteins since they are formed of only one cupin domain. As for Half PLP, it can be due to solubility problems deriving from the staring full PLP which was expressed in both supernatant and inclusion bodies.

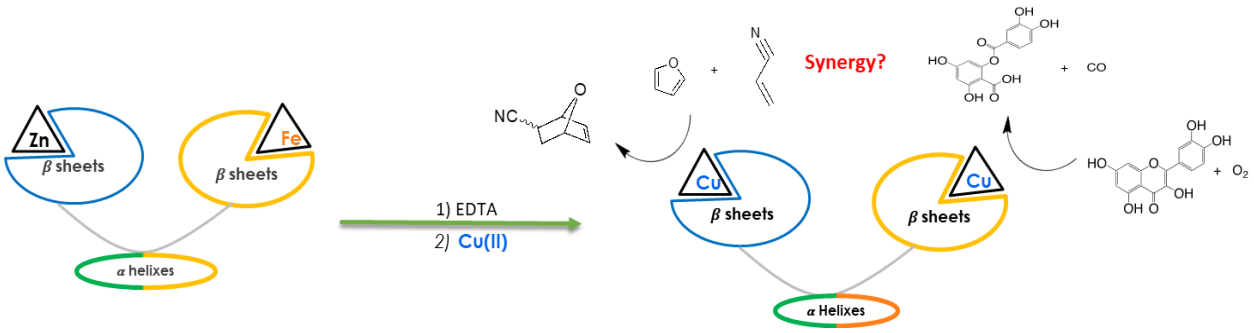
For Chimeric proteins, first, alternative methods of purification such as the use of denaturing agents and solubilization should be examined. Hence, the protein should be denatured which will require its subsequent refolding. Second, another possible solution leading to correctly folded chimera could be to reverse the order of the cupin domains. As mentioned earlier, the order of insertion influences the folding.[28] Inspired by natural and artificial enzymatic catalysis, the aim for preparing the chimeric PMI-PLP enzyme is the catalysis of reactions in cascade. Obtaining a correctly folded chimeric bicupin enzyme remains difficult, then mutating the metal-free domain of a bi-cupin is another option to achieve our aim. Indeed, inspired by the metal binding cupin domain of PLP, site directed mutagenesis may be used to create a metal binding site in the metal free cupin domain of PMI for example. This less challenging strategy could nevertheless be more successful to reach our goals. In fact, inspired by the different metal binding site of cupins, site directed mutagenesis can be used to adapt a binding site for a specific metal cation for a sought after activity. This could facilitate the preparation of enzymes with different activities while maintaining the same core of the bicupin domains.

Once such protein obtained, a starting point for analyzing catalysis will be to assay the isomerase activity deriving from the parent PMI so that the impact of the Fe(III) site on this reaction can be identified (scheme 24).



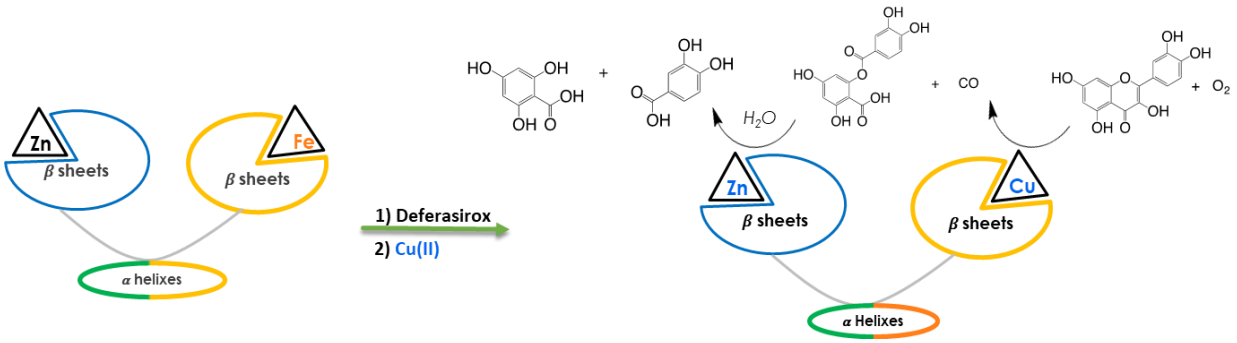
Scheme 24: Reversible isomerization of D-mannose-6-phosphate (M6P) to D-fructose-6-phosphate (F6P) catalyzed by the chimera.

Thereafter, the reconstitution of the two active sites Cu(II) will be considered to lead to the formation of an enzyme with two activities: Diels-Alderase and quercetinase, for example. Studies of these activities identify the potential synergy due to the presence of two Cu(II) sites in the chimera (scheme 25).



Scheme 25: Possible synergy of Diels-Alder cycloaddition and quercetinase activities catalyzed by the Cu(II)-Cu(II) containing chimera.

Finally, the reconstitution of the artificial enzyme by a different metal in each active site may be achieved by removing only the Fe(III) cation using a ligand that possesses high affinity for Fe(III) but low affinity for Zn(II) such as deferasirox (a drug that is 20 times more selective for Fe(III) than for Zn(II)). The reconstitution of the enzyme by a Cu(II) containing active site should provide the quercetinase activity while retaining the Zn(II) containing active site and its isomerase activity. These two very different active sites can then be used to catalyze different reactions, as well as reactions in cascade such as quercetin oxidation catalyzed by the artificial copper site of PMI, followed by the hydrolysis of the product catalyzed by the zinc of PMI (scheme 26).



Scheme 26: Two reactions in cascade catalyzed by heterobinuclear Cu(II) - Zn(II) containing chimera.

MATERIALS AND METHODS

I. CONSTRUCTION OF THE EXPRESSION VECTOR OF THE PROTEINS:

- **Plasmid purification:**

  - **For PMI :**

BL21 (DE3) *E. coli* cells incorporating a version of the pRSET-C expression vector (total 8 kb) containing the expression system of the His_6 tag PMI and the ampicillin resistance gene (see annex 1) were purchased from Sigma. The bacterial strain was first thawed and then grown at 37°C for 18 hours on plates coated with LB-agar (Lysogeny Broth-agar, Sigma) medium supplemented with ampicillin (100 µg/mL, Sigma).

A pre-inoculum from a single colony was grown in a 3 mL starter LB-Ampicillin culture at 37 °C (100 µg/mL). When the OD at 600nm (Cary 50, Varian) reached a value around 4 (after 6 hours), LB supplemented with 100 µg/mL ampicillin was inoculated with the preinocculum (final dilution 1/1000th) and the cells were grown at 37°C overnight (225 rpm).

The plasmid DNA was then purified by using the Xtra MIDI Kit (MACHEREY-NAGEL). The culture was centrifuged at 6000 g for 15 minutes, the supernatant was then discarded and the pellet was resuspended in resuspension RES buffer. Lysis buffer (sodium hydroxide<2%, sodium dodecylsulfate) was added to the suspension and incubated for 5 minutes at room temperature after inverting the tube 5 times. Neutralization Buffer NEU (acetic buffer) was added to the suspension, the lysate was mixed gently by inverting the tube until the blue colored samples turned completely colorless, and it was then incubated on ice for 5 minutes. The NucleoBond® Xtra Column was equilibrated as recommended, the lysate was then loaded onto the column, washed 3 times using the washing buffer (diluted ethanol <20%, potassium chloride) and eluted by adding ELU-EF. The plasmid DNA was precipitated by adding isopropanol. After centrifugation at 15000 g for 40 min at 4 °C, the supernatant was discarded and 70 % ethanol was added to the pellet. After centrifugation at 15000 g for 15 min at 4 °C, the supernatant was discarded and the pellets were dissolved in an appropriate volume of sterile nuclease-free water.

DNA concentration (3.2 µg/µL) and purity (A260 = 64,963; A280 = 34,621; A260/A280 = 1,88) were determined by measuring the UV-vis absorption spectra of the samples on a nanodrop spectrophotometer (Thermo Scientific) apparatus.

- **For PLP :**

The plasmid which provided ampicillin resistance and the expression vector for the His_6-PLP was requested from Kataoka's team in Japan (see annex 2).[14]

The lyophilized plasmid resuspended in sterile water was used to transform *E. coli* (DH5alfa) strain (Sigma). The cells were transferred from -70°C to wet ice and thawed for 10 minutes. The DNA was diluted to 10 ng/µL in water and then added directly to 100 µL competent cells in 14-ml BD Falcon polypropylene round bottom tube, that were mixed by gentle tapping, and incubated on ice for 30 minutes. The cells were transferred to a 42° C water bath for exactly 45 seconds and then back to ice for 2 minutes. 450 µl of SOC medium (Sigma) was added and the cells were incubated on a shaker incubator (225 rpm) at 37°C for 1 hour. Transformed cells were suspended into LB agar plates with ampicillin (75 µg/mL) and were incubated at 37°C overnight (225 rpm). A pre-inoculum from a single colony was grown in a 5 mL starter culture in LB-Ampicillin at 37 °C (75 µg/mL).

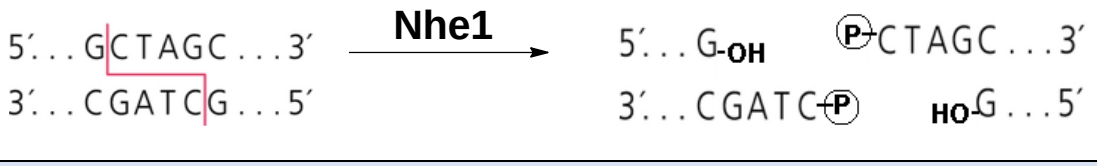
The plasmid DNA was then purified by using NucleoSpin Plasmid transfection-grade (MACHEREY-NAGEL): The culture was centrifuged at 11000 g for 1 minute, the supernatant was discarded, and the pellet was resuspended in Resuspension A1 buffer. Lysis buffer was added to the suspension and incubated for 5 minutes at room temperature by inverting the tube 5 times. Neutralization Buffer was added to the suspension, the lysate was mixed gently by inverting the tube until the blue colored samples turned completely colorless, and it was then centrifuged 10 minutes at 13000g. The clear supernatant was then loaded onto the given kit column and washed 3 times. The DNA was then eluted by adding 25 µL of sterile water.

DNA concentration (70 ng/µL) and purity (A260 = 64,7; A280 = 33.045; A260/A280 = 1,88) were determined by measuring the UV-vis absorption spectra of samples on a nanodrop spectrophotometer (Thermo Scientific) apparatus.

- **Plasmid digestion :**

  - **For PMI :**

Our plasmid was digested by using the restriction enzyme Nhe1 because its restriction site was located between the PMI sequence and the His_6 tag sequence. The Nhe1 specific restriction site and digestion products are in shown in scheme 27.



Scheme 27: Sequence of Nhe1 recognition site and its restriction products once DNA has been digested.

2 µL of the extracted plasmid (1.2 µg/µL) in water was mixed to 4 µL 10x tango buffer (33 mM Tris-acetate (pH 7.9 at 37°C), 10 mM magnesium acetate, 66 mM potassium acetate, 0.1 mg/ml BSA) (ThermoFisher), 2 µL Nhe1 (10 U/µL, ThermoFisher) and 32 µL water. The mixture was then incubated overnight at 37°C.

To inactivate the enzyme, the next day, the mixture was heated at 65°C for 25 minutes.

  - **For Chimera 2 :**

The pET-28b vector and the synthetic DNA sequence encoding for the chimera 2 were double digested with BamH1 (New England Biolab: NEB) and Pst1 (NEB). For this, 1µg DNA was added to 1µL BamH1, 1 µL Pst1, 5 µL Cutsmart buffer (50 mM Potassium Acetate. 10 mM Magnesium Acetate. 100 µg/ml BSA. pH 7.9) (NEB) and 33 µL sterile water. The mixtures were then incubated 4 hours at 37°C.

