ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE

FORTH
INSTITUTE OF MOLECULAR BIOLOGY & BIOTECHNOLOGY

Master Program

**"Protein Biotechnology"**

University of Crete

Master's thesis

# "Protein biotechnology-based approaches for the functional and structural characterization of insect enzymes"

**Sarafoglou Chara**

**Committee**

Prof. John Vontas

Prof. Michael Kokkinidis

Assist. Prof. Ioannis Pavlidis

**Supervisors**

Prof. John Vontas

Dr. Vasileia Balabanidou

Heraklion, Crete

October 2021

Στον παππού μου,

Χαράλαμπο Κουμπά

## Table of Contents

# Acknowledgments

First of all, I would like to thank Professor John Vontas, for giving me the opportunity to be a member of his lab and work on these exciting projects, believing in me, and specially for offering his scientific knowledge and guidance. In addition, I would like to express my gratitude to Dr. Vasileia Balabanidou who guided me during the whole year, took the time and patience to train me and provide advice about the experiments. Next, I owe my thanks to all members of the lab for their assistance and advice. Thank you for your kindness, hospitality, and assistance from the moment I started my work in the lab.

I want to thank the rest of my Assessment Committee, Professor Michael Kokkinidis for inspiring me and introducing me to structural biology, my field of choice, and his whole group with whom I worked together in the past and maintained a great friendship. I also appreciate Professor's Ioannis Pavlidis help through my thesis. He was kind enough to share his advice, lab equipment and some protocols as well.

I would also like to thank all collaborators for their support and contributions to the projects. I have received enormous support from the protein team, Vasileia Balabanidou, Maria Riga, Dimitra Tsakireli, Eva Katsavou, Konstantinos Parasyris and Stavroula Kaili for all matters of protein expression and more. A huge thanks to Professor Konstantinos Neochoritis and his group for assisting with *in silico* studies and chemical synthesis. It will be a pleasure to work together in future projects. Next, I would like to express my gratitude to Dr Giorgos Gouridis and Dr Yusran A. Muthahari for all the advice and scientific conversations, I really hope for more of those in the future.

Τέλος θα ήθελα να ευχαριστήσω την οικογένειά μου, Δημοσθένη, Μαρία και Γεωργία για την στήριξή τους σε αυτό το βήμα μου αλλά και σε όλα τα προηγούμενα. Τίποτα δεν θα ήταν εφικτό χωρίς την βοήθεια και την αγάπη τους. Θα ήθελα επίσης να ευχαριστήσω την παιδική μου φίλη Γωγώ Τσουκαλά που παρέμεινε δίπλα μου όλα αυτά τα χρόνια στις καλές και στις δύσκολες στιγμές αλλά και τους φίλους μου Χρύσα και Jordan για όλα.

It really was a very interesting year!

# Abstract

Since the practice of agriculture first begun, the loss of yield due to several insect pests was apparent. Even in recent years, with advances in agricultural sciences, losses due to pests and diseases range from 10-90%. Not limited to agricultural damage, insects are also responsible for disease transmission having a profound impact on human and livestock health. To address economy and health issues caused by these insects, various field techniques are used for insect control that can be grouped into three categories: mechanical, biological, and chemical.

Chemical methods are related to the use of insecticides targeting harmful insect species with the intention of eliminating them. Since the first applications of insecticides, insects, pressured by natural selection, started developing resistance mechanisms against them. This created the need of new methods of action in the form of new chemicals or new targets, forming this never-ending cycle.

Protein biotechnology-based approaches is a powerful tool used to win the race against insecticide resistance. *In vitro* techniques are used to isolate and characterize existing or novel insecticide targets. Furthermore, resistance mechanisms can be studied in order to identify patterns and eventually resolve problems by slightly altering existing commercially available insecticides or even by designing new chemicals that by-pass insect defenses. For the purpose of this work, two such examples are being examined.

As a new promising target, type III geranylgeranyl diphosphate synthase from *Helicoverpa armigera* was isolated and examined. After protein purification, an assay was adapted to measure activity and determine kinetic values. Through homology modeling, substrate binding patterns were evaluated, and new chemicals were designed using *in silico* techniques. These chemicals were later tested in the assay to determine potency by high-throughput screening.

Finally, Cytochrome P450 from *Bemicia tabaci*, an enzyme with detoxification function, was studied as it is implicated in insecticide resistance. In this case, the main objective was to establish a reproducible protocol of bacterial expression and purification of the enzyme while keeping it functional in lab conditions. Due to this enzyme's properties the above proved to be challenging and involved protein engineering strategies to improve efficiency.

# CHAPTER 1

## *In vitro* characterization of Geranylgeranyl Diphosphate Synthase from *Helicoverpa armigera*

# 1. INTRODUCTION

## *1.1* *Helicoverpa armigera* – a major pest of agriculture

### *1.1.1* Agricultural impact

Helicoverpa (Heliothis) armigera (Hübner) (Lepidoptera: Noctuidae) has a wide range of suitable hosts and can thrive in several climatic conditions. Adding its ability for rapid distribution due to long distance migration and quick reproduction makes it one of the most significant pests of agriculture (Feng et al. 2005; Head, McCaffery, and Callaghan 1998; Li et al. 2010). Important crops targeted by this species include cotton, maize, chickpea, pigeonpea, sorghum, sunflower, soyabean, and groundnuts (Fitt 1989).

This species comprises two sub-species; *Helicoverpa armigera armigera* is found widespread in central and southern Europe, Asia, and Africa while *Helicoverpa armigera conferta* is native to Australia, and Oceania (Jones et al. 2019). Recently, the former sub-species has also managed to successfully invade Brazil (Tay et al. 2013) and spread across South America and the Caribbean (Figure 1).

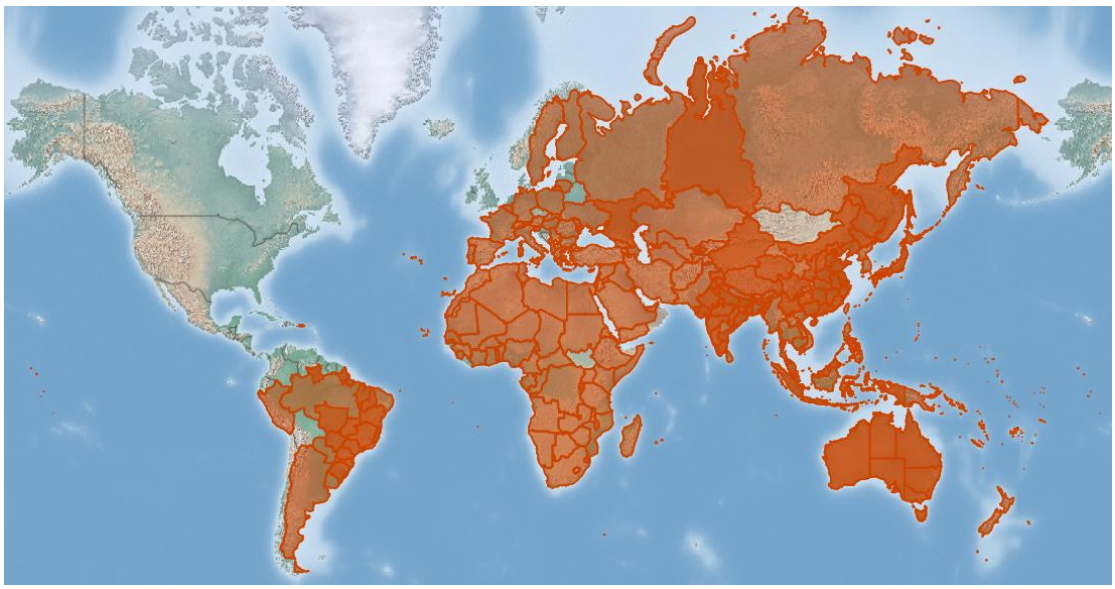


**Figure 1:** Global distribution of *Helicoverpa armigera* as economically important pest. Red coloring indicates countries that the pest is present and deeper color highlights countries that have higher numbers of it. *H. armigera* is new to the Western Hemisphere. It was first reported in Brazil in 2013, although it is suspected to have established earlier (CABI database, June 2021).

Females of the species lay hundreds of eggs each time, placing them in several places of important cultivated crops which provide the perfect nutritious value for the larvae development. This notorious pest causes huge economical damage because it constantly

feeds on fruits and flowers of such crops in its larvae state which usually lasts between 13 and 22 days after the egg hatches (Reed and Pawar 1982).

### *1.1.2* Insect control approaches

Since there are huge economical losses annually, many strategies of insect control were developed and applied to protect agricultural crops. While in general chemical approaches are the most popular for insect control, Helicoverpa armigera has a strong tendency to develop resistance making broad spectrum insecticides less effective. Instead, there are more successful cases of biological control through insect viruses release. Furthermore, in more recent years the use of transgenic plants to control insect population is gaining popularity fast.

Even though *Helicoverpa armigera* has developed resistance against most insecticide classes and particularly pyrethroids and organophosphates, there are still some newer classes of insecticides used successfully for pest control (Yang 2013). Spinosyns, diamides, and growth regulators are currently used and provide plant protection from the species. These new classes of insecticides do not target other beneficial insects making them perfect candidates (Durigan et al. 2017). Since *H. armigera* is a species known for the ability to rapidly develop resistance, the use of these insecticides is rotational to avoid targeting the same molecular mechanism in order to retard the rate of resistance development (Bueno et al. 2017, Pomari-Fernandes et al. 2015, Ahmad et al. 2019).

As an alternative to chemical means, biological control using Helicoverpa armigera nucleopolyhedrovirus (HearNPV) is a valid option. This viral pesticide is frequently used in China since 1993. Even though it is proven to be a viable option, it still has limitations associated with effectiveness only in specific instar larvae states and sensitivity to ultraviolet radiation. Even though initially the wild type viruses were isolated and commercialized, to address limitations, the interest has shifted to genetically modified viral strains, engineered to be more effective and stable (Sun 2015).

Nowadays, more and more genetically modified crops are used to achieve higher production yield or plant protection from pests in agriculture. A very well-known example of this eco-friendly strategy is the transgenic cotton plants expressing cry toxins derived from the soil bacterium *Bacillus thuringiensis* (James 2002, Barwale et al. 2004, Dong et al. 2005). As this approach effectively controls lepidopteran pests without having a negative impact on ecology or crop production rate, transgenic cotton has been one of the most rapidly adopted GM crops in the world (Gianessi and Carpenter 1999, Tabashnik et al. 2002). Even though most of the molecular mechanism of action of Cry toxins is understood (Hofte and Whitely 1989, Schnepf et al. 1998), there are still some challenges remaining to improve efficiency and delay the emergence of resistant strains. To achieve this, several protein assays have been developed to quantify concentration of toxins produced and the effectiveness on lepidopteran pests (Greenplate 1999, Holt et al. 2002).

### *1.1.3* Resistance mechanisms

As discussed before, there are major challenges concerning the control of the lepidopteran pest *Helicoverpa armigera,* that can be attributed to its rapid adaptation to stress induced by the excessive use of insecticides. Popular insecticides as pyrethroids, organophosphates and carbamates are not effective due to insensitive acetylcholinesterases and increased insecticide metabolism (Oppenoorth, 1984). Detoxification mechanisms often involve overexpression and/or mutation of cytochrome P450 genes that result in insecticide metabolism (Walsh, Thomas K et al. 2018, Joußen et al. 2012, Tian et al. 2021). Furthermore, recent studies show that esterases are also implicated in insecticide resistance to organophosphates and synthetic pyrethroids (Teese et al. 2013).

Cases of resistant lepidopteran populations to cry toxins expressed from transgenic cotton plants are also documented. Resistant strains are presumed to have developed due to extensive chemical use (Sayyed and Wright, 2006) and unapproved Bt cotton variety cultivation with inappropriate expression levels of the toxin (Alvi et al. 2012). Cry1Ac is the most popular Cry toxin, as the first generation Bt cotton produced in large scale expressed only this toxin (Tabashnik et al. 2011, Tabashnik et al. 2013). Since the molecular mechanism of action of the toxin has several steps, there are many ways for a pest population to gain resistance. Examples of such mechanisms are altered activation of midgut digestive proteases, toxin sequestration by glycolipid moieties or esterase, elevated immune response, and reduced binding of Cry toxins (Liliana et al. 2013, Bravo et al. 2011, Schnepf et al. 1998, Heckel et al. 2007)

As resistant strains threaten transgenic crop efficiency, the "pyramid" strategy was adopted, where transgenic plants express two different toxins. However, once again, resistant populations emerged with the ability to deal with more than one toxin (Brévault et al. 2013). Other toxins, like Bt Vip3Aa protein that was recently commercialized, tend to have great results when first introduced but after a few years become less effective due to increasing frequency of resistant alleles in lepidopteran species (Chakroun et al. 2016).