- **Agarose gel electrophoresis and ethidium bromide staining:**

The 1% agarose gel was prepared by dissolving agarose powder (1 g, Sigma) in 1x Tris Borate EDTA (TBE) buffer, pH = 8 (100 mL, Sigma) supplemented with ETBr (0.2-0.5 µg/mL, Sigma) and a comb was placed to create wells for loading samples. Once the gel was set, the comb was removed, leaving wells wherein DNA molecular weight ladder and samples of circular plasmid (3.2 µg/µL) and of the digesting mixture (1.2 µg/µL) were loaded mixed with low molecular weight dye (sigma).

After 30 minutes of migration at a voltage of 100 Volts, the dye reached the edge of the gel. The electric current was turned off and the gel was visualized with an ultraviolet (UV) DNA transilluminator (Biorad).

For PMI, a DNA band of 8 kb corresponding to the digested plasmid was cutout and purified.

For the construction of the chimera 2, a DNA band of 6379 pb corresponding to the digested pET-28B plasmid was cutout and purified. In addition, a DNA band of 1330 pb corresponding to the digested sequence of the Chimera 2 was also cutout and purified.

- **DNA extraction from agarose gels:**

The extraction from agarose gel was performed using the MACHERY-NAGEL kit (PCR clean up and gel extraction). 100 mg of gel containing the linear digested plasmid were added to 200 µL buffer NT1 (guanidine thiocyanate) and the sample was vortexed briefly until the agarose gel was completely dissolved (5–10 min at 50°C). The sample was loaded onto the provided NucleoBond® Xtra Column and centrifuged for 30 s at 11000 g. Buffer NT3 was then added, the sample centrifuged for 30 s at 11000 g. The flow through was discarded. The DNA was then eluted from the column by adding Buffer NE, incubated at room temperature (18–25 °C) for 1 min and then centrifuged for 1 min at 11000 g.

For the PMI, the linear plasmid concentration (33 ng/µl) and purity were also measured using the nanodrop spectrophotometer (Thermo Scientific) apparatus (A260 = 0.669; A280 = 0.366; A260/A280 = 1.83).

For the Chimera 2, the linear digested plasmid was eluted by adding 20 µL of sterile water. The DNA concentration (24 ng/µL) was determined by measuring the UV-vis absorption spectra of samples on a nanodrop spectrophotometer apparatus (ThermoFisher). On the other hand, 29 ng/µL DNA encoding for the synthetic sequence of the chimera 2 was eluted.

- **Modification of the plasmid:**

  - **For PMI:**

12 µl of the purified vector (33 ng/µl) was mixed with 4 µl of DNA insert (10 molar equivalents), 2 µl 10x T4 DNA ligase buffer (500 mM Tris, 100 mM MgCl2, 50 mM DTT,

10 mM ATP, pH=7.6, Sigma), 2 µl T4 ligase (1U/µL, Sigma), following the providers (Sigma) procedure. The mixture was then incubated overnight at 8°C.

The ligation mixture was used to transform *E. coli* (DH5alfa) strain (Sigma). The cells were transferred from -70 °C to wet ice and thawed for 10 minutes. 50 µL of the competent cells were aliquoted into 14-ml BD falcon polypropylene round bottom tube. 5 µL of the ligation mixture was added directly to competent cells, mixed by gentle tapping and was incubated in an ice bath for 30 minutes. The cells were transferred in a 45 °C water bath for exactly 45 seconds and then back in ice for 2 minutes. 960 µl of SOC medium (Sigma) were added and the cells were placed on a shaker incubator (225 rpm) at 37 °C for 1 hour. 100 µl of the mixture containing the transformed cells were coated onto LB agar plates with ampicillin (100 µg/mL) and the plates were incubated at 37 °C overnight (225 rpm).

- **For chimera 2:**

2.5 µl of the purified vector (24 ng/µl) was mixed with 2.5 µl of DNA insert (29 ng/µl) and 5 µl instant sticky ligase (NEB), following the providers (NEB) procedure. The mixture was then incubated 15 minutes in an ice bath.

The ligation mixture was added directly to 50 µl competent cells in 14-ml BD Falcon polypropylene round bottom tubes, mixed by gentle tapping and incubated in ice for 30 minutes. The cells were transferred in a 42 °C water bath for exactly 45 seconds and then back to ice for 2 minutes. 650 µl of SOC medium (NEB) were added and the resulting mixture was placed on a shaker incubator (225 rpm) at 37 °C for 1 hour. 100 µl of the mixture containing the transformed cells were coated onto LB agar plates with kanamycin (50 µg/mL) and incubated at 37 °C overnight (225 rpm).

- **Polymerase chain reaction (PCR):**

  - **For PMI :**

A PCR mixture was prepared using the Thermofisher Taq polymerase starter kit, by adding 2.5 µl Taq polymerase reaction Buffer, 4 µl $MgCl_2$ (C_i= 25 mM), 2.5 µl dNTPs (C_i=2mM), 25 µM primers, 1 µl nuclease free Taq DNA Polymerase (1U/µl), 15 µl water and finally a small amount of each colony on the tip of a wooden toothpick was added.

The reaction was gently mixed and the PCR tubes (50 µl) were transferred from ice to the PCR machine. The PCR was carried out in a Perkin-Elmer Thermocycler starting with

two minutes at 95 °C for Taq activation followed by 40 2-minutes cycles involving 30 seconds at 95 °C for DNA denaturation, 30 seconds at 48 °C for hybridization and one minute at 72 °C for extension.

The presence of the PCR amplicon and the size of the product were determined by electrophoresis alongside a DNA size marker on a precast 1.2% agarose gel.

- **For PLP:**

- Deletion of a small region of the lacZ promoter:

25µL of the PCR mixture was prepared using the Q5 site directed mutagenesis kit by adding 12.5 µL 2x Master Mix Q5 hot start kit (NEB) containing the polymerase and its buffer, 2.5 µL primers (final concentration 10 µM), 9 µL nuclease free water and 1 µL of the plasmid (10 pg - 1 µg).

The PCR was carried out using two primers (P1: 5' ATG GCC CAA CGG GAA ATT CTA TCC ATC ACG ACC GGC 3' and P2: 5' AGC TFT TTC CTG TGT GAA ATT GTT ATC CGC TCA CAA TTC CAC ACA ACA TAC GAG 3') in an applied biosystem thermocycler (StepOne Plus). It started with 30 seconds at 98 °C for Q5 Hot start high fidelity polymerase activation, followed by 25 3-minutes cycles involving 10 seconds at 98 °C for DNA denaturation, 20 seconds at 65 °C for hybridization, 30 seconds at 72 °C for extension and two minutes at 72 °C for final extension. The presence of the PCR amplicon and the size of the product were determined by electrophoresis alongside a DNA size marker on a precast 1.2% agarose gel.

Then, the PCR product was treated by Kinase, Ligase and DPN1 enzymes (KLD). For this, 1 µL of the PCR product was mixed with 1 µL KLD enzyme mix 10x, 5 µL 1X KLD reaction buffer 2x and 3 µL nuclease free water.

- TEV Protease site's insertion:

The TEV protease site (5' GAA AAC CTG TAT TTT CAG GGC 3') was inserted by PCR mutagenesis using two primers (P1: 5' TTT TCA GGG TCA TCA TCA TCA TCA TCA TTG AGG ATC CT 3' and P2: 5' TAC AGG TTT TCG GCC AGT GTC CCA TCG CGA AAA TC 3').

The PCR mixture was prepared using the Q5 site directed mutagenesis kit by adding 12.5 µL 2x Master Mix Q5 hot start kit (NEB) containing the polymerase and its buffer, 2.5 µL primers (final concentration 10 µM), 9 µL nuclease free water and 1 µL of the plasmid (10 pg - 1 µg). The PCR was carried out in an applied biosystem thermocycler (StepOne Plus). It started with 30 seconds at 98°C for Taq polymerase activation

followed by 25 4,5-minutes cycles including 10 seconds at 98°C for DNA denaturation, 20 seconds at 60°C for hybridation, 2 minutes at 72°C for extension and 2 minutes at 72 °C for final extension.

The presence of the PCR amplicon and the size of the product were determined by electrophoresis alongside a DNA size marker on a precast 1.2% agarose gel.

- **For Chimera 1:**

The pRSET-C vector for we previously used for expressing PMI was chosen as a vector (see annex 3). The suitable sequences of PMI and PLP were amplified. The PCR mixtures were prepared using the Q5 site directed mutagenesis kit by adding 12.5 µL 2x Master Mix Q5 hot start kit (NEB) containing the polymerase and its buffer, 2.5 µL primers (final concentration 10 µM), 9 µL nuclease free water and 1 µL of the plasmid (10 pg - 1 µg).

Each PCR was carried out in an applied biosystem thermocycler (StepOne Plus) using two primers: for PMI amplification: P1: 5′ CGA CTT CGC ATA AGG ATC CGA CCT CGA GAT CTG 3′ and P2: 5′ TAG GCC GGC CCA TTT TCT GTT TTT CAA CGC TTT CGT AGC T 3′, for PLP amplification: (P1: 5′ ACA GAA AAT GGG CCG GCC TAC ATC CG 3′ and P2: 5′ CG ATC CTT ATG CGA AGT CGC GGT AAC C 3′). Each PCR started with 30 seconds at 98°C for Taq polymerase activation. Then for PMI, by 25 4-minutes and 50-seconds cycles were performed including 10 seconds at 98°C for DNA denaturation, 20 seconds at 66°C for hybridation, 140 s at 72°C for PMI for extension and two minutes at 72 °C for primer final extension. For PLP, 25 2-minutes and 45-seconds cycles were performed including 10 seconds at 98°C for DNA denaturation, 20 seconds at 64°C for hybridation, 15 s at 72°C for PLP for extension and two minutes at 72 °C for primer final extension.

PCR products were purified and treated with Dpn1(Thermofisher, 10 u/ µL). 1 µg of eluted DNA in sterile water was mixed with 1µL of Dpn1 solution, 2 µL of 2x Tango buffer, the solution was completed to 20 µL with nuclease free water. The mixture was incubated 16 h at 37 °C after which the enzyme was inactivated at 80°C for 20 min. Then, the digested PCR products (1 µl PMI and 3 µl PLP) were assembled using the NEBuilder Hifi DNA assembly reaction kit by adding 10 µl Hifi DNA assembly Master Mix and 6 µl nuclease free water for 20 minutes at 50°C before transformation.

- **For Chimera 2:**

The synthetic sequence encoding for the chimera 2 was amplified using two primers:

5′ GCG CGG ATC CGG GTA TCA GAA TTA TGA TTG 3′ and 5′ GCG CCT GCA GTC AAG CAA AGT CCC GAT AAC CC 3′.

The PCR mixture was prepared using the Q5 site directed mutagenesis kit by adding 12.5 µL 2x Master Mix Q5 hot start kit (NEB) containing the polymerase and its buffer, 2.5 µL primers (final concentration 10 µM), 9 µL nuclease free water and 1 µL of the plasmid (10 pg - 1 µg). The PCR was carried out in an applied biosystem thermocycler (StepOne Plus). It started with 30 seconds at 98°C for Taq polymerase activation, followed by 25 4,5-minutes cycles including 10 seconds at 98°C for DNA denaturation, 20 seconds at 65°C for hybridization, 2 minutes at 72°C for extension and two minutes at 72 °C for primer final extension.

The presence of the PCR amplicon and the size of the product were determined by electrophoresis alongside a DNA size marker on a precast 1.2% agarose gel.

- **For Half PMI :**

The sequence was amplified using the Q5 HiFi PCR NEB kit. The DNA sequence of Half PMI was amplified using two primers:

P1 : 5′ CGC GGA TCC CAT TTT CTG TTT TTC AAC GCT TTC GT 3′ and P2 : 5′ CGC GGA TCC GAC TCG AGA TCT GCA GCT GGT A 3′.

The PCR was carried out in a applied biosystem thermocycler. It started with 30 seconds at 98°C for Taq polymerase activation, followed by 30 4-minutes and 50-seconds cycles including 10 seconds at 98°C for DNA denaturation, 20 seconds at 66°C for hybridization, 140 seconds at 72°C for extension and two minutes at 72 °C for primer final extension. Then, the PCR product was purified using the Macherey-Nagel kit. Concentration and purity were assayed on nanodrop spectrophotometer (Thermo Scientific) and the concentration was equal to 124 ng/µl.

8.1 µl of the purified product was then digested by the 1 µl BamH1 restriction enzyme (NEB) in 5 µl CutSmart Buffer and 35.9 µl sterile water. The optimal digestion time was empirically found to be 1 h. Digestion was stopped thermally by leaving the digestion mixture at 60°C for 20 minutes. It was followed by ligation using the instant sticky ligase master mix kit (NEB). 1.5 µl of the purified product (55 ng/µl) was mixed with 5 µl instant ligase mix (NEB) and 3.5 µl sterile water, following the providers (NEB) procedure. The mixture was then incubated 15 minutes in ice.

5 µl of the ligation mixture was added directly to 50 µl competent cells (NEB) in 14-ml BD Falcon polypropylene round bottom tube, mixed by gentle tapping and incubated

in ice for 30 minutes. The cells were then transferred into a 42° C water bath for exactly 45 seconds and then back into ice for 2 minutes. 650 µl of SOC medium (NEB) were added and the mixture were incubated in a shaker incubator (225 rpm) at 37°C for 1 hour. 100 µl transformed cells were coated on LB agar plates with ampicillin (100 µg/mL) and the plates were incubated at 37°C overnight (225 rpm).

- **For Half PLP:**

The sequence was amplified using the Q5 HiFi PCR NEB kit. The DNA sequence of Half PLP was amplified using four primers:

- P1: 5′ AGC TAT GGG CCG GCC TAC ATC CGA TG 3′ and P2: 5′ CAG GTT TTC TGC GAA GTC GCG GTA ACC CG 3′ for the sequence encoding for the PLP's active site.

- P3: 5′ ACT TCG CAG AAA ACC TGT ATT TTC AGA GTC ATC ATC ATC ATC ACA TTG AGG 3′ and P4: 5′ GGC CGG CCC ATA GCT GTT TCC TGT GTG AAA TTG AAA TTG TTA TCC GC 3′ for the sequence encoding for the pUC18 vector.

Each PCR was carried out in an applied biosystem thermocycler (StepOne Plus). It started with 30 seconds at 98°C for Taq polymerase activation, followed:

- For PMI, by 25 5-minutes and 10-seconds cycles, including 10 seconds at 98°C for denaturation, 20 seconds at 66°C (PMI's active site) and 20 seconds at 64°C (pUC18 vector) for hybridization, 140 s at 72°C for extension and two minutes at 72 °C for primer final extension.

- For PLP by 25 3-minutes and 5-seconds cycles including 10 seconds at 98°C for denaturation, 20 seconds at 66°C (PLP's active site) and 20 seconds at 64°C (pUC vector) for hybridization, 15 s at 72°C for extension and two minutes at 72 °C for primer final extension.

PCR products were purified using the Macherey-Nagel kit. Concentration and purity were assayed on nanodrop spectrophotometer (Thermo Scientific) and treated with Dpn1 enzyme (Thermofisher, 10 U/µL). 1 µg of DNA was mixed with 1µL of Dpn1 solution, 2 µL of 2x Tango buffer, the solution was completed to 20 µL with nuclease free water. The mixture was incubated 16 h at 37 °C and then the enzyme was inactivated at 80°C for 20 min. Then, the digested PCR products were assembled using the (NEB) HiFi DNA assembly reaction kit by adding 10 µl Hifi DNA assembly Master Mix and 6 µl nuclease free water for 20 minutes at 50°C before transformation.

- **Sequencing:**

For PMI, the sequencing was performed on purified plasmids at Eurofins using the Sanger method, in a single direction, starting from the expression T7 promotor. One direction sequencing was sufficient to confirm the PCR results since the remainder of the protein sequence was not modified. The single direction sequencing of PLP was performed following the same procedure, starting from the expression M13 Forward, which was sufficient to confirm the PCR results. The Sequencing of the chimeric enzymes was also performed on purified plasmids at Eurofins, using the Sanger method but in two directions using either T7 promotor or T7 Term.

- **Transformation of bacteria:**

To obtain large quantities of protein, the modified plasmid containing the DNA encoding for PLP was used to transform *E. coli* BL21(DE3) strain (Sigma-Aldrich), while for the modified plasmid containing the DNA encoding for PMI, *E. coli* BL21(DE3) pLysS strain was used. The transformation was performed according to the same procedure as that described above.

II. EXPRESSION AND PURIFICATION OF THE PROTEINS:

- TEV protease

Tobacco Etch Virus (TEV) proteases recognizes the seven-amino-acid sequence Glu-Asn-Leu-Tyr-Phe-Gln-Gly that it cleaves between Gln and Gly with high specificity. In this project, two TEV proteases have been used: His_6 tagged MBP-TEV protease that we have purified in our laboratory following the protocol described in the literature[108] and AcTEV Protease (His_6-AcTEV) which is commercial and can be purchased from Sigma. AcTEV Protease, an enhanced form of (TEV) protease, is highly site-specific and active, and it is more stable than native TEV protease.

First, the pYFJ16 plasmid which provided ampicillin resistance and the expression vector for the His_6 tagged MBP-TEV protease (His_6-Maltose Binding Protein (MBP)-TEV) was obtained in an E. *coli* DH5 alpha bacterial stab from Addgene. A subculturing was made on plates coated with LB medium supplemented with 75 µg/mL ampicillin and cells were grown overnight. Then, the plasmid was extracted using the MACHEREY-NAGEL kit as described above and its sequence was verified by sequencing. E. *coli* BL21-codon plus competent cells were then transformed with the sequenced plasmid and were grown overnight in 1L TB medium supplemented with 75 µg/ml ampicillin at 37°C, with shaking (200 rpm). The expression of the recombinant protein was induced with 0.1 mM IPTG. The purification of His_6-MBP-TEV was performed on a Ni-NTA affinity column following the protocol described below. Starting from a 800 mL culture, 60 mL of 70 µM TEV protease were obtained.

- Protein purification:

For PMI and PLP, a pre-inoculum from a single colony was grown in a starter culture in 20 mL LB supplemented with ampicillin (75 µg/mL), at 37 °C for 18 hours. 2L TB (Terrific Broth) supplemented with ampicillin (35 µg/mL) was inoculated with the preinocculum ($1/100^{th}$ dilution) and grown at 37 °C until the OD_{600} reached a value between 0.6 and 0.8 (after 3 hours). At this stage, a final concentration of 0.6 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added and incubation was continued at 20 °C for 18 h. Bacterial cells were pelleted by centrifugation and resuspended in lysis buffer (50mL Lysis buffer for 1L bacterial culture) (50 mM Tris HCl, 300 mM NaCl, 10% Glycerol, 1% Triton, 5 mM imidazole, pH = 8). After adding Lysozyme (10 mg/mL) and benzonase (3 µL benzonase for 30 mL lysis buffer) the sample became viscous. The sample was sonicated on ice (Amplitude = 50 %, 12 minutes, 15 s ON, 30 s OFF) and then centrifuged at 13000 g for 45 minutes at 4°C. The supernatant was collected for purification by column chromatography on an AKTA purification system. The

supernatant was applied onto a Ni-NTA column (HisTrap™ HP GE Healthcare Life Science; column volume, 5 mL), which had been equilibrated with buffer A (flow rate, 5 mL min-1) using an ÄKTA prime plus (GE Healthcare). The column was washed with 100 ml of buffer A at a flow rate of 5 mL min^{-1}. Elution of the appropriate protein was achieved using a linear gradient 50% of buffer A and 50% of B 0% at a flow rate of 2 mL min^{-1} (Buffer A: 75 mM Tris hydrochloride pH 7.5, 300 mM NaCl, and 20 mM imidazole; buffer B: 5 mM Tris hydrochloride pH 7.5, 300 mM NaCl, and 250 mM imidazole).

Starting from 1500 mL of culture, 40 ml of pure His tagged proteins (1g/L) were obtained.

The purity of the proteins was verified by SDS-PAGE analysis followed by Coomassie blue staining (see below). The eluted PMI was dialyzed against a 50 mM MOPS buffer, pH=7.4; whereas, the eluted PLP was dialyzed against a 50 mM MOPS buffer, 300 mM NaCl, pH = 7.6. His tagged proteins were then concentrated on vivaspin20 10 kDa cut off.

The concentrations were obtained from the absorbance at 280 nm measured on a Nanodrop spectrophotometer (Thermo Scientific) using extinction coefficients at 280 nm calculated from the protein sequence (https://web.expasy.org/protparam/) (28 880 $M^{-1}.cm^{-1}$ for His_6-PMI, 21 430 $M^{-1}.cm^{-1}$ for His_6-PLP, 75 190 $M^{-1}.cm^{-1}$for His_6-MBP-TEV protease and 39 880 $M^{-1}.cm^{-1}$ for His_6-Chimera 1).

- **Cleavage of the His_6 tag:**

  - By Active Tobacco Etch Virus (AcTEV) protease:

In order to cleave the His_6 tag, 1 mL of the purified and concentrated protein (3.48 mg/mL) was mixed with 20 µL of AcTEV protease (10 U/µl, ThermoFisher), 135 µL of 20x Buffer and 10 µL of 0.1M DTT and the resulting solution was incubated at 4°C overnight.

Following digestion, the AcTEV protease terminus was removed from the cleavage reaction mixture by purification on a Ni-NTA affinity column as described above. Pure tag free protein could not bind on the Ni-NTA affinity column and was thus recovered in the flow through. It was dialyzed against 50 mM MOPS buffer, pH=7.1 and concentrated via vivaspin20 cut off.

  - By His_6 tagged MBP TEV protease (His_6-MBP-uTEV3):

His-MBP-uTEV3 was first purified in our lab following the described protocol.[108]

In order to cleave the His_6 tag, 35 mL of the purified and concentrated protein (28 µM in 50mM MOPS buffer) was mixed with 2,8 mL TEV protease (70 µM) and 760 µL of 0.1M DTT ($C_{f\,DTT}$=2 mM) and the resulting solution was incubated at 4°C overnight.

Following digestion, AcTEV protease terminus was removed from the cleavage reaction mixture by purification on a Ni-NTA affinity column. The purification was performed according to the same procedure as that described above. Pure tag free protein could not bind on the Ni-NTA affinity column and thus was recovered in the flow through.

The concentrations of the proteins were obtained from the absorbance at 280 nm measured on a Nanodrop spectrophotometer (Thermo Scientific) using extinction coefficients at 280 nm of 27 390 $M^{-1}.cm^{-1}$ for PMI and 20 220 $M^{-1}.cm^{-1}$ for PLP as calculated from the protein sequence (https://web.expasy.org/protparam/).

- **Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and staining:**

The Proteins present in the samples could be separated by SDS-PAGE and visualized by Coomassie brilliant blue staining (InstantBlue from Expedeom). 60 µL commercial 2x Laemmli (4 % SDS, 20 % glycerol, 10 % 2-mercaptoethanol, 0.004 % bromophenol blue and 0.125 M Tris HCl, pH 6.8, from Sigma) were mixed with 60 µL sample and the resulting solution was boiled for 5 min. 15 µL aliquots of the final denatured samples were deposited in wells of a 12 % polyacrylamide gel. The addition of a mixture of control proteins of known molecular weight (SeeBleu – Thermofisher / Broad Range) in a control well, to make a molecular weight ladder, allowed the verification of the expected protein molecular weights. Migration was performed by vertical electrophoresis in Turbo reducing running buffer (0.6 M MOPS, 2% SDS, 1.2 M Tris base from C.B.S.) at 180 V (requiring about 110 mA) until the visible bromophenol blue reached the top of the gel. The gel was stained by incubation in Coomassie brilliant blue solution for 30 min, washed 5 times with water at 15 min intervals, and fixed with 5 % acetic acid. The amount and the purity of the proteins was estimated from the intensity of the bands revealed on the gel.