## *1.2* Geranylgeranyl diphosphate synthase

### *1.2.1* Natural role of geranylgeranyl diphosphate synthase enzymes

Geranylgeranyl diphosphate synthase (GGPPS) is a short-chain trans-prenyltransferase that catalyzes the condensation of the non-allylic diphosphate, isopentenyl diphosphate (IPP; $C_5$), with allylic diphosphates to generate the 20-carbon prenyl chain (geranylgeranyl diphosphate, GGPP) used for protein prenylation and diterpenoid biosynthesis. GGPPSs are classified into 3 types depending on substrate preference and catalytic mechanism. Type I GGPPS is present in archaeal, type II GGPPS is found in eubacteria and plants, while yeast and the rest of eukaryote have type III GGPPS (Ling et al. 2007). Insects fall into the third category having a preference of forming GGPP using farnesyl diphosphate (FPP) as a substrate.

This enzyme function plays a major role in insects as the product, GGPP, is a precursor essential for the synthesis of diterpenoid (C20) compounds used for insect chemical defence. (Hojo et al., 2007). Moreover, GGPP is also a donor for protein prenylation, a post translational modification where a moiety (GGPP) is transferred to a carboxyl-terminal cysteine of a protein associating it with membranes and facilitating protein-protein interactions (Day et al. 2008). Finally, there are cases that GGPP acts also as a precursor to insect pheromones like frontalin that is used from pine beetles to communicate, initiate, and modulate the mass attack of host trees (Keeling et al. 2013).

GGPPS is found in most organisms and the reaction that it catalyzes is part of the isoprenoid biosynthetic pathway. In insects, the mevalonate (MVA) pathway (Figure 2) leads to isoprenoid synthesis. It begins with two condensations of acetyl-CoA catalyzed by thiolase to produce acetoacetyl-CoA which then reacts with a third molecule of acetyl-CoA resulting in 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). The HMG-CoA is then reduced to mevalonate by HMG-CoA reductase. Mevalonate is further phosphorylated by mevalonate kinase and phosphomevalonate kinase to 5-diphosphomevalonate. Finally, after decarboxylation IPP is formed which is used as substrate together with allylic diphosphates by short-chain trans-prenyltransferases to create isoprenoids (Dhar et al. 2013).

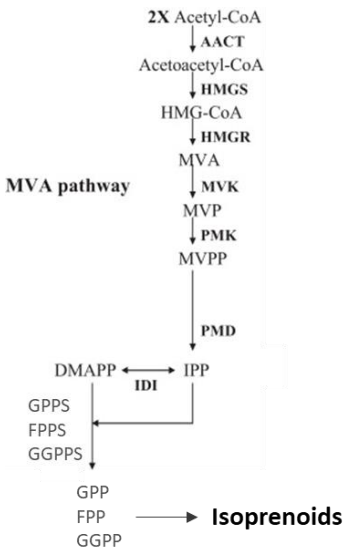



**Figure 2:** Isoprenoid biosynthesis through the mevalonate pathway in insects. The abbreviations used in the figure are: MVA: mevalonic acid, HMG-CoA: hydroxymethylglutaryl CoA, MVP: 5-phosphomevalonate, MVPP: 5-diphosphomevalonate, IPP: isopentenyl pyrophosphate, DMAPP: dimethylallyl pyrophosphate, GPP: geranyl diphosphate, FPP: farnesyl pyrophosphate, GGPP: geranylgeranyl diphosphate, AACT: acetoacetyl-CoA thiolase, HMGS: HMG-CoA synthase, HMGR: HMG-CoA reductase, MVK: mevalonate kinase, PMK: phosphomevalonate kinase, PMD: mevalonate 5-diphosphate decarboxylase, IDI: IPP isomerase, GPPS: geranyl diphosphate synthase, FPPS: FPP synthase, GGPPS: geranylgeranyl diphosphate synthase. Figure adapted from Dhar et al. 2013.

The general catalytic mechanism of short chain prenyltransferases is chain elongation by coupling of an allylic isoprenoid diphosphate (DMAPP, GPP, or FPP) with IPP through electrophilic alkylation of its carbon–carbon double bond (figure 3). When both substrates

bind to the enzyme's active cite, a carbocation is formed at the C1' position of the allylic substrate. For this step the presence of $Mg^{2+}$ or $Mn^{2+}$ is crucial as these cations act as activators. The carbocation electrophilically attacks the C4 position of IPP resulting in formation of a C–C bond between IPP and the allylic substrate (Figure 3). The product is then released from the active site (Koyama et al. 1999).

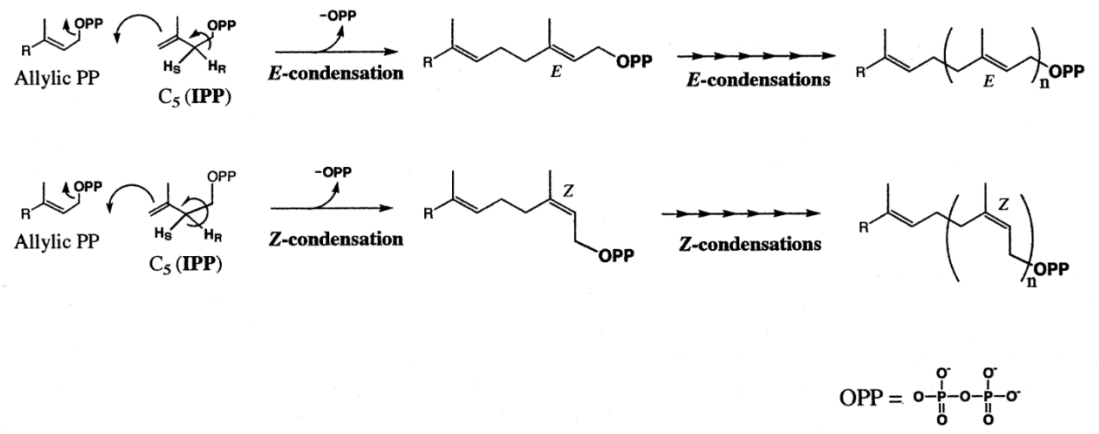



**Figure 3:** Chain elongation mechanism of short chain prenyltransferases. Figure adapted from Koyama et al. 1999.

### *1.2.2* Recent studies of geranylgeranyl diphosphate synthase

GGPPSs are extensively studied in several organisms as they are great drug target candidates. Owing to this, there is an abundance of information about the enzyme's structure, kinetics, inhibition, and general behavior in lab conditions.

Farnesyl diphosphate synthase (FPPS), another short chain prenyltransferases from the mevalonate pathway, is a known inhibition target for cancer treatment. However, in recent years the interest has shifted towards GGPPS and human GGPPS inhibitors are explored as anti-cancer agent candidates. The activity of GGPPSi therapy was evaluated against pancreatic ductal adenocarcinoma (Haney et al. 2019), breast cancer (Dudakovic et al. 2011), and prostate cancer (Weissenrieder et al. 2019). In addition to their application as anti-cancer drugs, GGPPS inhibitors may be of use for the treatment of fibrotic lung disease (Osborn-Heaford et al. 2015).

In plants GGPPS is very well studied as well. Its product, GGPP is the direct precursor for carotenoid biosynthesis, as well as for many other important plastidial isoprenoids, such as gibberellins, chlorophylls, tocopherols, and plastoquinones. In this case, instead of the MVA pathway, in similar manner, the plastidial 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway is used to synthesize GGPP (Botella-Pavía et al. 2004).

In insects this protein is not as well researched and characterized. Despite information about GGPPS is still scarce, there are some studies that highlight the importance of the enzyme in insect organisms. As mentioned before, this enzyme function is linked to protein

prenylation, a procedure of vital importance to insects, and chemical or pheromone production, crucial for insect behavior and communication. For that reason, its potential to be used for pest control is becoming more apparent over the years.

The first key research about the essentiality of GGPPS to insects was in 1998 where recessive lethality was observed at the drosophila first larvae instar when expression of the ggpps gene was reduced by a nonsense mutation (Lai et al. 1998). In contrast, a more recent study in aphids (*Aphis gossypii*), show that RNAi silencing of *ggpps* gene through third instar larvae feeding assays had no obvious cost on either body size or the survival of the treated aphids (Sun et al. 2018). In a different work in aphids (*Acyrthosiphon pisum*) silencing of GGPPS through RNAi injections in adults showed decreased expression of some carotenoid biosynthesis-related genes, including carotenoid synthase/cyclase genes and carotenoid desaturase genes (Ding et al. 2019).

### *1.2.3* Structural characterization

Even though there is no structural information considering insect GGPPS, homologous proteins from several organisms have been crystalized and can provide an insight on structure and protein dynamics. The percentage identity of some species is even high enough to generate homology-based prediction models, especially in areas near the active site that are highly conserved between species.

Studies of *Saccharomyces cerevisiae* GGPPS (also classified as type III) show that it shares a 38% identity with lepidopteran *Helicoverpa armigera*. Results show a tight dimer formation facilitated by an N-terminal α' helix (Figure 4). Each monomer can bind the substrates and catalyze a reaction independently of each other suggesting that the dimer is formed mainly for stability reasons. The substrate binding pocket is a deep cleft with two conserved DDXXDD motifs in the opening. The active site includes polar, mostly positively charged residues, including Lys37, Arg40, His68, Arg84, and Arg85, which are hypothesized to be responsible for attracting or directly interacting with the negatively charged pyrophosphate head groups of the substrates. The $Mg^{2+}$ cofactors are coordinated by Asp75 and Asp79 from the first DDXXDD and four water molecules to obtain the correct orientation (Chang et al. 2006).

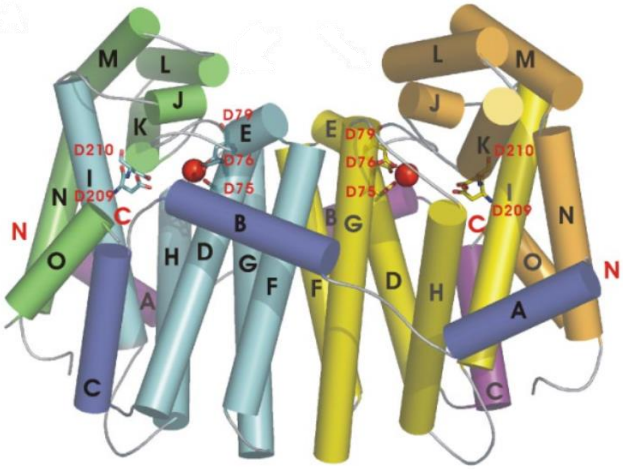



**Figure 4:** The structural model of *S. cerevisiae* GGPPS shown in cylinder diagram. N-terminal helices in the outer layer are shown in blue and purple for the two individual subunits, whereas the six helices in the inner layers are shown in cyan and yellow, respectively. Two small peripheral domains are colored green and orange. $Mg^{2+}$ ions observed near the DDXXDD motifs are shown as red spheres and the Asp residues in the two DDXXDD motifs are presented as sticks. Figure adapted from Chang et al. 2006. (PDB code: 2DH4).

Human GGPPS structure has also been solved by X-ray diffraction (figure 5). Crystallographic data revealed an hexameric quaternary structure with each monomer capable of performing a reaction (Figure 5). The structure is very similar with the yeast GGPPS even in regions with amino acid sequence differences. Interestingly enough, the regions involved in hexamer formation are largely conserved for mammalian and insect GGPPS, suggesting insect GGPPS may be hexameric as well (Kavanagh et al. 2006).

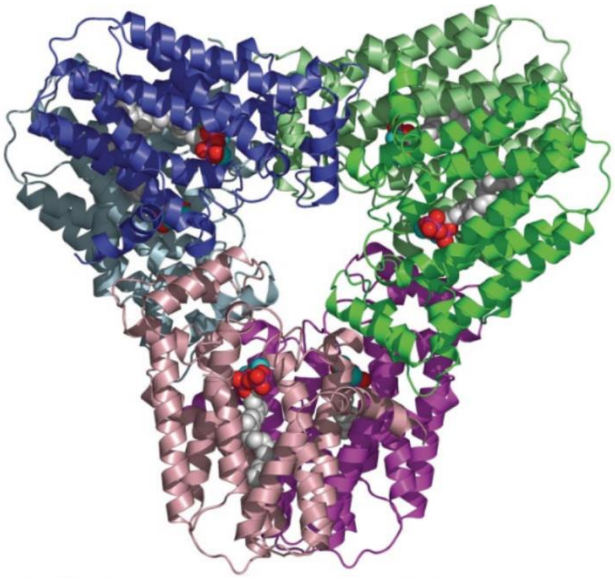


**Figure 5:** The structure of human GGPPS (PDB code: 2Q80) shown in a ribbon diagram. Each monomer is colored differently. $Mg^{2+}$ ions in the catalytic pocket are shown as cyan spheres and GGPP substrates (white, carbons; red, oxygens; pink, phosphorus) are rendered in space-filling representation. Figure adapted from Kavanagh et al. 2006.