- **Western Blot:**

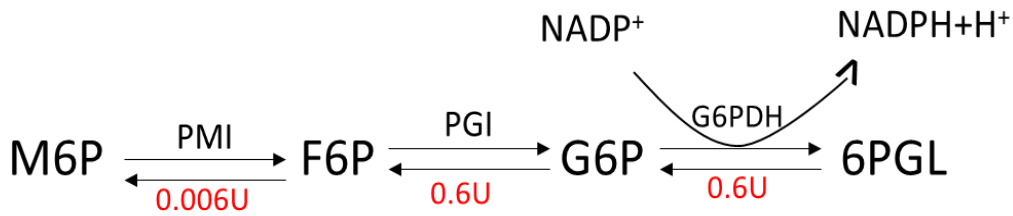
After SDS-PAGE the protein bands were electro transferred horizontally onto a nitrocellulose membrane 70 minutes in migration buffer (25mM Trizma base, 192 mM Glycine, 20% ethanol, pH= 8.4) at 72V (requiring 220 mA). To avoid false positives due to non-specific staining of the nitrocellulose membrane by antibodies, the membrane was soaked in a 3 % BSA solution supplemented with the primary anti-His_6 mouse

antibody (dilution:1:250) and incubated at 4 °C overnight with shaking. The next day, the membrane was washed three times with PBS every 15 minutes and was then incubated with the secondary anti-mouse antibody conjugated with horseradish peroxidase HRP (0.3 µg/mL) in the presence of 20 mL PBS 1X at 4 °C for 2 hours at 4 °C with shaking. The membrane was finally washed 3 times with PBS every 15 minutes. The bands could then be revealed by adding the chemiluminescent substrate provided by Thermo Scientific containing solutions of hydrogen peroxide and luminol which generated light that could be detected on an Image Quant LAS 500 apparatus.

III. CHARACTERIZATION OF THE PROTEINS:

- Enzyme activity assay of PMI

To assay the catalytic activity of PMI for the isomerisation of M6P into F6P, a test involving three steps and two other enzymes PGI (phosphoglucose isomerase) and G6PDH (glucose-6-phosphate dehydrogenase) was set up in the lab (Scheme 7.1).[13] PGI first transforms F6P formed by PMI into G6P, which is further transformed into 6PGL (6-phosphogluconolactone) by G6PDH. The later reaction produces NADPH that can be detected by measuring the absorbance at 340 nm (ε = 6220 $M^{-1}.cm^{-1}$, see scheme 28). PGI and G6PDH are introduced in large excess (100 equiv.) with respect to PMI, so that the rate determining step for forming NADPH becomes the reaction catalyzed by PMI. The amount of NADPH formed is thus proportional to the amount of M6P isomerize by PMI.



Scheme 28: Three reaction system for measuring PMI activity.

PGI (from rabbit muscle) and G6PDH (from yeast) were commercial enzymes provided by Sigma Aldrich. For PGI, the commercial solution was stored at 4 ^{0}C. 500 µL of distilled water were added to the commercial enzyme solution. This solution was aliquoted by 15 µL. The aliquots were stored in the freezer at -20 ^{0}C. From these aliquots, solutions PGI/100 were then prepared with 50 mM HEPES buffer pH 7.4. For G6PDH, the commercial solution was stored at 4 ^{0}C. 500 µL of distilled water were added to the

commercial enzyme solution. This solution was aliquoted by 15 µL. The aliquots were stored in the freezer at -20 ^{0}C. From these aliquots, solutions G6PDH/100 were then prepared with 50 mM HEPES buffer pH 7.4. The PGI/100 and the G6PDH/100 solutions were used for the PMI activity assay.

All the substrates and enzymes except PMI (Table 4) were first added in a 1 cm quartz cell for UV-visible spectrocopy. After 7 minutes of incubation at 25°C, increasing volumes of 12 µM PMI (30-120 µL) were added and the resulting solution was mixed by pipetting up and down. The kinetics was followed by measuring the absorbance at 340 nm (A_{340}) every ten seconds for 10 minutes. The initial rate of the reaction catalyzed by PMI could be determined from the slope (s) at the origin of the curve A_{340} = f(t) : Vi(mM/min) = s/6.22). Vi was measured at various PMI volumes (V) and the slope of the Vi = f(V) provided the specific activity of the per µL of protein U/µL.

Table 4: Various chemicals and enzymes used in the test set up to measure the specific PMI activity. PGI and G6PDH volumes were added so that their final concentration was 100 fold higher than that of PMI.

0.1 M M6P in water	20 µL
0.5 M $MgCl_2$ in water	10 µL
40 mM $NADP^+$ in water	10 µL
V(PGI/100)	40 µL
V(G6PDH/100)	100 µL
12 µM PMI	30, 60, 90 or 120 µL
50 mM MOPS, pH= 7.1	up to 1mL

- **Zinc removal:**

The removal of zinc from the active site of the PMI was performed by incubating 10 mL of a solution of the protein (3 µM in 50 mM MOPS buffer, pH = 7.4) with 10^6 equiv. EDTA (Ethylenediamine-tetraacetic acid) overnight at 4°C with shaking. Then, the protein was washed 3 times through a vivaspin20 10 kDa cut off by 30 mL of a 50 mM MOPS buffer pH 7.4 and was finally dialyzed against 2 L of a 50 mM MOPS buffer, pH 7.4 for 2 days.

- **Iron removal:**

The removal of iron from the active site of the PLP was performed by incubating 10mL of 10 µM of a solution of the protein (in 50 mM MOPS, 300 mM NaCl buffer pH = 7,6) with 10^3 equiv. EDTA for 6 hours at 4°C with shaking. Then, the protein was washed 3 times through a vivaspin20 by 30 mL of a 50 mM MOPS buffer pH 7.4 and was finally dialyzed against 2 L of 50 mM MOPS buffer 300 mM NaCl, pH 7.6 for 2 days.

- **Protein intrinsic fluorescence:**

The fluorescence of the proteins was measured on a FluoroMax-4 spectrofluorometer (Excitation : 280 nm – Emission: 300 nm to 500 nm).

To prepare the artificial enzyme derived from PMI, up to 50 equiv. of Cu $(NO_3)_2$ (C_i=1.2mM in water) were added to 900 µL of a 4 µM solution of PMI in 50 mM MOPS buffer pH 7.4. The interaction between PMI and the metal complex was characterized by measuring the protein intrinsic fluorescence at λ max = 340 nm. The curve giving the variation of the fluorescence as a function of the concentration of added copper(II) salt could be fitted with the following equation, which allowed the determination of the affinity constant of the protein for copper(II):

$$\Delta F = \Delta F_{max} \frac{(n[P]_0 + [L]_0 + K_D) - \sqrt{(n[P]_0 + [L]_0 + K_D)^2 - 4n[P]_0[L]_0}}{2n[P]_0}$$

Here F is the intensity of the fluorescence signal at 340 nm, F0 the intensity in the absence of ligand, [P]0 and [L]0 are total protein and total copper(II) concentrations, respectively, n is the stoichiometry (number of ligand-binding sites per protein); deltaFmax (=F0-Fmax; Fmax is the intensity when the protein was saturated with copper) represents the asymptotic value for DeltaF; DeltaFmax, K_D, and n were calculated from the fitting equation.

To prepare the artificial enzyme derived from PLP, up to 50 equiv. of Cu $(SO_4)_2$ (C_i=1.2mM in water) were added to 900 µL of a 4 µM solution of PLP in 50 mM MOPS buffer, 300 mM NaCl pH 7,4. The interaction between PLP and the metal complex was first characterized by ITC (see following paragraph). It was also characterized by measuring the protein intrinsic fluorescence λ max = 340 nm. The curve giving the variation of the fluorescence as a function of the concentration of added copper(II) salt could be fitted with the above written equation, which allowed the determination of the affinity constant of the protein for copper(II)

- **EPR**

X-band EPR spectra were recorded on a Bruker ELEXSYS 500 spectrometer equipped with a Bruker ER 4116DM X band resonator, an Oxford Instrument continuous flow ESR 900 cryostat, and an Oxford ITC 503 temperature control system.

To prepare the artificial copper metalloenzymes, 1 µL of a 3.5 mM solution of Cu $(NO_3)_2$ in water (12 mg $Cu(NO_3)_2$ in 20 mL water) were added to 50 µL of 80 µM enzyme in 50 mM MOPS, pH=7,4, with 15 % glycerol, (final concentration 80 µM). The EPR spectrum of the proteins was measured at 10 K with 9.3 GHz (MW power = 1.0 mW, MA = 8 Gauss, Mod. Freq = 100 KHz, Gain = 50 db Scans = 10)

- **Isothermal Titration Calorimetry:**

ITC measurements were performed on Microcal Peaq-IT apparatus (Malvern).

To determine the thermodynamic parameters for the binding of metal cations to the proteins, up to 12 equiv. of a solution of metal salt 360 µM in 50 mM MOPS buffer, 300 mM NaCl, pH 7.6 were added to 200 µL of a 30 µM solution of PLP in 50 mM MOPS buffer, 300 mM NaCl, pH 7.6. The interaction between PMI and the metal complex was characterized by measuring the heat variation upon the addition of increasing amounts of metal salt.

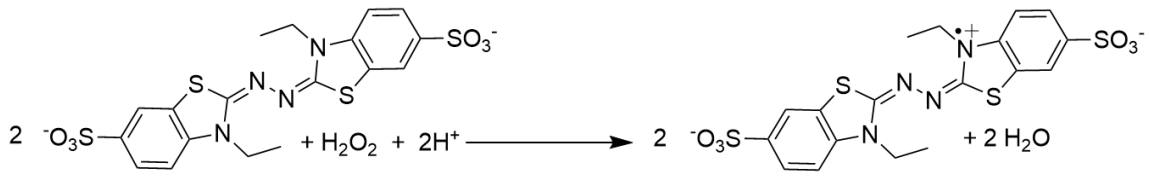
ΔS, ΔH and K_d were calculated from the fitting of the experimental data.

- **TYCHO :**

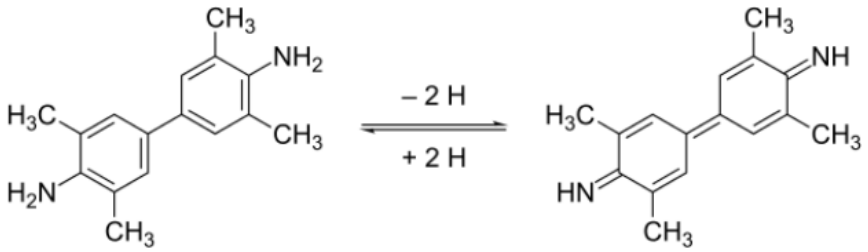
To examine the protein's structural integrity and to determine its folding, measurements were done on Tycho NT.6 (NanoTemper). Changes in intrinsic fluorescence at 350 nm and 330 nm, due to tryptophan and tyrosine residues in the protein were assessed while a heat ramp was applied. Transitions in the folding state of a protein were indicated by changes in the fluorescence signal.

- **Peroxidase activity:**

The oxidation of ABTS and TMB (scheme 29 and scheme 30), was followed by monitoring the absorbance respectively at 414 and 650 nm of the corresponding products obtained after the oxidation by H_2O_2 catalyzed by the artificial copper(II)-PMI metalloenzyme.



Scheme 29: Oxidation reaction of ABTS by H_2O_2.



Scheme 30: Oxidation of TMB by H2O2.

The reactions were carried out in 200 µL wells of a 96 wells non-binding microplate as follows. In each well, 90 µL of 12 µM enzyme in 50 mM MOPS, pH 7,6 were added to 90 µL of 50 mM MOPS, pH 7,6. 10 µL of a 2.5 mM to 10 mM solution of the desired co-substrate in water (TMB and ABTS) were then added to each well. The reaction was then started by the addition of 10 µL of a 10 mM H_2O_2 solution and the absorbance at resp. 415 and 650 nm for ABTS and TMB was monitored as a function of time at 30°C, using a TECAN infinite 200 apparatus. Between each measurement, the microplate was orbitally shaken during 3 seconds with 3 mm of amplitude.

BIBLIOGRAPHY

1. Schmid, A. *et al.* Industrial biocatalysis today and tomorrow. *Nature* **409**, 258–268 (2001).
2. Straathof, A. J. ., Panke, S. & Schmid, A. The production of fine chemicals by biotransformations. *Current Opinion in Biotechnology* **13**, 548–556 (2002).
3. Di Meo, T. *et al.* αRep A3: A Versatile Artificial Scaffold for Metalloenzyme Design. *Chemistry - A European Journal* **23**, 10156–10166 (2017).
4. Steinreiber, J. & Ward, T. R. Artificial metalloenzymes as selective catalysts in aqueous media. *Coordination Chemistry Reviews* **252**, 751–766 (2008).
5. Matsuo, T. & Hirota, S. Artificial enzymes with protein scaffolds: Structural design and modification. *Bioorganic & Medicinal Chemistry* **22**, 5638–5656 (2014).
6. Kuah, E., Toh, S., Yee, J., Ma, Q. & Gao, Z. Enzyme Mimics: Advances and Applications. *Chemistry - A European Journal* **22**, 8404–8430 (2016).
7. Patel, R. N. 9.12 Industrial Applications of Biocatalytic Hydrolysis (Esters, Amides, Epoxides, Nitriles) and Biocatalytic Dynamic Kinetic Resolution. in *Comprehensive Chirality* 288–317 (Elsevier, 2012). doi:10.1016/B978-0-08-095167-6.00913-7.
8. William Strohl. Chimeric Genes,Proteins. *Elsevier* (2017).
9. Harper, S. & Speicher, D. W. Expression and Purification of GST Fusion Proteins. *Current Protocols in Protein Science* **9**, (1997).
10. Li, C.-J. & Trost, B. M. Green chemistry for chemical synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13197–13202 (2008).
11. Dunwell, J. M., Purvis, A. & Khuri, S. Cupins: the most functionally diverse protein superfamily? *Phytochemistry* **65**, 7–17 (2004).
12. Nanda, V. & Koder, R. L. Designing artificial enzymes by intuition and computation. *Nature Chemistry* **2**, 15–24 (2010).
13. Ahmad, L. *et al.* Crystal structure of phosphomannose isomerase from *Candida albicans* complexed with 5-phospho- D -arabinonhydrazide. *FEBS Letters* (2018) doi:10.1002/1873-3468.13059.
14. Widiatningrum, T., Maeda, S., Kataoka, K. & Sakurai, T. A pirin-like protein from Pseudomonas stutzeri and its quercetinase activity. *Biochemistry and Biophysics Reports* **3**, 144–149 (2015).
15. Sommer, M., Xie, H. & Michel, H. Pseudomonas stutzeri as an alternative host for membrane proteins. *Microb Cell Fact* **16**, 157 (2017).
16. Ghattas, W. *et al.* Receptor-Based Artificial Metalloenzymes on Living Human Cells. *J. Am. Chem. Soc.* **140**, 8756–8762 (2018).
17. Lin, Y.-W. Rational design of metalloenzymes: From single to multiple active sites. *Coordination Chemistry Reviews* **336**, 1–27 (2017).
18. Wilson, M. E. & Whitesides, G. M. Conversion of a protein to a homogeneous asymmetric hydrogenation catalyst by site-specific modification with a diphosphinerhodium(I) moiety. *Journal of the American Chemical Society* **100**, 306–307 (1978).

19. Mahy, J.-P., Avenier, F., Ghattas, W., Ricoux, R. & Salmain, M. Current Applications of Artificial Metalloenzymes and Future Developments. in *Enzymes for Solving Humankind's Problems* (eds. Moura, J. J. G., Moura, I. & Maia, L. B.) 363–411 (Springer International Publishing, 2021). doi:10.1007/978-3-030-58315-6_12.
20. Reetz, M. T. Artificial Metalloenzymes as Catalysts in Stereoselective Diels-Alder Reactions. *The Chemical Record* **12**, 391–406 (2012).
21. Minard, P. L'évolution dirigée des protéines. *médecine/sciences* **35**, 169–175 (2019).
22. Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angewandte Chemie International Edition* **57**, 4143–4148 (2018).
23. Ghattas Wadih & Mahy Jean-Pierre. L'Actualité Chimique. in 9–10 (2018).
24. Bonde, J. & Bülow. Chimeric Genes, Proteins. *Elsevier* **1**, 333–334 (2013).
25. Davis, H. J. & Ward, T. R. Artificial Metalloenzymes: Challenges and Opportunities. *ACS Cent. Sci.* **5**, 1120–1136 (2019).
26. El Bakkali-Tahéri, N. *et al.* Characterization of Cu(II)-reconstituted ACC Oxidase using experimental and theoretical approaches. *Archives of Biochemistry and Biophysics* **623**–**624**, 31–41 (2017).
27. Ghattas, W. *et al.* Cu^{II} -Containing 1-Aminocyclopropane Carboxylic Acid Oxidase Is an Efficient Stereospecific Diels–Alderase. *Angew. Chem. Int. Ed.* **58**, 14605–14609 (2019).
28. Yu, K., Liu, C., Kim, B.-G. & Lee, D.-Y. Synthetic fusion protein design and applications. *Biotechnology Advances* **33**, 155–164 (2015).
29. Wang, W., Zhang, J., Alvarez, C., Llopart, A. & Long, M. The Origin of the Jingwei Gene and the Complex Modular Structure of Its Parental Gene, Yellow Emperor, in Drosophila melanogaster. *Molecular Biology and Evolution* **17**, 1294–1301 (2000).
30. Wu, C. *et al.* Simultaneous targeting of multiple disease mediators by a dual-variable-domain immunoglobulin. *Nat Biotechnol* **25**, 1290–1297 (2007).
31. George, R. A. & Heringa, J. An analysis of protein domain linkers: their classification and role in protein folding. *Protein Engineering, Design and Selection* **15**, 871–879 (2002).
32. Chiang, J., Li, I., Pham, E. & Truong, K. FPMOD: A Modeling Tool for Sampling the Conformational Space of Fusion Proteins. in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society* 4111–4114 (IEEE, 2006). doi:10.1109/IEMBS.2006.259224.
33. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* **54**, (2016).
34. Martí-Renom, M. A. *et al.* Comparative Protein Structure Modeling of Genes and Genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325 (2000).
35. Fiser, A., Do, R. K. G. & Šali, A. Modeling of loops in protein structures. *Protein Science* **9**, 1753–1773 (2000).
36. Yang, J. & Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* **43**, W174–W181 (2015).

37. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
38. Aroul-Selvam, R., Hubbard, T. & Sasidharan, R. Domain Insertions in Protein Structures. *Journal of Molecular Biology* **338**, 633–641 (2004).
39. Voigt, C. A., Martinez, C., Wang, Z.-G., Mayo, S. L. & Arnold, F. H. Protein building blocks preserved by recombination. *Nat. Struct Biol.* (2002) doi:10.1038/nsb805.
40. Arnold, U. Incorporation of non-natural modules into proteins: structural features beyond the genetic code. *Biotechnol Lett* **31**, 1129–1139 (2009).
41. Iturrate, L., Sánchez-Moreno, I., Oroz-Guinea, I., Pérez-Gil, J. & García-Junceda, E. Preparation and Characterization of a Bifunctional Aldolase/Kinase Enzyme: A More Efficient Biocatalyst for C▪C Bond Formation. *Chem. Eur. J.* **16**, 4018–4030 (2010).
42. Wang, R., Xue, Y., Wu, X., Song, X. & Peng, J. Enhancement of engineered trifunctional enzyme by optimizing linker peptides for degradation of agricultural by-products. *Enzyme and Microbial Technology* **47**, 194–199 (2010).
43. Choudhury, R., Tsai, Y. S., Dominguez, D., Wang, Y. & Wang, Z. Engineering RNA endonucleases with customized sequence specificities. *Nat Commun* **3**, 1147 (2012).
44. Choi, K.-Y. *et al.* Novel iron-sulfur containing NADPH-Reductase from Nocardia farcinica IFM10152 and fusion construction with CYP51 lanosterol demethylase. *Biotechnol. Bioeng.* **109**, 630–636 (2012).
45. Meister, G. E. & Joshi, N. S. An Engineered Calmodulin-Based Allosteric Switch for Peptide Biosensing. *ChemBioChem* **14**, 1460–1467 (2013).
46. Liu, R. M. *et al.* Synthetic chimeric nucleases function for efficient genome editing. *Nat Commun* **10**, 5524 (2019).
47. *Fusion Protein Technologies for Biopharmaceuticals: Applications and Challenges*. (John Wiley & Sons, Inc., 2013). doi:10.1002/9781118354599.
48. Cornelia Halin, Verena Gafner & Maria Elena Villani. Synergistic therapeutic effects of a tumor targeting antibody fragment, fused to interleukin 12 and to tumor necrosis factor alpha. *Cancer Res* (2003).
49. Cortez-Retamozo, V. *et al.* Efficient Cancer Therapy with a Nanobody-Based Conjugate. *Cancer Research* **64**, 2853–2857 (2004).
50. E W Thompson & B G Lane. Relation of protein synthesis in imbibing wheat embryos to the cell-free translational capacities of bulk mRNA from dry and imbibing embryos. *J. Biol. Chem.* (1980).
51. Faye, L. & Chrispeels, M. J. Common antigenic determinants in the glycoproteins of plants, molluscs and insects. *Glycoconjugate J* **5**, 245–256 (1988).
52. Jaikaran AS. Covalently bonded and adventitious glycans in germin. *J. Biol. Chem.* (1990).
53. Bernier, F., Lemieux, G. & Pallotta, D. Gene families encode the major encystment-specific proteins of Physarum polycephalum plasmodia. *Gene* **59**, 265–277 (1987).

54. Dunwell, J. M. Cupins: A New Superfamily of Functionally Diverse Proteins that Include Germins and Plant Storage Proteins. *Biotechnology and Genetic Engineering Reviews* **15**, 1–32 (1998).
55. Jim M. Dunwell et al. Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities. *Nat. Struct Biol.* (2000).
56. Richard Uberto & Ellen W. Moomaw. Protein Similarity Networks Reveal Relationships among Sequence, Structure, and Function within the Cupin Superfamily. *PLoS ONE* (2013).
57. Gracy, R. W. & Noltmann, E. A. Studies on Phosphomannose Isomerase: II. Characterization As A Zinc Metalloenzyme. *Journal of Biological Chemistry* **243**, 4109–4116 (1968).
58. Jensen, S. O. & Reeves, P. R. Domain organisation in phosphomannose isomerases (types I and II). *Biochim. Biophys. Acta* **1382**, 5–7 (1998).
59. Cleasby, A. *et al.* The x-ray crystal structure of phosphomannose isomerase from Candida albicans at 1.7 angstrom resolution. *Nat. Struct. Biol.* **3**, 470–479 (1996).
60. Shinabarger, D., Berry, A. & May, T. B. Purification and characterization of phosphomannose isomerase-guanosine diphospho-D-mannose pyrophosphorylase. A bifunctional enzyme in the alginate biosynthetic pathway of Pseudomonas aeruginosa. *Biochimica Et Biophysica Acta* **1172**, 329–331 (1991).
61. PATTERSON, J. H., WALLER, R. F., JEEVARAJAH, D., BILLMAN-JACOBE, H. & McCONVILLE, M. J. Mannose metabolism is required for mycobacterial growth. 10 (2003).
62. Payton, M. A., Rheinnecker, M., Klig, L. S., DeTiani, M. & Bowden, E. A novel Saccharomyces cerevisiae secretory mutant possesses a thermolabile phosphomannose isomerase. *J Bacteriol* **173**, 2006–2010 (1991).
63. Smith, D. J., Proudfoot, A. E. I., Detiani, M., Wells, T. N. C. & Payton, M. A. Cloning and heterologous expression of theCandida albicans genePMI 1 encoding phosphomannose isomerase. *Yeast* **11**, 301–310 (1995).
64. Wills, E. A. *et al.* Identification and characterization of the Cryptococcus neoformans phosphomannose isomerase-encoding gene, MAN1, and its impact on pathogenicity: Isolation of the MAN1 gene in Cryptococcus neoformans. *Molecular Microbiology* **40**, 610–620 (2001).
65. Smith, D. J. & Payton, M. A. Hyphal tip extension in Aspergillus nidulans requires the manA gene, which encodes phosphomannose isomerase. *Molecular and Cellular Biology* **14**, 6030–6038 (1994).
66. Schmidt, M., Arnold, W., Niemann, A., Kleickmann, A. & Pühler, A. The Rhizobium meliloti pmi gene encodes a new type of phosphomannose isomerase. *Gene* **122**, 35–43 (1992).
67. Proudfoot, A. E. I., Payton, M. A. & Wells, T. N. C. Purification and characterization of fungal and mammalian phosphomannose isomerases. *J Protein Chem* **13**, 619–627 (1994).