In a similar manner to yeast GGPPS, human GGPPS active site is capped by charged and polar residues, including the aspartic reach motifs DDXXDD. The pocket where the substrates bind, has the form of a an approximately 25 -Å- long channel surrounded by mainly aliphatic and aromatic sidechains of residues Arg28, Leu31, Phe35, His57, Leu122, Leu155, Phe156, Ala159, Val160, and Phe184 (Kavanagh et al. 2006).

### *1.2.4* Insect GGPPS *in vitro* characterization

Even though there are not many studies of insect GGPPS, there are some cases that the enzyme was isolated and characterized. The enzyme from *Choristoneura fumiferana* was successfully isolated through bacterial heterologous expression and purification via metal affinity chromatography. The purified protein was easily produced in lab conditions, was soluble, stable, and also active. To measure enzyme activity radiolabeled IPP ([$^{14}$C]IPP) was used to trace chain elongation. In the end, products were extracted and measured for radioactivity via scintillation counter (Barbar et al. 2013).

Even though the use of radiolabeled substrates is a sensitive and accurate technique for kinetic studies, a simpler *in vitro* spectrophotometric assay was proposed for GGPPS activity measurement indirectly through PPi quantification. This method was developed using *Arabidopsis thaliana* GGPPS as a case study (Barja et al. 2020).

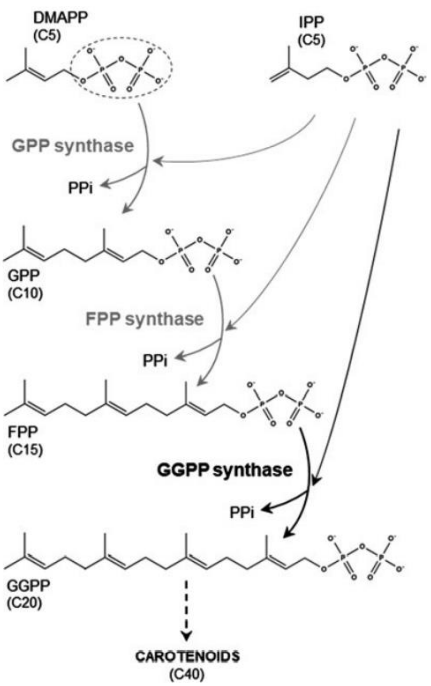



**Figure 6:** Biosynthesis of short-chain prenyl diphosphates by short-chain prenyltransferases. Each IPP condensation involves the elongation of the prenyl diphosphate molecule with the subsequent release of one PPi molecule (dashed circle). Abbreviations: DMAPP: dimethylallyl diphosphate, IPP: isopentenyl diphosphate, PPi: inorganic pyrophosphate, GPP: geranyl diphosphate, FPP: farnesyl diphosphate,

GGPP: geranylgeranyl diphosphate, GPPS: geranyl diphosphate synthase, FPPS: farnesyl diphosphate synthase, GGPP: geranylgeranyl diphosphate synthase. Figure adapted from Barja et al. 2020.

This method uses the already commercially available pyrophosphate quantification kits and couples them with optimal conditions for GGPPS activity. In the final reaction the only PPi produced originates from the catalytic activity of the enzyme. In the end, the measured quantity of PPi can be correlated with product quantity (in this case, GGPP). Since absorbance is measured in the course of time, kinetic values ($K_M$, $V_{max}$) can also be calculated (Barja et al. 2020). Even though this assay has not been used for insect GGPPS before, the activity mechanism is the same, so this method should perform in the same way.

## *1.3* Inhibitor design for geranylgeranyl diphosphate synthase

### *1.3.1* Modern drug design

The goal of drug design is to identify a chemical compound that is capable of interacting with a protein target and disturb its natural function. In the past, new drugs were mostly found by trial and error since there were not enough technological advancements. With new technologies discovered in molecular biology and biotechnology the rationalization of protein-ligand relationships improved drastically making the orthological drug design process possible. Large-scale techniques quickly became popular due to being fast and inexpensive and are being improved constantly over the years (Schedler, 2006).

Bioinformatic and chemoinformatic methods are also widely popular for simulating protein-ligand interactions. *In silico* simulations give a huge advantage to drug design as whole chemical libraries can be tested easily and be scored for a specific protein interaction. These techniques can find potential drugs based on predictions. The more information available as input, the more reliable the prediction is.

“Structure-based drug design” is an approached used when there is a specific protein selected as a drug target (Figure 7). The first step is to obtain a reliable 3D model of the protein that the simulation will later be based on. If the protein structure is known through X-ray crystallography, nuclear magnetic resonance, or other techniques, it is used as a base for the screening. However, if there is no solved structure of the protein, a homology-based model is generated. These models can differ from reality, but drugs usually target protein regions that are conserved between species like the active center of an enzyme, making the simulation more accurate (Cavasotto et al. 2009). Molecular dynamic simulations and docking analysis are methods often used for screening (Ivanov et al. 2006). By calculation binding energy and in some cases other molecular factors such as membrane permeability, potent ligands are scored and can be used for future optimization (Wang et al. 2010, J.W. et al. 2005).

From the process described previously, a list of compounds that can interact in some way with the target is formed. These compounds are called ‘‘hits’’, chemicals that can potentially be optimized and used as drug targets (Wunberg et al. 2006). Hits are turned to leads by biochemical evaluation and limited optimization. Chemical tractability, binding mechanism, pharmacokinetic properties, and freedom of operation (patentability) are usually

important parameters that need to be checked before the chemical moves further ahead the design process (Annis et al. 2004). For lead optimization, new analogs with improved potency, reduced off-target activities, and physiochemical/metabolic properties suggestive of reasonable in vivo pharmacokinetics are designed build on the starting lead compounds. Even though software predictions can be helpful up to a point, this process heavily rely on biochemical feedback (Ferreira et al. 2019). The most promising candidate(s) can then procced to clinical development.

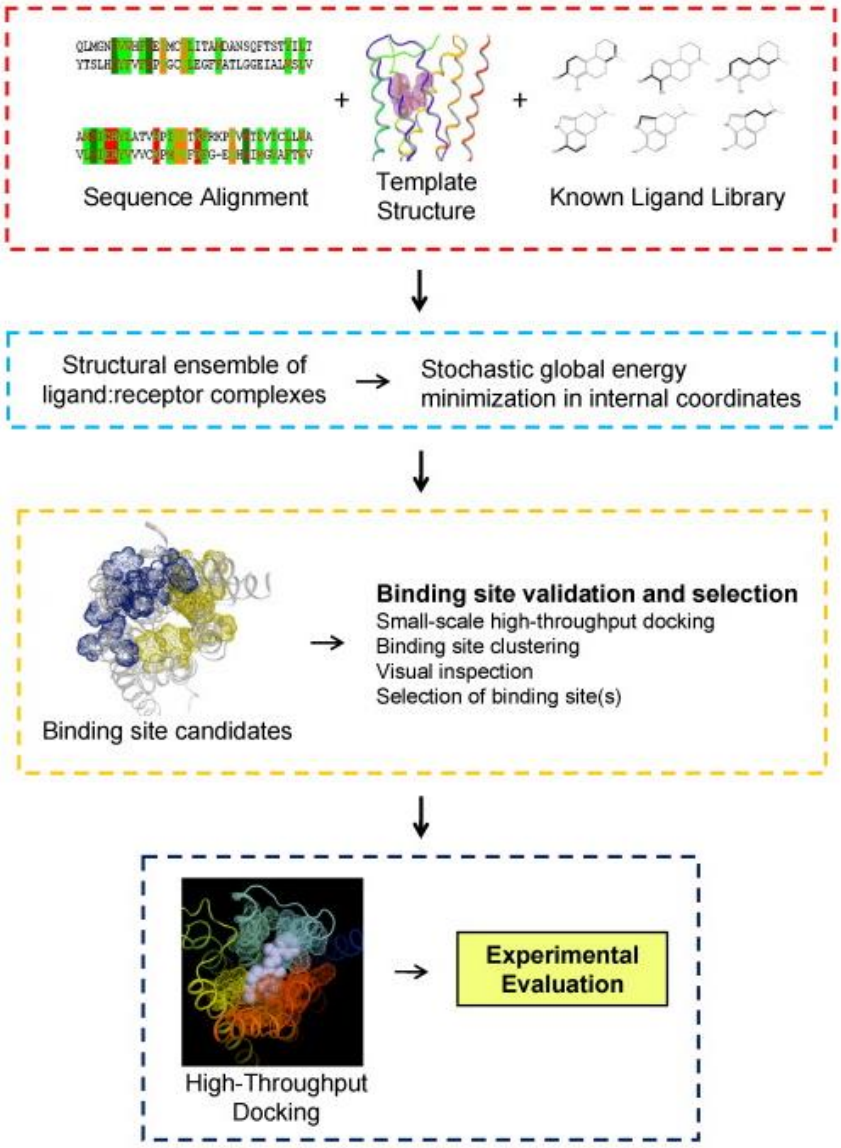



**Figure 7:** Schematic representation of the "Structure-based drug design" method. In the red bordered box, there are the information available in databases about the protein target and ligands, light blue box includes initial structural protein-ligand models generated through an energy minimization technique, yellow box includes small-scale screening after binding site validation and lastly the dark blue box has to do with the final high-throughput docking analysis. Figure adapted from Cavasotto et al. 2009.

### *1.3.2* Known inhibitors for geranylgeranyl diphosphate synthase

Geranylgeranyl diphosphate synthase can be inhibited by bisphosphonate drugs. The binding sites of *Saccharomyces cerevisiae* GGPPS have also been determined based on crystallographic evidence. Examples of such inhibitors are zoledronate and minodronate, which are also known for inhibiting farnesyl diphosphate synthase (FPPS). These two chemicals are bound in the catalytic pocket of the enzyme by forming hydrogen bonds with the conserved residues, Arg 84, Lys 169, Asp 209, and Lys 223 (Figure 8). They are also interacting with the aspartic residues responsible for the $Mg^{2+}$ cofactor coordination (Guo et al. 2007).

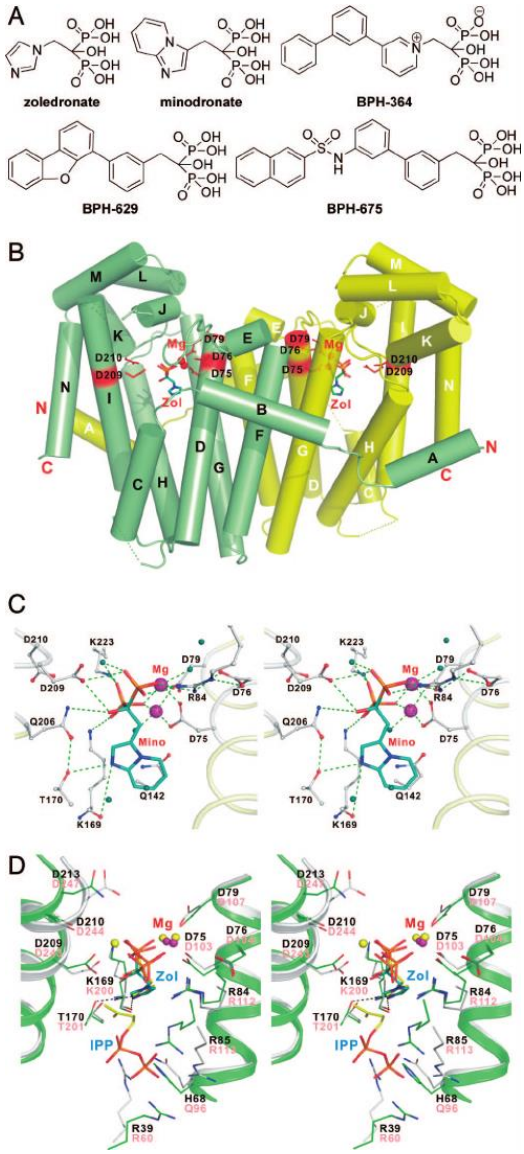



**Figure 8:** Bisphosphonates and GGPPS structures. (A) Structures of bisphosphonates investigated as GGPPS inhibitors. (B) GGPPS structure containing zoledronate (PDB code: 2E91) showing dimer structure. (C) Stereoview of minodronate bound to GGPPS (PDB code: 2E92). (D) Stereoview of

zoledronate/GGPPS (PDB ID code: 2E91) superimposed on zoledronate/IPP/FPPS structure (PDB code: 2F8C). Figure adapted from Guo et al. 2007.