68. Proudfoot, A. E., Turcatti, G., Wells, T. N., Payton, M. A. & Smith, D. J. Purification, cDNA cloning and heterologous expression of human phosphomannose isomerase. *Eur. J. Biochem.* **219**, 415–423 (1994).
69. Islam, M. A. *et al.* Experimental validation of *in silico* model-predicted isocitrate dehydrogenase and phosphomannose isomerase from *D ehalococcoides mccartyi*: Functional characterization of *D. mccartyi* proteins. *Microbial Biotechnology* **9**, 47–60 (2016).
70. Roux, C. *et al.* The reaction mechanism of type I phosphomannose isomerases: New information from inhibition and polarizable molecular mechanics studies: Type I Phosphomannose Isomerases Mechanism. *Proteins* **79**, 203–220 (2011).
71. Swan, M. K., Hansen, T., Schönheit, P. & Davies, C. A Novel Phosphoglucose Isomerase (PGI)/Phosphomannose Isomerase from the Crenarchaeon Pyrobaculum aerophilum Is a Member of the PGI Superfamily. *Journal of Biological Chemistry* **279**, 39838–39845 (2004).
72. Malaisse-Lagael, F., Willem, R., Penders, M. & Malaisse, W. Dual anomeric specificity of phosphomannoisomerase assessed by 2D phase sensitive 13C EXSY NMR. *Molecular and Cellular Biochemistry* **115**, (1992).
73. Duan, Y. *et al.* An efficient and high-throughput protocol for Agrobacterium-mediated transformation based on phosphomannose isomerase positive selection in Japonica rice (Oryza sativa L.). *Plant Cell Reports* **31**, 1611–1624 (2012).
74. Bruns, F. H. & Noltmann, E. Phosphomannoisomerase, an SH-Dependent Metal-Enzyme Complex. *Nature* **181**, 1467–1468 (1958).
75. Gracy, R. W. & Noltmann, E. A. ISOLATION, HOMOGENEITY MEASUREMENTS, AND DETERMINATION OF SOME PHYSICAL PROPERTIES. *Journal of Biological Chemistry* **243**, 3161–3168 (1968).
76. Gracy, R. W. & Noltmann, E. A. Studies On Phospomannose Isomerase: A Mechanism For Catalysis And For The Role Of Zinc In The Enzymatic And The Nonenzymatic Isomerization. *Journal of Biological Chemistry* **243**, 5410–5419 (1968).
77. Fang, W. *et al.* Characterization of the Aspergillus fumigatus phosphomannose isomerase Pmi1 and its impact on cell wall synthesis and morphogenesis. *Microbiology* **155**, 3281–3293 (2009).
78. Everse, J. Heme Proteins. in *Encyclopedia of Biological Chemistry* 354–361 (Elsevier, 2004). doi:10.1016/B0-12-443710-9/00304-5.
79. Lewandowski, Ł., Kepinska, M. & Milnerowicz, H. Inhibition of copper-zinc superoxide dismutase activity by selected environmental xenobiotics. *Environmental Toxicology and Pharmacology* **58**, 105–113 (2018).
80. Wendler, W. M. F., Kremmer, E., Förster, R. & Winnacker, E.-L. Identification of Pirin, a Novel Highly Conserved Nuclear Protein. *Journal of Biological Chemistry* **272**, 8482–8489 (1997).
81. Soo, P.-C. *et al.* Pirin Regulates Pyruvate Catabolism by Interacting with the Pyruvate Dehydrogenase E1 Subunit and Modulating Pyruvate Dehydrogenase Activity. *Journal of Bacteriology* **189**, 109–118 (2007).

82. Liu, F. *et al.* Pirin is an iron-dependent redox regulator of NFκB. *Proceedings of the National Academy of Sciences* **110**, 9722–9727 (2013).
83. Solomon, E. I. *et al.* Copper Active Sites in Biology. *Chem. Rev.* **114**, 3659–3853 (2014).
84. Tranchimand, S., Brouant, P. & Iacazio, G. The rutin catabolic pathway with special emphasis on quercetinase. *Biodegradation* **21**, 833–859 (2010).
85. Fetzner, S. Ring-Cleaving Dioxygenases with a Cupin Fold. *Appl Environ Microbiol* **78**, 2505–2514 (2012).
86. Nianios, D. *et al.* Nickel quercetinase, a "promiscuous" metalloenzyme: metal incorporation and metal ligand substitution studies. *BMC Biochemistry* **16**, (2015).
87. Talà, A. *et al.* Pirin: A novel redox-sensitive modulator of primary and secondary metabolism in Streptomyces. *Metabolic Engineering* **48**, 254–268 (2018).
88. Arai, T. *et al.* Pirin: a potential novel therapeutic target for castration-resistant prostate cancer regulated by *miR-455-5p*. *Mol Oncol* **13**, 322–337 (2019).
89. Porath, J. Immobilized metal ion affinity chromatography. *Protein Expression and Purification* **3**, 263–281 (1992).
90. Barbosa, M. E. M. *et al.* Synthesis and ITC characterization of novel nanoparticles constituted by poly(-benzyl L-glutamate)-cyclodextrin. *Journal of Molecular Recognition* **21**, 169–178 (2008).
91. Janecki, D. J. & Reilly, J. P. Denaturation of metalloproteins with EDTA to facilitate enzymatic digestion and mass fingerprinting. *Rapid Communications in Mass Spectrometry* **19**, 1268–1272 (2005).
92. Müller, M. & Haeberli, A. pH-dependent formation of ethylenediaminetetraacetic acid supramolecular aggregates. *FEBS Letters* **340**, 17–21 (1994).
93. William Studier, F., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. [6] Use of T7 RNA polymerase to direct expression of cloned genes. in *Methods in Enzymology* vol. 185 60–89 (Elsevier, 1990).
94. Sahdev, S., Khattar, S. K. & Saini, K. S. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem* **307**, 249–264 (2007).
95. Leman, J. K. *et al.* Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nature Methods* **17**, 665–680 (2020).
96. Ghosh, S. *et al.* Method for enhancing solubility of the expressed recombinant proteins in *Escherichia coli*. *BioTechniques* **37**, 418–423 (2004).
97. Swartz, J. R. Advances in Escherichia coli production of therapeutic proteins. *Current Opinion in Biotechnology* **12**, 195–201 (2001).
98. Dubendorf, J. W. & Studier, F. W. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *Journal of Molecular Biology* **219**, 45–59 (1991).
99. Firat Duzenli, O., Okay, S., 1 Department of Pharmaceutical Biotechnology, Faculty of Pharmacy, Hacettepe University, Ankara, Turkey & 2 Department of Vaccine Technology, Vaccine Institute, Hacettepe University, Ankara, Turkey. Promoter

engineering for the recombinant protein production in prokaryotic systems. *AIMS Bioengineering* **7**, 62–81 (2020).
100. Bhandari, P. & Gowrishankar, J. An Escherichia coli host strain useful for efficient overproduction of cloned gene products with NaCl as the inducer. *Journal of Bacteriology* **179**, 4403–4406 (1997).
101. Esposito, D. & Chatterjee, D. K. Enhancement of soluble protein expression through the use of fusion tags. *Current Opinion in Biotechnology* **17**, 353–358 (2006).
102. KAZUYO NISHIHARA, TAKASHI YURA, MASAAKI KANEMORI & MASANARI KITAGAWA. Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE and GroEL-GroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2, in Escherichia coli. in *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* vol. 64 1694–1699 (1998).
103. Mileo, E. *et al.* Emerging fields in chaperone proteins: A French workshop. *Biochimie* **151**, 159–165 (2018).
104. Gragerov, A. *et al.* Cooperation of GroEL/GroES and DnaK/DnaJ heat shock proteins in preventing protein misfolding in Escherichia coli. *Proceedings of the National Academy of Sciences* **89**, 10341–10344 (1992).
105. Kusukawa, N. & Yura, T. Heat shock protein GroE of Escherichia coli: key protective roles against thermal stress. *Genes & Development* **2**, 874–882 (1988).
106. Mukhopadhyay, A. Inclusion bodies and purification of proteins in biologically active forms. in *Biotreatment, Downstream Processing and Modelling* vol. 56 61–109 (Springer Berlin Heidelberg, 1997).
107. Marston, F. A. O. & Hartley, D. L. [20] Solubilization of protein aggregates. in *Methods in Enzymology* vol. 182 264–276 (Elsevier, 1990).
108. Sanchez, M. I. & Ting, A. Y. Directed evolution improves the catalytic efficiency of TEV protease. *Nat Methods* **17**, 167–174 (2020).

LIST OF FIGURES:

LIST OF SCHEMES

RESUME ETENDU EN FRANCAIS

Dans le contexte de la chimie verte durable, les réactions chimiques catalysées par des enzymes représentent le meilleur scénario écologique en synthèse organique. Les enzymes peuvent être utilisées dans des conditions écocompatibles (en solution aqueuse, à température ambiante, sous pression atmosphérique) pour obtenir des produits chiraux spécifiques et éviter la formation de produits secondaires indésirables.

Les enzymes naturelles sont classées parmi les catalyseurs les plus puissants des organismes naturels. Elles peuvent catalyser efficacement différents types de réactions.

Pourtant, les enzymes ne peuvent catalyser que des types limités de réactions et les produits souhaités stéréoisomères ne sont pas toujours accessibles. Pour surmonter les limitations des enzymes, le domaine des métalloenzymes artificielles a évolué.

Dans le contexte de la chimie écologique, les métalloenzymes artificielles résultent de l'incorporation d'un cofacteur métallique catalytiquement compétent dans une protéine. La construction de ces biocatalyseurs hybrides combinant des composants biologiques et des composants chimiques synthétiques comprend la mutation génétique, la modification chimique et/ou la combinaison de ces deux méthodes.

Ces biocatalyseurs hybrides présentent plusieurs avantages tels que : une large gamme de substrats et de réactions, la possibilité de catalyser des réactions dans un environnement chiral fourni par le résidu de la protéine et la capacité d'être optimisés par voies chimique et biologique.

Avec les progrès rapides de la recherche sur les métalloenzymes artificielles, ainsi que leurs avantages, elles ont le potentiel de devenir des concurrents légitimes aux enzymes naturelles pour leur utilisation dans un large éventail de domaines allant des industries chimiques et alimentaires aux diagnostics et médecine. Ainsi, les métalloenzymes artificielles sont classées dans la production de carburants, dans les réactions biologiques (mimiques de la ribonucléase A, mimiques de l'ATPase) et dans la chimie verte et durable (Diels Alder, Suzuki-Miyaura et autres pour la synthèse chimique).

Différentes stratégies ont été décrites au cours des années pour développer des métalloenzymes artificielles. Mais dans ce projet, une nouvelle stratégie nous a semblé la plus intéressante : la construction d'enzymes chimériques.

Au cours de la dernière décennie, de nombreux domaines tels que la biochimie, la chimie des protéines, le clonage moléculaire, la mutagenèse aléatoire et dirigée, l'évolution dirigée des biocatalyseurs... ont permis d'utiliser une large gamme d'enzymes et de cultures microbiennes en tant qu'outils de synthèse organique.