Zoledronate and minodronate are not the most potent inhibitors for *Saccharomyces cerevisiae* GGPPS. Instead, when the binding of three other chemicals (BPH-364, BPH-629, and BPH-675) was tested, these compounds appeared to be more potent inhibitors. This is likely caused by the addition of a large hydrophobic moiety that increases the affinity with the protein target. While BPH-629 has a similar binding position as zoledronate and minodronate, BPH-364 and BPH-675 are positioned deeper in the catalytic pocket and interacting with more residues (Guo et al. 2007).

### *1.3.3* Drug design and selectivity

As discussed before, targeting insect GGPPS needs to be done selectively to avoid toxicity in other non-target organisms. Since this enzyme is present to the majority of organisms including humans, selectivity is very important. However, to make a drug strictly selective to one species is a major challenge. A general strategy that could be followed, looks like an extension to "Structure-based drug design". The first step is to obtain sequence and structural data of targets and non-targets, compare the two categories and identify consistent changes in the regions that ligands are known to interact. After candidate lead chemicals are found, a superposition of the 3-D models of the two categories are made and differences are rationalized in the context of the ligand. A combination of computational simulation and biochemical screening is then tested until a compound is interacting only with the targeted species protein. Rationalizing this system and predicting the outcome by ''building'' a ligand in the catalytic center is highly challenging and can be time consuming (Pastor et al. 1995).

## *1.4* The purpose of this work

The lepidopteran species *Helicoverpa armigera* is known to be a major pest of agriculture and for a good reason. Almost worldwide spread and with an impact on a wide variety of crops of agricultural importance, this insect pest has managed to evolve resistance to the majority of insecticides currently in the market. Since the economical damage is very significant, it is clear that the numbers of this pest must be somehow controlled. In the process of new potential insecticide targets are being explored, geranylgeranyl diphosphate synthase showed high potential as it is a known drug target in other organisms.

This study seeks to understand and explore insect GGPPS as a potential insecticide target by *in vitro* characterization. Briefly, the outline of this work can be divided into three parts. The first objective was the heterologous expression of *Helicoverpa armigera* GGPPS gene using an *E. coli* expression system and its purification by metal affinity chromatography. The second objective was to establish a reproducible method of testing GGPPS activity and to calculate available enzymes' kinetic values. The last objective was to experiment with some inhibitors. Known inhibitors were tested and new inhibitors were designed and evaluated using molecular docking simulations.

# 2. MATERIALS AND METHODS

## 2.1 Bioinformatic tools and other software

### 2.1.1 Visualization software

The protein sequences used for alignments were obtained from UniProt and NCBI (see appendix A for more details). Multiple sequence alignment was performed with Mafft v7.310 (Katoh & Standley, 2013) using the default parameters. Unipro UGENE v40.0 bioinformatic toolkit was used for protein sequence analysis and visualization (http://ugene.net/).

PyMOL v2.5 is a molecular visualization system that was used to generate all protein figures (Martinez et al. 2019). This software was also used to calculate atom distance and evaluate the results of molecular docking (https://pymol.org/2/).

### 2.1.2 Molecular docking software

GOLD, a protein-ligand docking software of Cambridge Crystallographic Data Centre (CCDC) was used for virtual screening (Merk et al. 2018), (available at: https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/, as seen 10/10/2021). The structure of *Saccharomyces cerevisiae* GGPPS (PDB code: 2E8W) was used as the target protein for all the chemicals virtual docking as a reliable crystallographic model of a homologous protein. Docking simulations were targeted to the enzyme catalytic pocket based on previous inhibition models.

### 2.1.3 Graphs and kinetic values calculation

GraphPad prism was used for all statistical analysis and graph generation. The Michaelis–Menten curve fitting and calculation of the enzyme kinetics were also made by this software (available at https://www.graphpad.com/scientific-software/prism/, as seen 10/10/2021). For graph generation, XY tables were generated including all replicates (when available).

For kinetic values, an XY table was made by selecting “Enzyme kinetics—Michaelis–Menten” as sample data and adding the triplicates of the enzyme activities obtained for each substrate concentration. To obtain the regression curve and the kinetic parameters from the generated graph a “Nonlinear regression” analysis was performed by selecting “Enzyme Kinetics— Substrate vs. Velocity” and “Michaelis–Menten equation.” As a result of the analysis, the regression curve was superimposed on the graph and a table with the values of Vmax and Km together with the statistical parameters was retrieved.

## 2.2 Protein expression and purification protocols

### 2.2.1 HaGGPPS gene cloning

The HaGGPPS gene was cloned into pET16b(+) vector. Briefly, primers matching the predicted ORF of HaGGPPS were designed having NdeI and BamHI restriction sites to the 5' end. Phusion polymerase was used to amplify the gene through polymerase chain reaction (PCR) from *H. armigera* cDNA prepared from the L3 midgut. Fragments produced were A tailed and cloned into the pGem-T Easy vector and sequenced. Finally, the selected construct was prepped and digested with NdeI and BamHI and cloned into pET16b(+) vector.

### 2.2.2 HaGGPPS expression

The HaGGPPS gene was cloned to pET16b(+) (CoIE1 plasmids) vector under the regulation of the T7, an IPTG-inducible promoter, carrying a N-terminal 10-His tag. For the initial expression tests the selected construct was used to transform three *E. coli* strains:

1. BL21(DE3) star (Thermo Fisher Scientific)
2. BL21(DE3) pLysS. (Thermo Fisher Scientific)
3. BL21 codon plus (DE3)-RIPL (Agilent)

**BL21** strains are the most widely used hosts for protein expression from pET recombinants and have the advantage of being deficient in both *lon* and *ompT* proteases.

**DE3** indicates that the host is a lysogen of λDE3, and therefore carries a chromosomal copy of the T7 RNA polymerase gene under control of the *lacUV5* promoter. Such strains are suitable for production of protein from target genes cloned in pET vectors by induction with IPTG.

**pLysS** strains express T7 lysozyme, which further suppresses basal expression of T7 RNA polymerase prior to induction, thus stabilizing pET recombinants encoding target proteins that affect cell growth and viability.

**Codon plus** strains dramatically improve protein expression in E. coli by overcoming codon bias.

**Transformation protocol**

All procedures were performed in sterilized conditions, under flame. 100 µL of competent cells were transformed with 50 ng of plasmid vector. After a 30 min incubation in ice, cells were heat socked at 42$^{o}$C for 1 min. Tubes were placed again in ice for 2 min. 900 µL of LB liquid nutrient medium was added to the transformed cells and after incubation at 37$^{o}$C for 1 h shaking at 220 rpm, 100 µL of the transformation mix were transferred into a petri dish with LB solid nutrient medium that had the antibiotics/selection markers needed (30 µg/mL ampicillin for BL21(DE3) star, 30 µg/ml ampicillin and 34 µg/mL chloramphenicol for

BL21(DE3) pLysS and codon plus strains). The final transformation step was an overnight incubation at 37°C.

**Expression test protocol**

After bacterial transformation, a single colony was selected and left to grow on 10 mL LB liquid culture (one for each bacterial strain), supplemented with 30 µg/mL ampicillin (and 34 µg/mL chloramphenicol for BL21(DE3) pLysS and codon plus strains), for 18 h (preculture). For every bacterial strain one 20 mL LB liquid culture, supplemented with the appropriate selection markers as mentioned above, was inoculated with 200 µL of preculture (1 % v/v inoculum volume). Cultures were grown at 30°C and monitored for their OD at 600 nm. When OD reached 0.5, IPTG was added at final concentration of 0.5 mM. After induction, incubation was continued for 4 h (25°C, shaking) and 18 h (25°C, shaking). One mL culture sample, before and after induction (-IPTG, +IPTG), was collected in order to check expression.

The protocol above was repeated with BL21 codon plus bacterial strains and tested in different temperature conditions. Induction temperature was 30°C in one culture and 20 °C in the second. Again, 1 mL culture sample, before and after induction (-IPTG, +IPTG), was collected in order to check expression.

**LB liquid culture**

10 g Tryptone
5 g of Yeast extract
10 g of sodium chloride
up to 1000 mL of $H_2O$.

The pH of the buffer is adjusted to 7.0 – 7.4 by the addition of NaOH.

**Final expression protocol**

After bacterial transformation of BL21 codon plus strain, a single colony was selected and left to grow on a 20 mL liquid culture supplemented with 30 µg/mL ampicillin and 34 µg/mL chloramphenicol, for 16 h (preculture). Two 1 L LB liquid cultures, supplemented with the appropriate selection markers as mentioned above, were inoculated with 10 mL of preculture (1 % v/v inoculum volume). Cultures were grown at 30°C and monitored for their OD at 600 nm. When OD reached 0.5, IPTG was added at a final concentration of 0,5 mM. After induction, incubation was continued for 18 h at 20 °C while shaking at 180 rpm. One mL culture sample, before and after induction (-IPTG, +IPTG), was collected in order to check expression. Cultures were then centrifuged at 6000 rpm for 25min at 4°C. The cell pellet was weighed and stored at -20°C.

Results were evaluated using Sodium Dodecyl Sulphate–polyacrylamide gel electrophoresis (SDS-PAGE, table 1). Protein samples were loaded on a 10% 10-well SDS gel and upon separation, the gel was stained with Coomassie Brilliant Blue. In general, the bacterial cells loaded from each sample were approximately the same (for 1mL bacterial culture of 0.5 OD, 50 µL of 5X sample buffer was used).

**5X sample buffer (10mL)**

3.5 mL Tris, pH 6.8 (from 1M stock)
3.0 mL glycerol (from 100% stock)
1 g SDS
6 mg bromophenol blue Add water up to 8 mL

Aliquot 0.8 mL adding 0.2 mL β-mercaptethanol (resulting in a **5X sample buffer**)

**10X Running buffer preparation (Stock)**

30.0 g of Tris base
144.0 g of glycine
10.0 g of SDS
up to 1000 mL of $H_2O$.
The pH of the buffer should be 8.3 and no pH adjustment is required.

**Table 1:** Reagents for making one SDS-PAGE gel

| **Stacking** | |
|---|---|
| $dH_2O$ | 3 mL |
| 0,5M Tris-HCl pH 6.8 | 1,25 mL |
| Acrylamide 30% | 665 µL |
| 10% SDS | 50 µL |
| 10% APS | 50 µL |
| TEMED | 5 µL |

| | **6%** | **7%** | **10%** | **12%** | **15%** | **17%** |
|---|---|---|---|---|---|---|
| **Running** | | | | | | |
| $dH_2O$ | 5,35 mL | 5 mL | 4 mL | 3,35 mL | 2,35 mL | 1,65 mL |
| 1,5M Tris-HCl pH8,8 | 2,5 mL | 2,5 mL | 2,5 mL | 2,5 mL | 2,5 mL | 2,5 mL |
| Acrylamide 30% | 2 mL | 2,33 mL | 3,33 mL | 4 mL | 5 mL | 5,7 mL |
| 10% SDS | 100 µL | 100 µL | 100 µL | 100 µL | 100 µL | 100 µL |
| 10% APS | 50 µL | 50 µL | 50 µL | 50 µL | 50 µL | 50 µL |
| TEMED | 5 µL | 5 µL | 5 µL | 5 µL | 5 µL | 5 µL |

### Cell lysis

Cell pellets stored at -20 °C were thawed out, using 20mL ice cold lysis buffer for every 1 g of bacterial cell pellet, mixing by pipetting and gentle vortexing. After the samples were completely homogenized, 5 mM β-mercaptethanol, 1 mM EDTA and 2 mM PMSF were added. The sample was then sonicated for 15 min (20'' pulse – 40'' rest). Intact cells were removed by a 10-min centrifugation at 4,000 rpm. The soluble fraction was then centrifuged at 18.000 rpm, 4 °C for 30 minutes.

**Lysis buffer**: 50 mM Tris-HCl pH 7.5, 10% glycerol, 100mM NaCl

### Purification via Q or SP Sepharose ion-exchange chromatography

Q or SP Sepharose ion-exchange chromatography was carried out at 25°C. HaGGPPS lysates (from previous cell lysis step), were diluted until final buffer concentration of 50 mM Tris-HCl pH 7.5, 12.5 mM NaCl and 10% glycerol and loaded on 1 mL of pre-equilibrated Q or SP Sepharose resin (Sigma-Aldrich) (pre-equilibrated with 10 column volumes equilibration buffer) with a flow rate 2 mL/min. The column was washed with 4 column volumes (4 * 1 mL= 4 mL) of wash buffer. Then, the bound proteins were eluted with 2 c.v. (2 ml) of elution buffer and 5 elution fractions, 400 µL each, were collected.

Results were evaluated using Sodium Dodecyl Sulphate–polyacrylamide gel electrophoresis (SDS-PAGE). Protein samples were taken from each purification step and ran on a 12% SDS gel. Upon separation the gel was stained with Coomassie Brilliant Blue to estimate protein purity.

**Equilibration buffer**: 50 mM Tris-HCl pH 7.5, 12.5 mM NaCl, 10% glycerol, 5 mM β-mercaptethanol

**Wash buffer**: 50 mM Tris-HCl pH 7.5, 12.5 mM NaCl, 10% glycerol, 5 mM β-mercaptethanol

**Elution buffer**: 50 mM Tris-HCl pH 7.5, 1 M NaCl, 10% glycerol, 5 mM β-mercaptethanol

### Purification via $Ni^{2+}$-NTA affinity chromatography

$Ni^{2+}$-NTA affinity chromatography was carried out at 4°C. HaGGPPS lysates (from previous cell lysis step) supplemented also with 5 mM imidazole, were loaded on 1 mL of pre-equilibrated $Ni^{2+}$-NTA resin (QIAGEN) (pre-equilibrated with 10 column volumes equilibration buffer) with a flow rate 0.5 mL/min. The column was washed with 10 column volumes (10 * 1 mL = 10 mL) of Wash 1 buffe r followed by wash with 10 c.v. (10 mL) of Wash 2 buffer. Then, the bound proteins were eluted with 5 c.v. (5 mL) of elution buffer and 10 elution fractions, 500µl each, were collected.

Results were evaluated using Sodium Dodecyl Sulphate–polyacrylamide gel electrophoresis (SDS-PAGE). Protein samples were taken from each purification step and ran on a 12% SDS gel. Upon separation the gel was stained with Coomassie Brilliant Blue to estimate protein purity.

**Equilibration buffer**: 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% glycerol, 5 mM β-mercaptethanol

**Wash 1 buffer**: 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% glycerol, 20 mM imidazole, 5 mM β-mercaptethanol.

**Wash 2 buffer**: 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% glycerol, 40 mM imidazole, 5 mM β-mercaptethanol.

**Elution buffer**: 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% glycerol, 200 mM imidazole, 5 mM β-mercaptethanol.

**Bradford Protein Assay**

After purification, protein concentration was measured by Bradford assay.

**Table 2:** Preparation of BSA standards to obtain a standard curve for the Bradford Assay

| µg of protein /20 µL | BSA stock (5 mg/mL) | $dH_2O$ | Final concentration |
|---|---|---|---|
| 2 µg | 20 µL | 980 µL | 100 µg/mL |
| 5 µg | 50 µL | 950 µL | 250 µg/mL |
| 10 µg | 100 µL | 900 µL | 500 µg/mL |
| 15 µg | 150 µL | 850 µL | 750 µg/mL |
| 20 µg | 200 µL | 800 µL | 1 mg/mL |

Bradford reagent was diluted with $dH_2O$ (1:4 dilution). The BSA standards were made by adding 20 µL of each one of the standard stock solutions and up to 1 mL Bradford reagent (table 2). For each one of the samples, between 1-20 µL protein were added in the Bradford reagent (again up to 1 mL). Absorption was measured at 595 nm. The estimation of protein concentration of each sample is based on the linear part of the curve that the BSA standards are giving.

### 2.2.3 HaGGPPS activity assay and inhibitor screening

Purified HaGGPPS protein was used for all the activity assays. All reactions were set according to M. Victoria Barja *in vitro* assay (Barja et al. 2020) to measure the activity of Geranylgeranyl Diphosphate Synthase and Other Short-Chain Prenyltransferases using the EnzChek Pyrophosphate Assay Kit (Thermo Fisher, cat# E6645).

**Table 3:** Reagents and volumes to prepare a standard 200 µL HaGGPPS enzyme reaction

| Reagent | Stock concentration | Assay concentration | Volume added |
|---|---|---|---|
| 20X reaction buffer | 1M Tris-HCl | 50mM Tris-HCl | 10µl |
| | 20mM $MgCl_2$ | 1mM $MgCl_2$ | |
| | pH 7 | pH 7 | |
| MESG | 1mM | 0,2mM | 10 µl |
| PNP | 100U/ml | 1U/ml | 2µl |
| Inorganic pyrophosphatase | 3U/ml | 0,03U/ml | 2µl |
| IPP | 1mM | 150mM | 18µl |
| DMAPP | 1mM | 50mM | 9µl |
| Enzyme | 1,5µg/µl | Varying | |
| Water | | | Up to 200µl |

**Protocol of a standard reaction**

All 200 µL enzyme reactions were carried out in 96-well plates. The reaction mixture was prepared by adding 10 µL of 20X reaction buffer, 40 µL of 1 mM MESG substrate, 2 µL of 100 U/mL PNP and 2 µL of the 3 U/mL inorganic pyrophosphatase. Enough water to reach a final total volume of 200 µL was added minus the volumes of the enzyme, IPP and DMAPP that were added later. A no-pyrophosphatase control was also added (replacing the 3 U/mL inorganic pyrophosphatase with water in the reaction mixture) to control for possible PPi contamination. This reaction should absorb as the no substrate control and if it is higher, it must be subtracted from the experimental samples. The enzyme was added to the reaction (always the same volume not depending on concentration). The reactions, before measured, were preincubated at room temperature for 10 min. The substrates IPP and DMAPP were added (concentration depending on the experiments and mentioned in results) and mixed by up and down pipetting. A control assay was performed in parallel in which solvent buffer was added instead of the experimental substrates (no-substrate control). Reactions were set in a 96-well plate which was then inserted in the spectrophotometer. The absorbance of the reactions was measured at 360 nm as a function of time during 1 h (reads every 2 minutes) at 25ºC. When analyzing the data, the no-substrate control values were subtracted from the corresponding experimental samples.

**PPi standard curve**

The linear range for the quantification of PPi using the EnzChek Kit is from 1 µM to 75µM. Following the standard reaction, increasing amounts of PPi standard solution were added while omitting the volumes of the substrates (IPP and DMAPP) and the experimental enzyme. A no-PPi control without PPi (0 µM) was prepared as well. After mixing all the reagents the plate was incubated at 25 ºC for 10 min. The absorbance was measured at 360 nm while the no-PPi control absorbance value was subtracted from each experimental reaction and the PPi standard curve was built by plotting the absorbance at 360 nm as a function of PPi concentration.

### pH curve

The Kit allows to perform the activity assay over a pH range of 6.5 – 8.5, In this case, five identical reactions were prepared as in a standard 200µl reaction, but the pH of each reaction was prepared separately to have the pH of interest in order to test optimal pH conditions. The pH values tested were 6.5, 7, 7.5, 8, 8.5.

### Activity assay to measure kinetic parameters

To measure the kinetic parameters of HaGGPPS, standard 200µl GGPPS enzyme reactions were made keeping the enzyme concentration constant while increasing the substrate of the enzyme. Since every GGPP molecule is formed from three molecules of IPP and only one of DMAPP Km had to be measured for each substrate individually by keeping the second constant and in excess.

Basic kinetic parameters ($V_{max}$ and $K_M$) for DMAPP substrate were calculated by testing a range of DMAPP concentration (10, 25, 50, 100 and 150 µM) while fixing IPP at 450 µM. Reactions were set in a 96-well plate which was then inserted in the spectrophotometer. The absorbance of the reactions was measured at 360 nm as a function of time during 1 h (reads every 2 min) at 25 $^{o}$C. For data analysis, the absorbance of the no-substate controls at 1 h were subtracted from the ones of the reactions including the substrates (see materials and methods 2.1.3 for kinetic values calculation).

### Inhibition screening

To test the effect of a possible inhibitor on the HaGGPPS catalysis, standard 200µl GGPPS enzyme reaction were made. Protein and substrate concentration (substrate concentrations were approximately at Km values) remained constant while increasing concentration of the inhibitors added. Each experiment had no-substrate controls for every inhibitor concentration used.

Reactions were set in a 96-well plate which was then inserted in the spectrophotometer. The absorbance of the reactions was measured at 360 nm as a function of time during 1 h (reads every 2 minutes) at 25$^{o}$C. For data analysis, the absorbance of the no-substate controls at 1h were subtracted from the ones of the reactions including the inhibitor. The resulting values were compared to the no-inhibitor reaction and a percentage of activity was obtained for every inhibitor concentration. The reaction for every inhibitor concentration war repeated 3 times (three technical replicates).

For a quick and simple inhibitor screening (to see if a chemical has the potential to inhibit the reaction), reactions were set as above but only one concentration of inhibitor was tested (5 µM of inhibitor) and in one replicate. The results are presented as percentages of enzymatic activity.

# 3. Results

## 3.1 Bioinformatics

### 3.1.1 Multiple sequence alignment

Geranylgeranyl diphosphate synthase enzymes are mostly studied in human and yeast. Since in these cases there is an abundance of information about structure and function, these sequences are used in order to find evolutionary relationships between the insect GGPPSs and to identify shared patterns since these are close related genes. Furthermore, by comparing the amino acid sequences, motifs that are related to function can be identified in the insect enzymes.

Multiple sequence alignment (MSA) was performed to compare insect GGPPSs with those of human and yeast (Figure 9). In order to identify differences in amino acid sequence between lepidopteran pests and pollinators three species of each category was chosen for the MSA. Specifically, *Helicoverpa armigera*, *Spodoptera litura* and *Choristoneura fumiferana* were chosen as major lepidopteran pest, while *Osmia lignaria*, *Bombus terrestris* and *Apis mellifera* are the pollinator representatives.

In general, excluding gaps, *Helicoverpa armigera* shares 84% - 95% identity with lepidopteran pest, 63% - 68% identity with pollinators, 55% identity with human GGPPS and 38% identity with yeast GGPPS. Moreover, motifs found in trans-prenyltransferases, and thus are associated with catalytic activity, are highly conserved in all species comparisons. The six insect species presented below, share the same amino acids in motifs I to V. The only exception, is observed in motif III where in position 149, as shown in figure, isoleucine is more frequent in lepidopteran pests, compared to the methionine in all pollinators. Even in this case, the two amino acids share the same properties (both having hydrophobic side chains), with methionine being slightly bulkier than isoleucine.