Les protéines de fusion, également appelées protéines chimériques, sont produites par la traduction d'un gène qui combine deux ou plusieurs gènes qui codaient auparavant pour des protéines différentes ou similaires. En effet, lorsque ce gène de fusion est traduit, il produit un polypeptide unique dont les caractéristiques fonctionnelles sont dérivées de celles des protéines d'origine.

De plus, les protéines chimériques sont avantageuses car elles permettent de réduire les coûts en remplaçant plusieurs enzymes par une seule enzyme capable de catalyser deux réactions ou plus. Elles sont donc avantageuses pour l'économie d'atomes qui représente le deuxième des 12 principes de la chimie verte, qui vise à optimiser l'incorporation de tous les atomes des matériaux de départ dans le produit final souhaité sans générer de produits secondaires.

La biocatalyse utilisant des enzymes naturelles est la plus souvent limitée aux substrats naturels des enzymes employées et aux réactions biotiques. Les métalloenzymes artificielles combinent à la fois les avantages des métalloenzymes naturels tels que la catalyse stéréosélective en solution aqueuse, dans des conditions douces, et les avantages des catalyseurs à base de complexes métalliques synthétisés chimiquement tels que la catalyse inventée par le chimiste et la compatibilité avec les substrats artificiels.

Dans ce projet, nous avons préparé une protéine chimérique artificielle qui possède plusieurs sites actifs par la fusion des domaines catalytiques de deux enzymes apparentées. Cette enzyme chimérique serait utilisée pour catalyser des réactions en cascade dans des conditions respectueuses de l'environnement.

Pour préparer l'enzyme chimérique, nous nous sommes appuyés sur la superfamille des protéines cupines. Ainsi, l'enzyme chimérique combine les domaines catalytiques de deux protéines bicupines similaires : La Phosphomannose isomérase **PMI** et la Pirin Like Protein **PLP**. Ceci devrait éviter la dénaturation et conduire à un repliement correct.

Avant la préparation des métalloprotéines chimériques, nous avons procédé à l'expression et à la purification de chacune des deux enzymes de départ, la PMI et la PLP, et nous avons testé leurs activités avant et après l'échange des cations métalliques de leur

site actif respectif. Nous prévoyons de comparer ces activités avec celles catalysées par les métalloprotéines chimériques PMI-PLP.

Les PMI sont des métalloenzymes qui catalysent l'isomérisation réversible du D-mannose-6-phosphate en D-fructose-6-phosphate ; leur site actif contient le Zinc dans le coté C – terminal des feuillets β. Puisque les enzymes artificielles formées en incorporant des métaux de transition ou leurs complexes dans des feuillets β sont rares, nous avons préparé et caractérisé une métalloprotéine artificielle basée sur la PMI. Le vecteur d'expression codant pour la PMI His tagguée que nous avons pu cliver en introduisant un site de clivage de la TEV protéase était disponible dans notre laboratoire. Le plasmide a été digéré et un site d'hydrolyse spécifique par la protéase TEV a été introduit avec succès, entre la séquence de la PMI et celle du His tag. Après purification sur une colonne Ni-NTA, l'enzyme a été séparée du tag par digestion avec la TEV protéase. Des analyses sur TYCHO ont montré que la dénaturation thermique de la PMI se produit à 60°C. La PMI contient un zinc(II) par molécule d'enzyme et l'élimination du zinc(II) du site actif a été réalisée par incubation avec 10^6 équivalent d'EDTA.

Pour préparer la métalloenzyme artificielle, du cuivre(II) a été ajouté et la PMI a été transformée avec succès en une protéine artificielle contenant du cuivre(II) qui a été caractérisée par une constante de dissociation relativement bonne (KD = 3 µM). Le spectre RPE du complexe PMI-Cu a confirmé la présence d'un seul type de cuivre d'une symétrie axiale dans le site actif. Le complexe PMI-Cu n'a pas pu catalyser la réaction de Diels-Alder, très probablement parce que le site actif était enfoui dans un canal étroit et n'était pas accessible aux substrats caractéristiques de Diels-Alder. Comme perspectives, des substrats de Diels-Alder plus petits seront utilisés avec cette protéine comme catalyseur. Une activité peroxydase est également envisagée pour la protéine artificielle PMI-Cu.

La PLP est une métalloenzyme qui ressemble à la PMI. Elle appartient à la superfamille des enzymes cupines, et son site actif est situé dans le coté N-terminal. Le rôle biologique de la PLP n'est toujours pas clair jusqu'à présent. Une étude récente a montré que la PLP pouvait être transformée en une métalloenzyme artificielle en substituant son cation ferrique par un cation métallique divalent tel que Cu^{2+}. Dans le but de préparer une protéine chimérique PMI-PLP, nous avons d'abord préparé et caractérisé une métalloprotéine artificielle basée sur la PLP. Le vecteur d'expression codant pour la PLP His tagguée nous a été fourni par l'équipe Kataoka du Japon. Le séquençage de l'ADN fourni a montré une co-expression de la séquence de la PLP avec une partie de LacZ. La co-expression a été confirmée par un western blot anti-His6 qui a révélé deux bandes au

lieu d'une seule. Pour éviter cette co-expression, le plasmide d'expression a été modifié en délétant une partie de LacZ.

Le plasmide a été digéré et un site de clivage de la TEV protéase a été introduit entre la séquence de l'enzyme et de celle de l'His tag. Après expression et purification de la PLP sur une colonne Ni-NTA, l'enzyme a été séparée du tag His par digestion avec la TEV protéase. La PLP produite contient un cation Fe(II) par molécule d'enzyme, par conséquent l'élimination du Fe(II) du site actif a été effectuée en incubant la protéine avec 10^3 équivalent d'EDTA. Pour préparer la métalloprotéine artificielle, Ni^{2+}, Zn^{2+} et Cu^{2+} ont été ajoutés à la protéine sans fer. Une constante de dissociation relativement bonne (KD = 0,5 µM) a été mesurée par extinction de la fluorescence lors de l'addition de quantités croissantes de sulfate de cuivre. Le spectre RPE du complexe PLP-Cu a confirmé la présence d'une seule type de cuivre. Les analyses sur TYCHO ont montré que la dénaturation thermique a eu lieu à 75°C et que des concentrations élevées d'EDTA et de sulfate de cuivre pouvaient altérer le repliement de la PLP. La PLP a donc été transformée avec succès en une protéine artificielle contenant du cuivre(II). Puisque la PLP contenant du cuivre a catalysé l'oxydation de la quercétine, nous nous attendons à ce qu'elle catalyse également la cycloaddition de Diels-Alder après de futurs travaux.

Les deux enzymes de départ ayant été exprimées, purifiées, transformées en métalloprotéines artificielles et caractérisées par différentes techniques, nous avons procédé au développement d'une métalloenzyme chimérique artificielle qui possède plusieurs sites actifs par la fusion des domaines catalytiques de ces deux enzymes apparentées. Tout d'abord, la stratégie que nous avons suivie pour préparer la chimère bicupine PMI-PLP était basée sur des études de modélisation structurelle et sur la détermination du point de liaison entre les deux domaines catalytiques de la PMI et de la PLP. Le serveur elNémo a été utilisé pour simuler les modes normaux de vibration en partant de la basse fréquence de la PMI et de la PLP. Les mouvements de tous les atomes/domaines de la PMI et de la PLP ont ainsi été étudiés et visualisés. Le 1er et le 2ème mode de vibration de la PMI montrent que les deux domaines de cupine vibrent indépendamment. Le 3ème mode normal de vibration montre une cupure au niveau du linnker qui relie les 2 domaines. Donc l'association de deux domaines de cupine différents peut donc se produire par fusion au centre de ce linker qui relie les deux dmaines. Les séquences codant pour le site actif PMI et le site actif PLP ont été clonées par PCR. Les produits PCR ont ensuite été purifiés et ligaturés via la méthode d'assemblage de Gibson.

Le vecteur d'expression construit contenant l'ADN codant pour la chimère PMI-PLP a été utilisé pour transformer les E.*coli* BL21. La purification initiale n'a pas été satisfaisante car la chimère n'est pas pure. Comme la modélisation moléculaire à l'aide de MODELLER a révélé un éventuel clash possible entre une boucle de la PMI et une partie de la PLP, une deuxième enzyme chimérique, Chimère 2, a été construite en remplaçant la boucle de la PMI responsable du clash par une boucle analogue de la PLP qui est beaucoup plus petite. La chimère 2 a également été exprimée mais, comme pour la chimère 1, la purification a de nouveau échoué. En effet, les protéines ont été trouvées principalement dans les fractions insolubles, très probablement dans des corps d'inclusion. La co-expression de protéines chaperonnes avec la chimère 2 n'a pas permis de solubiliser la chimère.

Dans le but d'évaluer la conservation de l'activité initiale d'une bicupine formée par un seul domaine contenant le site actif en absence du second domaine cupin, nous avons procédé à la préparation de chaque moitié des deux enzymes de départ (Half PMI et Half PLP) constituées d'un domaine de feuillets β qui contiennent le métal natif dans leur site actif. Ainsi, les vecteurs d'expression contenant l'ADN codant pour les demi-protéines ont été construits et les protéines ont été exprimées dans E.coli BL21. Le Western blot a confirmé que les deux demi-protéines étaient effectivement exprimées mais les purifications initiales n'étaient pas non plus satisfaisantes. Nous soupçonnons que ces protéines ont également été produites dans des corps d'inclusion, ce qui reste à vérifier par analyse sur gel d'électrophorèse SDS PAGE avant et après la lyse cellulaire. Cela peut être dû à des problèmes de repliement pour les deux demi-protéines puisqu'elles ne sont formées que d'un seul domaine de cupine. Pour ce qui est de la demi-PLP, cela peut être dû à des problèmes de solubilité dérivant de la PLP complète qui a été exprimée à la fois dans le surnageant et dans les corps d'inclusion.

Pour les protéines chimériques, il convient tout d'abord d'examiner les méthodes alternatives de purification telles que l'utilisation d'agents dénaturants afin de solubiliser les corps d'inclusion. Deuxièmement, une autre solution possible menant à une chimère correctement repliée pourrait être d'inverser l'ordre des domaines de la cupine. Comme mentionné dans la littérature, l'ordre d'insertion influence le repliement.

Inspiré par la catalyse enzymatique naturelle et artificielle, l'objectif de la préparation de l'enzyme chimérique PMI-PLP est la catalyse de réactions en cascade. L'obtention d'une enzyme bicupine chimérique correctement repliée reste difficile, alors la mutation du domaine sans métal d'une bi-cupine est une autre option pour atteindre notre objectif. En effet, en s'inspirant du domaine cupin de la PLP qui se lie au métal, la mutagenèse

dirigée peut être utilisée pour créer un site de liaison au métal dans le domaine cupin sans métal de la PMI par exemple. Cette stratégie moins difficile pourrait néanmoins être plus efficace pour atteindre nos objectifs.

Une fois la protéine chimérique obtenue, un point de départ pour l'analyse de la catalyse sera de tester l'activité isomérase dérivant de la PMI parentale afin de pouvoir identifier l'impact du site Fe(III) sur cette réaction.

Par la suite, on considérera que la reconstitution des deux sites actifs Cu(II) conduit à la formation d'une enzyme à deux activités : Diels-Alderase et quercétinase, par exemple. L'étude de ces activités permet d'identifier la synergie potentielle due à la présence de deux sites Cu(II) dans la chimère.

Enfin, la reconstitution de l'enzyme artificielle par un métal différent dans chaque site actif peut être réalisée en éliminant uniquement le cation Fe(III) de la PLP à l'aide d'un ligand qui possède une forte affinité pour le Fe(III) mais une faible affinité pour le Zn(II) comme le déférasirox (un médicament qui est 20 fois plus sélectif pour le Fe(III) que pour le Zn(II)). La reconstitution de l'enzyme par un site actif contenant du Cu(II) devrait fournir l'activité quercétinase tout en conservant le site actif contenant du Zn(II) et son activité isomérase. Ces deux sites actifs très différents peuvent alors être utilisés pour catalyser différentes réactions, ainsi que des réactions en cascade comme l'oxydation de la quercétine catalysée par le cuivre du site actif de la PLP, suivie par l'hydrolyse du produit ainsi formé catalysée par le zinc de la PMI.

ANNEXES

ANNEX 1

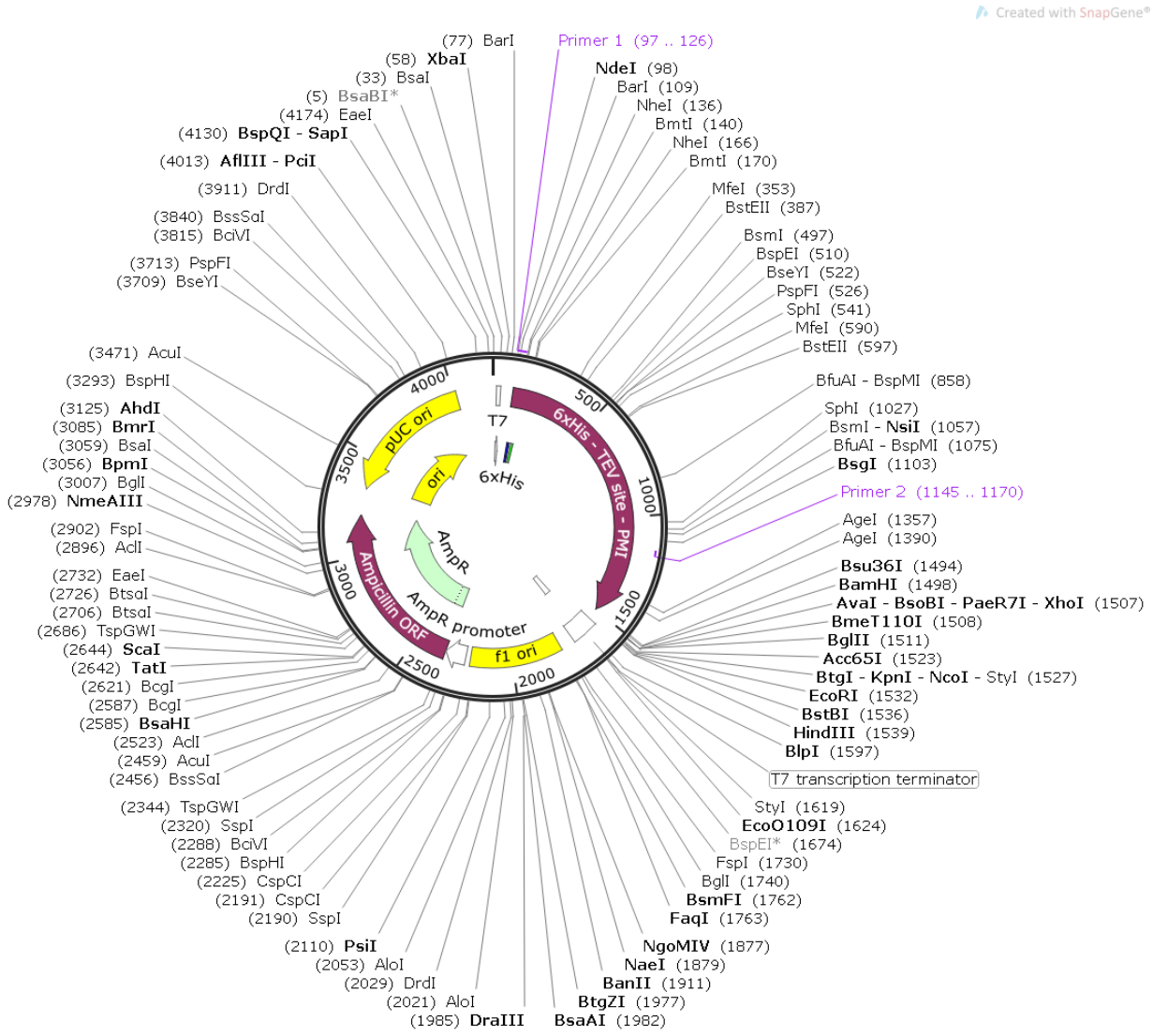


Figure 95: Vector map of pRSET-C showing its important features and containing PMI and His_6 tag flanking $TEV_{c/site}$.

ANNEX 2

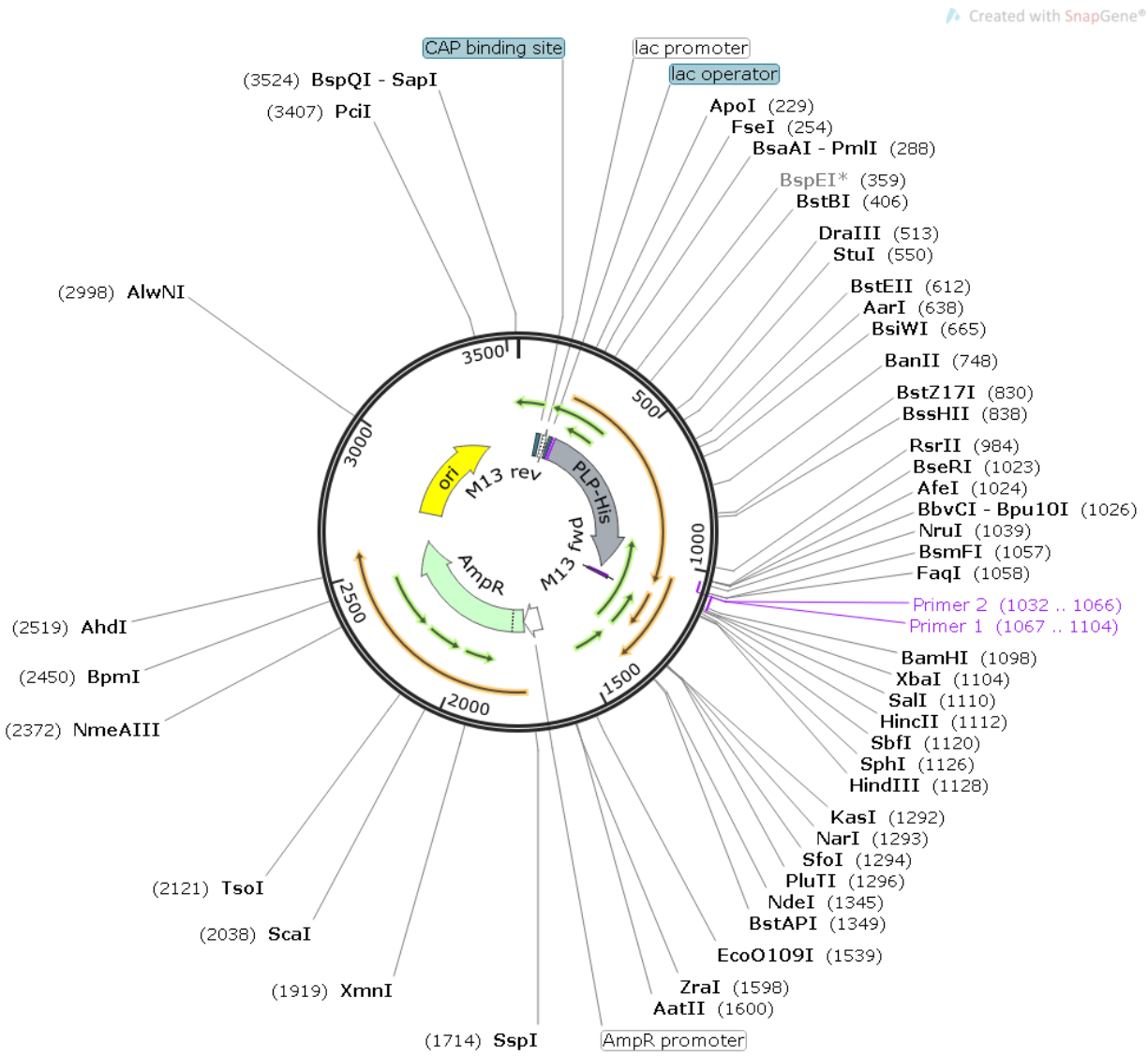


Figure 96: Vector map of pUC18 showing its important features and containing PLP and His_6 tag flanking $TEV_{c/site}$.

ANNEX 3

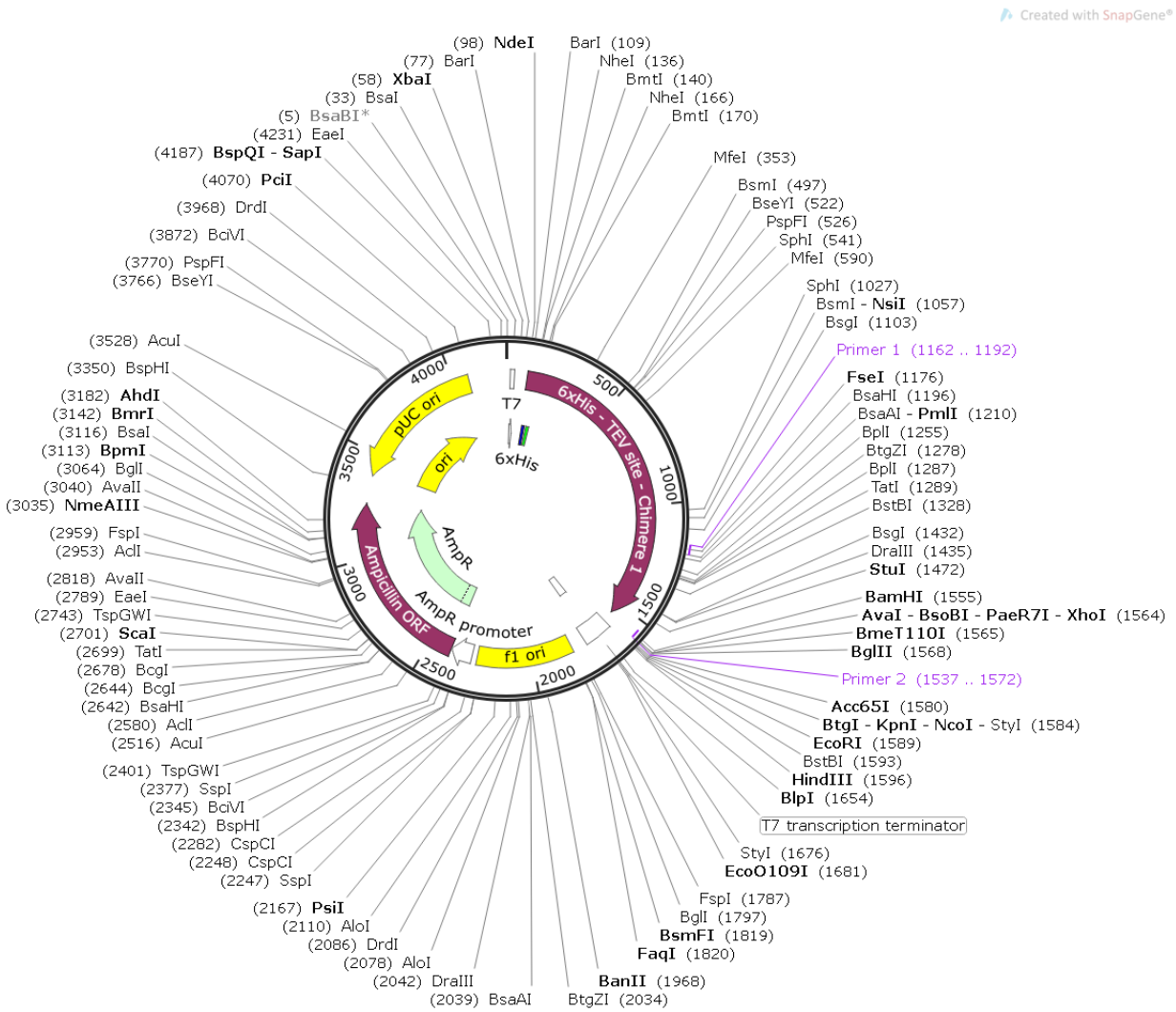


Figure 97: Vector map of pRSET-C showing its important features and containing the first chimeric PMI-PLP protein and His_6 tag flanking $TEV_{c/site}$.