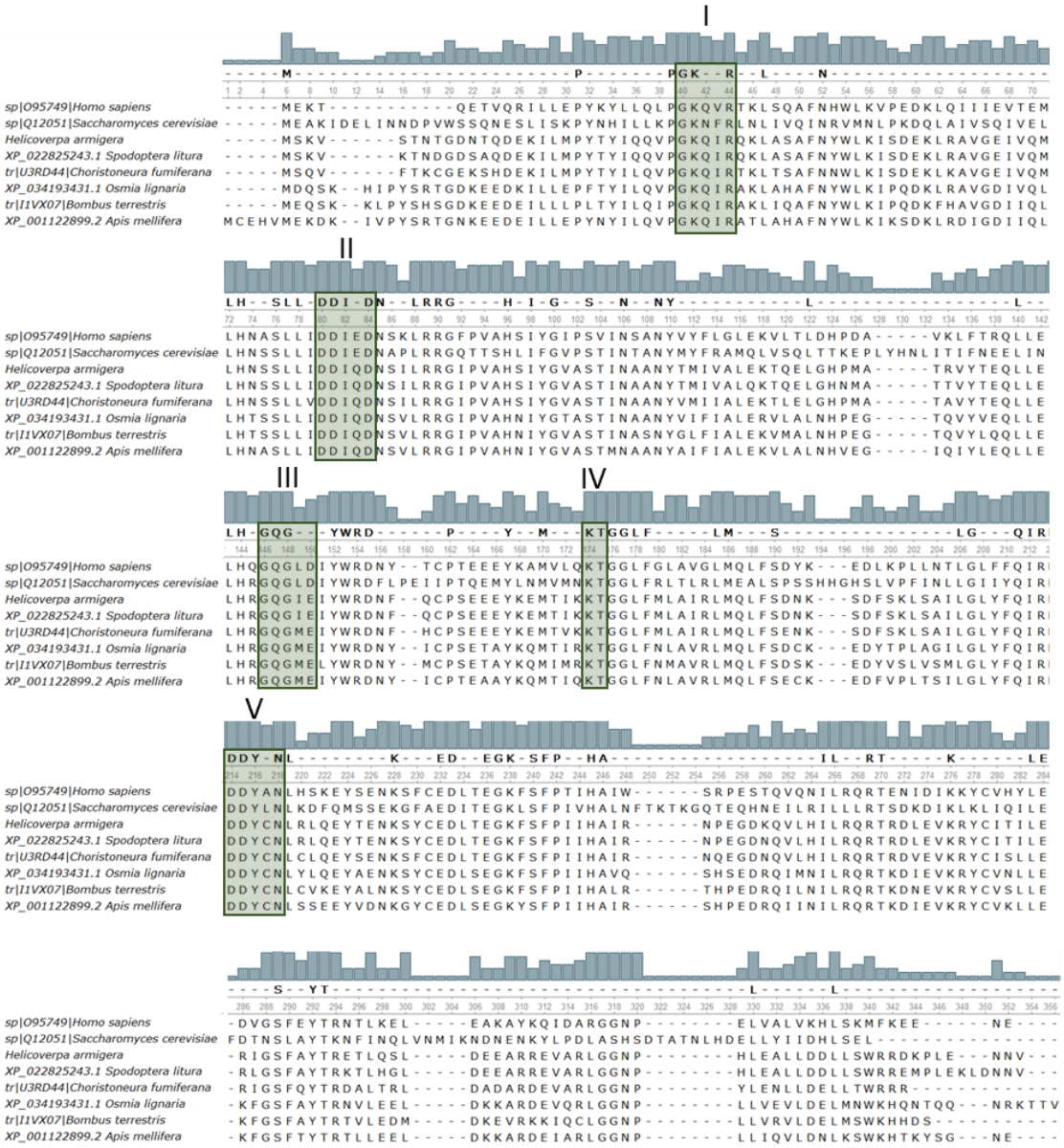


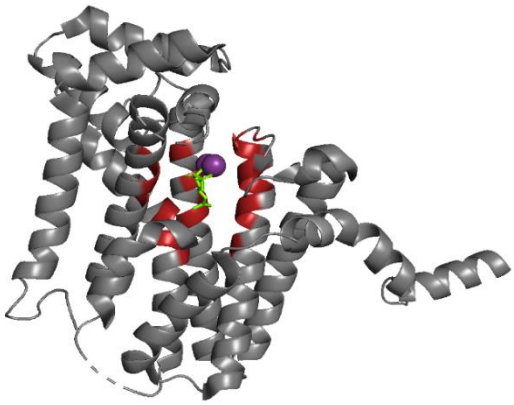

**Figure 9:** Multiple sequence alignment of eukaryotic GGPPSs. Sequences of the genes aligned are from *Homo sapiens*, *Saccharomyces cerevisiae*, from three major lepidopteran pests, *Helicoverpa armigera*, *Spodoptera litura* and *Choristoneura fumiferana* and from three pollinators, *Osmia lignaria*, *Bombus terrestris* and *Apis mellifera*. Conserved sequences motifs found in trans-prenyltransferases are labeled I to V are highlighted in green boxes. Motifs II and V are the characteristic first and second aspartate-rich motifs, FARM and SARM, respectively. Bars above the sequence indicate the level of amino acid conservation.

### 3.1.2 Identifying important amino acids for drug design

Current inhibitors in literature are bound in the catalytic pocket of GGPPS (competitive inhibition) and there is crystallographic evidence of amino acid residues that directly interact with the ligand (substrate, inhibitor, or product). For de novo inhibitor design selectivity needs to be taken into account. In contrast to GGPPSs from insect pests, pollinator GGPPSs must not be inhibited, thus amino acid sequence differences between the two groups must be considered when designing a new inhibitor.

Due to the highly conserved regions of the active site, the crystallographic model of *Saccharomyces cerevisiae* in complex with isopentenyl pyrophosphate (IPP) (PDB code: 2E8W) can be used as a reference for protein – ligand interactions. Since there are no significant amino acid differences in motifs associated with the enzyme activity the search is expanded to residues inside the catalytic pocket and in close proximity with the substrate. Using the substrate as the center, all amino acids with atoms within a distance of 8Å are selected (Figure 10). Because of their proximity to the substrate, these amino acids could be used to design selective inhibitors.

A



B

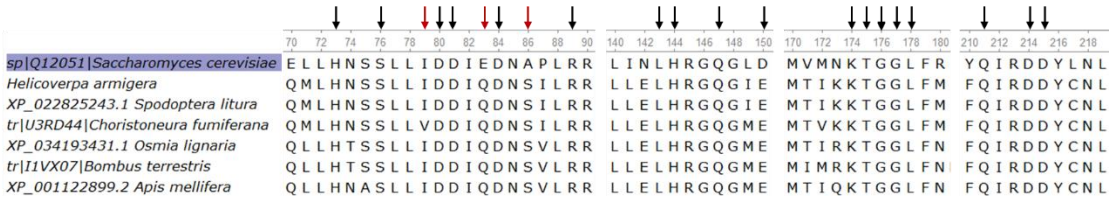
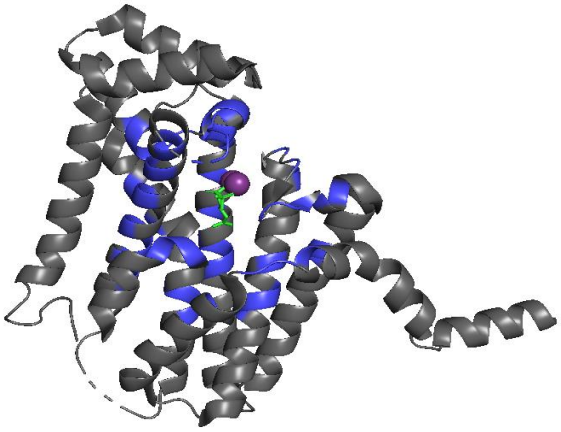



**Figure 10:** A) Crystallographic model of *Saccharomyces cerevisiae* geranylgeranyl diphosphate synthase (ScGGPPS, PDB code: 2E8W) in complex with IPP substrate (in green color) and magnesium cofactors (in purple-colored spheres). The red area includes amino acids in the catalytic pocket that are within 8 Å from the IPP substrate. B) Catalytic pocket sequence comparison 8 Å from the IPP substrate of *Saccharomyces cerevisiae,* lepidopteran, and pollinator GGPPS. Black arrows indicate identities while red arrows indicate amino acid changes between fungal and insect sequence.

Twenty-one amino acids and the magnesium cofactors were within 8Å of the substrate (figure 10). Even though there were a few differences between yeast GGPPS and insect GGPPSs there were no consistent differences observed between insect species.

Since differences in amino acids were not detected in radius of 8Å, the search is expanded to 12 Å (Figure 11) In this case, sixty-one amino acids were found within 8Å to 12Å, six of which were different between pest and pollinator groups (table 4). Most of them share the same properties while there were no significant size differences. Especially interesting cases are residues No.169 in the *Saccharomyces cerevisiae* sequence, where the polar glutamine of lepidopteran pests is substituted by charged glutamic acid in pollinators, and No.180, where the hydrophobic methionine of pests is replaced by polar asparagine in pollinators.

A



B

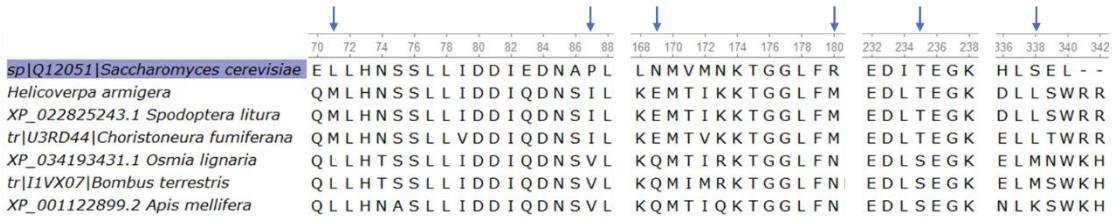



**Figure 11:** A) Crystallographic model of Saccharomyces cerevisiae geranylgeranyl diphosphate synthase (ScGGPPS, PDB code: 2E8W) in complex with IPP substrate (in green color) and magnesium cofactors (in purple-colored spheres). The blue area includes amino acids in the catalytic pocket that are within 8 Å to 12 Å from the IPP substrate. B) Catalytic pocket sequence comparison within 8 Å to 12 Å from the IPP substrate of *Saccharomyces cerevisiae,* lepidopteran, and pollinator GGPPS. Blue arrows indicate amino acid residues that are different between pest and pollinator sequences.

**Table 4:** Amino acids in the catalytic pocket that are within 8Å to 12Å from the IPP substrate and differ between pest and pollinator groups.

| Pest GGPPSs | Pollinator GGPPSs | Residue No. |
|---|---|---|
| M | L | 71 |
| I | V | 87 |
| E | Q | 169 |
| M | N | 180 |
| T | S | 235 |
| L | M/K | 338 |

Inhibitors of GGPPS enzymes that are described in recent literature are decreasing catalytic activity by disturbing the magnesium cofactor interactions. Since this is a conserved, common mechanism between GGPPSs of all organisms, such inhibitors would not be able to target a specific insect species. Instead, new inhibitors could be designed by using these different amino acids to increase inhibitor affinity to the lepidopteran pest enzyme and decrease it in the case of pollinators.

## 3.2 *Helicoverpa armigera* geranylgeranyl diphosphate synthase expression and purification

### 3.2.1 Expression test

The HaGGPPS gene was cloned to pET16b (CoIE1 plasmids) vector, carrying a N-terminal 10-his-tag, under the regulation of the T7, an IPTG-inducible promoter. For the recombinant protein expression three *E. coli* BL21 DE3 strains were tested [BL21(DE3) pLysS, BL21(DE3) star, BL21-CodonPlus] in small scale (20 mL cultures) under the same conditions (see materials and methods) to establish the most suitable system for expression.

From this experiment it is deduced that even though all bacterial strains tested are able to successfully express the target protein, BL21-CodonPlus cells performed the best and thus are used for all further experiments. Furthermore, extending the induction time to eighteen hours also resulted in higher levels of protein expression (Figure 12).

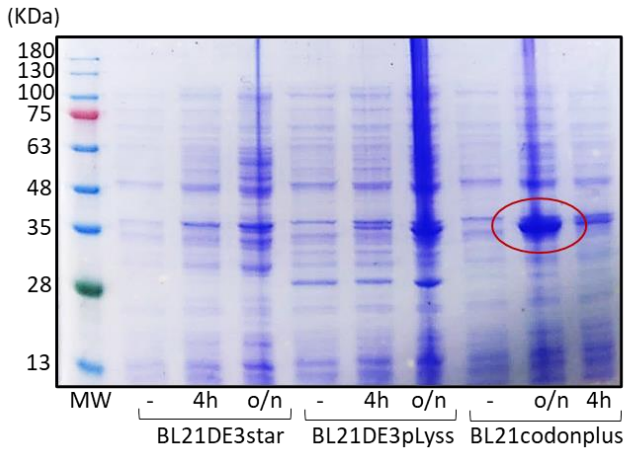



**Figure 12:** Expression patterns of his-HaGGPPS in different E. coli strains. Proteins were separated with SDS-PAGE (12%) and stained with Coomassie Brilliant Blue – Calculated MW of his-HaGGPPS is 38,47 kDa. For each bacterial strain [BL21(DE3) pLysS, BL21(DE3) star and BL21-CodonPlus] a before induction total bacterial extract (-), an after 4 h induction total bacterial extract (4 h) and an after overnight (18 h) induction total bacterial extract (o/n) sample is presented. The condition with the highest level of protein expression is highlighted in a red circle.

As a second round of expression test, the induction temperature was tested. Two experiments were conducted keeping the same conditions (see materials and methods) and only varying in the induction temperature (20°C and 30°C), (Figure 13).

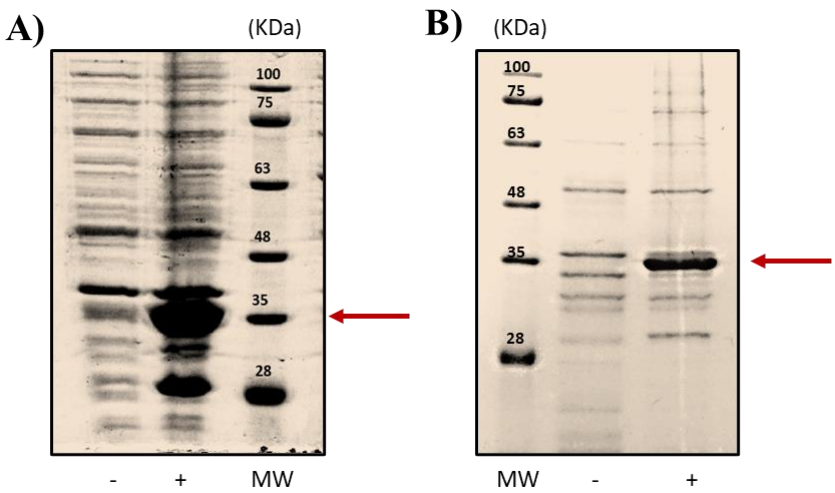



**Figure 13:** Expression patterns of his-HaGGPPS in different induction temperatures. Proteins were separated with SDS-PAGE (10% or 12%) and stained with Coomassie Brilliant Blue – Calculated MW of his-HaGGPPS is 38,47 kDa (red arrows indicating expected gel height). A) HaGGPPS expressed in BL21-CodonPlus bacterial strain, with induction temperature at 30°C. A before induction total bacterial extract (-) and an after overnight (18 h) induction total bacterial extract (+) sample is presented. B) HaGGPPS expressed in BL21-CodonPlus bacterial strain, with induction temperature at 20°C. A before induction total bacterial extract (-) and an after overnight (18 h) induction total bacterial extract (+) sample is presented.

The results of this experiment show that by raising the induction temperature, higher levels of soluble protein expression are achieved. Initially, in further purification experiments the 30°C induction temperature was used as it was proved to have a higher yield. However, protein aggregation problems were observed due to high protein concentration during affinity chromatography and again after dialysis. For that reason, the induction temperature of 20 °C was better, and thus used for the following experiments, as it solved the protein precipitation problems.

### 3.2.2 Purification via ion exchange chromatography

To measure protein activity in an assay, the HaGGPPS must first be purified. The first purification attempt was carried out using Ion exchange chromatography. Bacterial cell pellets that expressed the target protein (from 1 L bacterial culture) were gathered and after cell lysis with sonication the soluble fraction was gathered. This sample was loaded in Q and SP Sepharose resin keeping all chromatography buffers the same in both cases (Figure 14). Since the two resins have opposite charge, the protein was also expected to behave differently (bind or not) in the two resins.

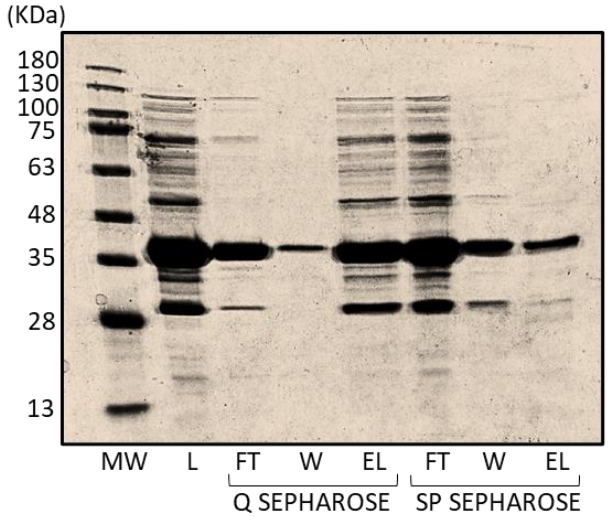



**Figure 14:** First purification attempt of his-HaGGPPS via ion exchange chromatography. Total soluble bacterial extracts were loaded in a Q and SP Sepharose resin. Proteins were separated with SDS-PAGE (10%) and stained with Coomassie Brilliant Blue – Calculated MW of his-HaGGPPS is 38.47 kDa. MW: molecular weight marker, L: total protein loading material, FT: unbound proteins W: wash and EL: second elution fraction of bound proteins are shown.

The calculated PI of his-HaGGPPS is 6.55. Purification was carried out using buffers having a pH of 7.5 so the target protein is expected to bind to an anion exchange resin (Q Sepharose). As shown in the results above, GGPPS tends to bind in the Q Sepharose resin while is lost in the unbound protein material fraction of the SP Sepharose resin followed by most of the other bacterial proteins (low percentage of purity).

Because of the higher level of purity, the unbound and wash fraction of Q Sepharose (FT, W) were pooled together and dialyzed. After dialysis, the protein sample (8 mL total) concentration was measured using Bradford assay. The protein concentration was found on an average of 0.45 mg/mL giving a final yield of 3.6 mg of HaGGPPS / 1 L bacterial culture.

### 3.2.3 Purification via metal affinity chromatography

The second purification attempt was carried out using metal affinity chromatography. Bacterial cell pellets that expressed the target protein (from 1L bacterial culture) were gathered and after cell lysis with sonication the soluble fraction was gathered. Lysate was loaded in $Ni^{2+}$-NTA resin. As mentioned before, HaGGPPS was cloned and expressed fused with a 10X-his-tag which makes the protein bind with a high affinity to Nickel resin. In the end the protein expected to be bound to the Nickel resin is eluted with imidazole.

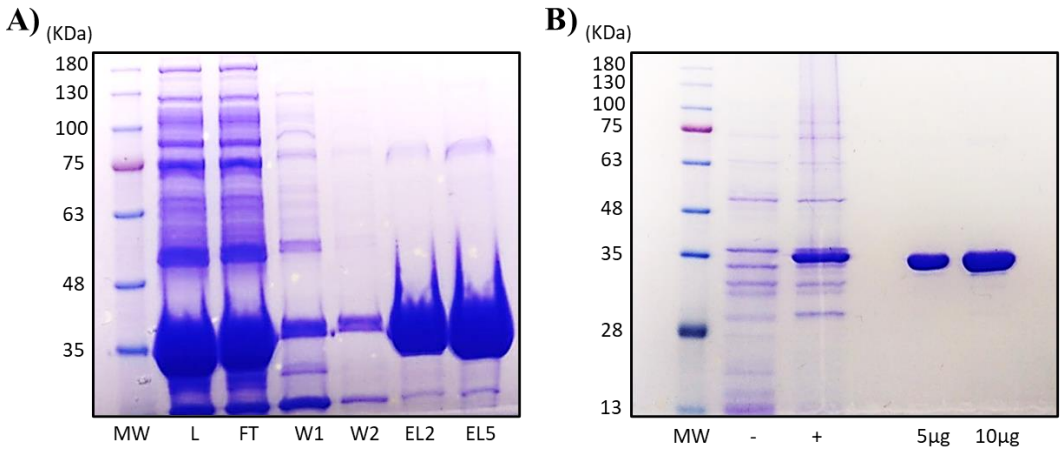



**Figure 15:** Second purification attempt of his-HaGGPPS via metal affinity chromatography. Total soluble bacterial extracts were loaded in a $Ni^{2+}$-NTA resin. Proteins were separated with SDS-PAGE (10%) and stained with Coomassie Brilliant Blue – Calculated MW of his-HaGGPPS is 38.47 kDa. A] All purification steps of metal affinity chromatography. MW: molecular weight marker, L: total protein loading material, FT: unbound proteins W1: first wash, W2: second wash, EL2: second elution fraction of bound proteins, EL5: fifth elution fraction. B] Final purification result after dialysis. MW: molecular weight marker, (-) before induction total bacterial extract, (+) after overnight (eighteen-hour) induction total bacterial extract, (5 µg) 5 µg of purified his-HaGGPPS, (10 µg) 10 µg of purified his-HaGGPPS.

As shown in figure 15, the protein did bind with high affinity to the nickel resin and a high level of purity was achieved. Large amount of the target protein was also lost to the unbound protein fraction of the purification indicating that the protein quantity was exceeding the resin capacity. The elution fractions were overloaded in the gel to assess protein purity and in the end all elution fractions were pooled and dialyzed.

After dialysis, the protein sample (8 mL total) concentration was measured using Bradford assay. The protein concentration was found on an average of 1.3 mg/mL giving a final yield of 10.4 mg of HaGGPPS / 1 L bacterial culture.

From the two purification attempts, metal affinity chromatography has better results in terms of the amount of protein obtained in the end as well as the level of purity. Since the protein was purified in order to be characterized and measure its enzymatic activity, high level of purity is essential. The following experiments were conducted using the protein obtained after purification via $Ni^{2+}$-NTA metal affinity chromatography.

## 3.3 HaGGPPS in vitro characterization

### 3.3.1 Setting the assay

The method followed to measure protein activity is based on the fact that geranylgeranyl diphosphate synthase releases a pyrophosphate molecule (PPi) in every reaction cycle. Using the commercially available EnzChek Pyrophosphate Assay Kit (E-6645), the PPi release can be quantified indirectly by spectrophotometry. Briefly, the PPi produced is converted into two molecules of inorganic phosphate (Pi) by inorganic pyrophosphatase (first enzyme provided by the kit) which then react with the 2-amino-6-mercapto-7-mehtylpurine ribonucleoside (MESG, substrate provided by the kit) by a purine nucleoside phosphorylase (PNP, second enzyme provided by the kit). The final products are ribose 1-phosphate and 2-amino-6-mercapto-7-methyl-purine, which can be monitored at 360nm (Figure 16).

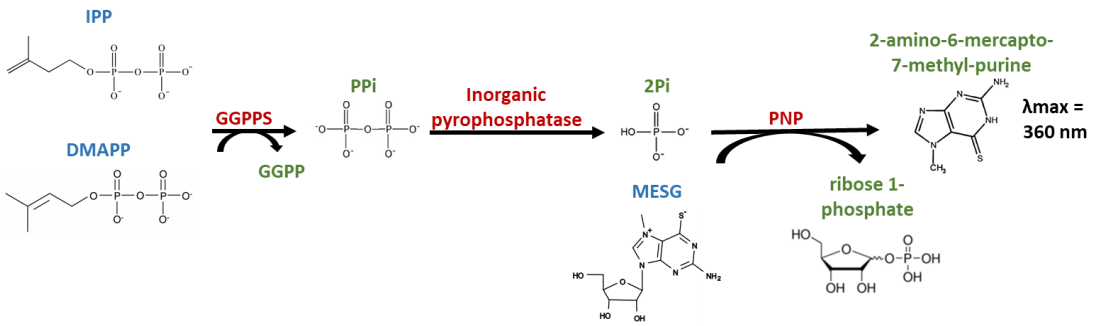



**Figure 16:** Schematic representation of the chain reaction taking place in the GGPPS assay. All substrates are indicated in blue color, enzymes in red and products in green. The final product, 2-amino-6-mercapto-7-methyl-purine, can be measured at 360nm.

The first step of setting the assay was to build a pH curve. As indicated, the assay can work in a pH range of 6,5 to 8,5. For this reason the assay was repeated with the same conditions (50 µg HaGGPPS, 50 mM Tris-HCl, 1mM MgCl2, 0,2 mM MESG, 1 U/mL PNP, 0,03 U/mL Inorganic pyrophosphatase, 150 mM IPP, 50 mM DMAPP, total reaction volume 200 µL) but in different pH values (6.5, 7, 7.5, 8, 8.5) to determine when the enzyme is most active.

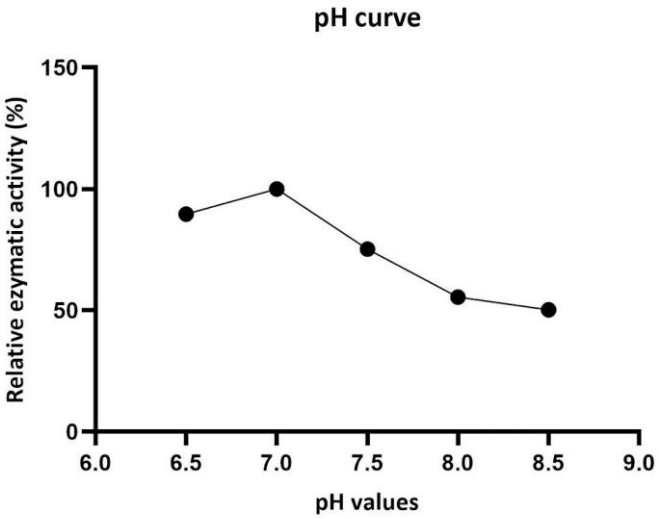



**Figure 17**: The effect of pH on the activity of *Helicoverpa armigera* GGPPS in conditions of 50ug purified enzyme, 50 mM Tris-HCl of varying pH values, 1 mM $MgCl_2$, 0,2 mM MESG, 1 U/mL PNP, 0,03 U/mL Inorganic pyrophosphatase, 150 mM IPP, 50 mM DMAPP, total reaction volume 200 µL at 25°C. In the graph and table, the results are presented as the relative activity (setting the highest activity as 100%) of the enzyme in different pH values.

From the pH curve it is clear that the purified HaGGPPS has a higher activity when the reaction pH is set to 7.0. In pH values of 6.5 and 7.5 the enzyme performance is also relatively high, however as the pH of the reaction becomes more basic (see pH values of 8 and 8.5) the enzymatic activity drops significantly. For all future experiments using this enzymatic activity measurement system, pH 7.0 is selected as it proved to be the most efficient (Figure 17).

The way that this assay is designed, the enzymatic activity calculations are based on the absorbance of the last product of the reaction chain at 360 nm. To quantify the amount of product that HaGGPPS is releasing in a certain time period a standard curve is needed. To create the standard curve, various concentrations of PPi are used. The reaction remains the same as it would be to measure enzymatic activity but the initial substrates (IPP and DMAPP) and HaGGPPS enzyme are missing. This way a correlation between PPi concentration and final absorbance obtained is established.

To obtain the standard curve, PPi concentration varied between 5 µM and 70 µM. After 1 h the PPi concentration was plotted versus the absorbance at 360 nm and the linear part of the equation was selected to determine the range of the absorbance that can be converted to PPi concentration. A linear equation was created which is used to quantify the PPi produced in each time point (Figure 18).

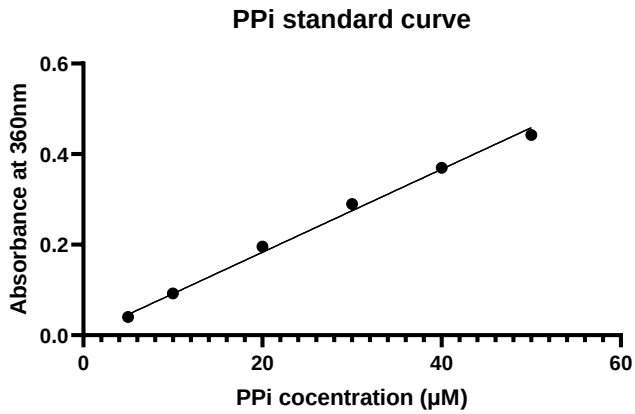



**Figure 18**: Standard curve using PPi substrate. The linear equation y=0.0092x with $R^2$=0.998 is displayed deriving from PPi concentrations of range 5 µM to 50 µM.

### 3.3.2 Kinetic characterization of HaGGPPS

The next step was to determine if the purified enzyme showed consistent activity. For the next experiment all conditions were kept constant (50 mM Tris-HCl pH 7.0, 1 mM $MgCl_2$, 0,2 mM MESG, 1 U/mL PNP, 0,03 U/mL Inorganic pyrophosphatase, 150 mM IPP, 50 mM DMAPP, total reaction volume 200 µL) except from the enzyme concentration. Increasing concentration of the enzyme HaGGPPS should result in increased product formation (given that the substrates were in excess). Enzyme concentrations that demonstrated this fact the best were 10, 20, 40 and 80 µg of protein per 200 µL reaction (Figures 19 and 20).

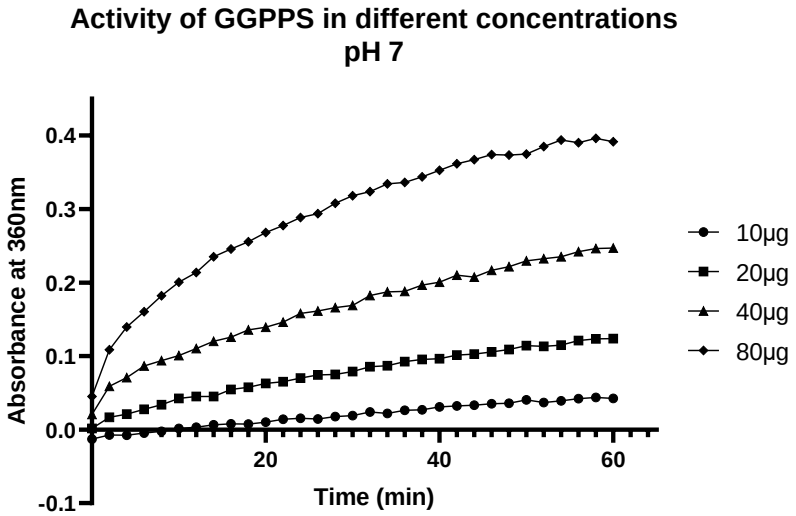



**Figure 19**: PPi produced versus time curve of different GGPPS concentrations. Increasing protein concentration results in increased product production. The absorbance values obtained by the assay were converted to PPi concentration using the standard curve

.

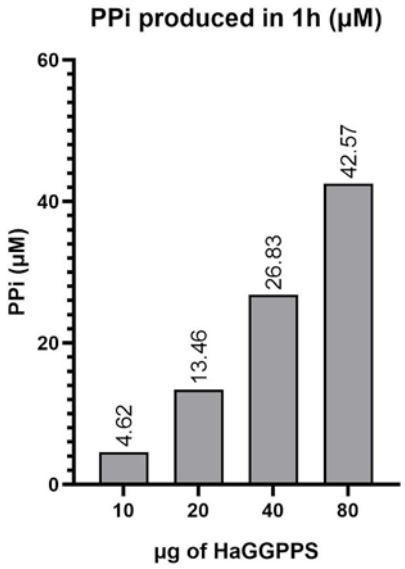



**Figure 20**: PPi produced in 1 h of reaction time. Different GGPPS concentrations are used. Increasing protein concentration results in increased product production. The absorbance values obtained by the assay were converted to PPi concentration using the standard curve.

As expected, the production of PPi increased when HaGGPPS increased. This indicated that the assay was suitable for measuring enzymatic activity. These results were reproducible as this experiment was replicated in different days giving similar results (biological replicates). An interesting finding was also the fact that HaGGPPS was highly stable when stored at -20ºC and there was no activity decrease observed for at least 6 months after protein purification.

As the assay proved to be suitable for measuring GGPPS activity, a more in-depth kinetic study needed to be done. More specifically, Michaelis Menten constant $V_{max}$ and $K_M$, needed to be calculated. The way that the activity assay was set for HaGGPPS, the enzyme produces 1 molecule of GGPP using as substrates 3 molecules of IPP and one molecule of DMAPP. Since there are two substrates, the Km of DMAPP is calculated by keeping the other substrate (IPP) constant and in excess.

In order for $V_{max}$ and $K_M$ values to be calculated considering GGPP as the product of GGPPS, some parameters must be taken into account. Firstly, the products of GGPPS are 3 molecules of PPi and one molecule of GGPP. Since the assay measures PPi production, the results need to be divided by 3 as stoichiometry proposes. HaGGPPS dimeric state should also be considered. However, since each monomer has its own catalytic pocket that is able to catalyze the reaction, and assuming that the reaction happens independently in each monomer, HaGGPPS is treated as a monomer in kinetic calculations.

For kinetic parameter determination, all reagents of the assay remained constant (20 µg HaGGPPS, 50 mM Tris-HCl pH 7.0, 1 mM $MgCl_2$, 0.2 mM MESG, 1 U/mL PNP, 0.03 U/mL Inorganic pyrophosphatase, 450 mM IPP, total reaction volume 200 µL) except from the DMAPP substrate that varied in concentration (25, 50, 100, 150 and 200 µM). For every concentration there were three replicates of the reaction.

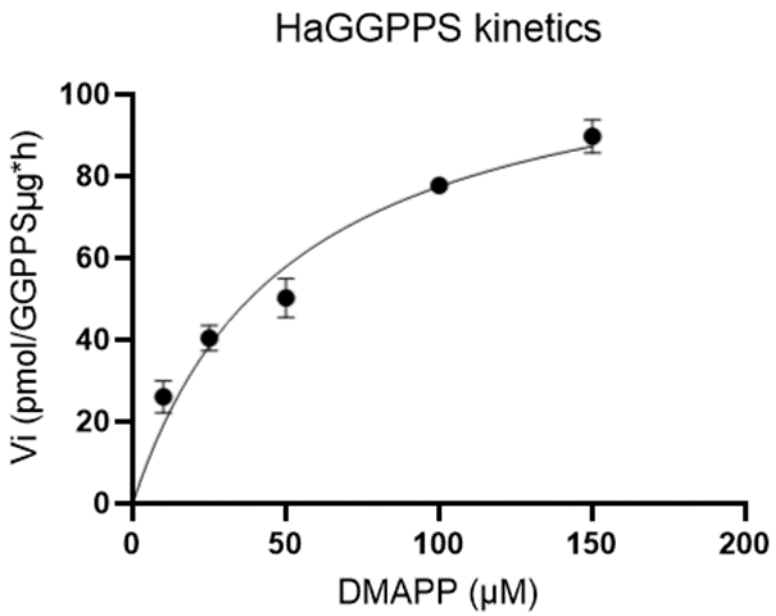



**Figure 21**: Non-linear fitting of activity date of HaGGPPS. Curve and kinetic calculations were obtained using GraphPad Prism. The curve shows the reaction velocity versus increasing concentrations of the substrate (DMAPP). $V_{max}$ and $K_M$ were calculated by GraphPad Prism 117.1 pmol/GGPPSµg*h and 51.03 µM respectively.

The three replicates in general did not have large deviation and were fitted to a curve of the Michaelis-Menten equation Y=Vmax*X/(Km+X) which was also used to calculate the constants $V_{max}$ and $K_M$. $V_{max}$ was found to be 117.1 pmol/GGPPSµg*h and DMAPP $K_M$ 51.03 µM for 20 µg of HaGGPPS enzyme (Figure 21).

## 3.4 HaGGPPS inhibition

### 3.4.1 Testing known inhibitors

Taking the first steps towards understanding how HaGGPPS is inhibited, two known geranylgeranyl diphosphate synthase inhibitors from literature were tested. Zoledronic acid and minodronic acid were both found inhibiting the yeast (*Saccharomyces cerevisiae)* GGPPS by binding in the catalytic pocket imitating the substrate and interacting with the two magnesium cofactors (competitive inhibition).

Since zoledronic acid and minodronic acid both inhibit a HaGGPPS homologous protein and the inhibition mechanism is very well characterized supported by crystallographic data, amino acids found to interact with the inhibitor were compared between the two proteins. All those residues were the same in both proteins, so the inhibitors were tested using the in-vitro protein assay expecting to see inhibition.

The inhibitors were tested in the in-vitro assay keeping all conditions constant (20 µg HaGGPPS, 50 mM Tris-HCl pH 7.0, 1 mM $MgCl_2$, 0.2 mM MESG, 1 U/mL PNP, 0.03 U/mL Inorganic pyrophosphatase, 150 µM IPP, 50 µM DMAPP, total reaction volume 200 µL) with the natural substrates at approximately $K_M$ values and the inhibitors varying in concentration.

**Table 5**: GGPPS inhibition screening performed at pH 7.0. The natural substrates IPP and DMAPP are in concentrations of 50 µM and 150 µM respectively, while the inhibitors are increasing in concentration. HaGGPPS concentration was set at 2.6 µM.

| Inhibitor | Concentration (µM) | % relative activity |
|---|---|---|
| *Zoledronic acid* | 2,6 | 73,9 ± 28,8 |
| | 5,2 | 54,2 ± 5,9 |
| | 10,4 | 20,7 ± 1,8 |
| | 26 | 12,1 ± 4,4 |
| *Minodronic acid* | 2,6 | 33,1 ± 1,7 |
| | 5,2 | 18,6 ± 0,8 |
| | 10,4 | 22,2 ± 1,3 |
| | 26 | 11,4 ± 3,7 |

Inhibitor concentration of 2.6 µM was in a 1:1 ratio with HaGGPPS. The maximum inhibitor concentration tested was ten times the amount of the enzyme which in both cases decreased the enzymatic activity by approximately 90%. The two chemicals had similar inhibition patterns and these results are in agreement with yeast GGPPS inhibition pattern observed in literature. A difference observed between the two chemicals was that minodronic acid inhibited the protein when in lower concentrations while zoledronic acid had to reach concentration of 10.4 µM for over 50% inhibition (table 5).

### 3.4.2 Inhibitor design

Since zoledronic acid and minodronic acid are not the most potent inhibitors, more inhibitors had to be tested. Insect GGPPS could be a very appealing insecticide target so a method for designing new inhibitors had to be followed. Based on structure-activity relationship methods, the crystal structure of *Saccharomyces cerevisiae* GGPPS was used to further examine this dynamic. As a reference, the structure of GGPPS in complex with the substrate isopentenyl pyrophosphate (IPP) (PDB code: 2E8W) was used and all molecular docking simulations were based on it.

For the initial screening, a list of chemical compounds was chosen based on selected characteristics that are present in inhibitors already tested. As shown in figure 23, new chemicals for screening had to have a way of interacting with the $Mg^{2+}$ cofactors, a hydrophobic group and one or more atoms capable of hydrogen bond formation. The library included chemicals with phosphoric groups and some bioisosteric replacements like

(di)carboxylic acids and tetrazoles, several hydrophobic groups of aromatic compounds with varying lengths combined with nitrogen or oxygen atoms (Figure 22).

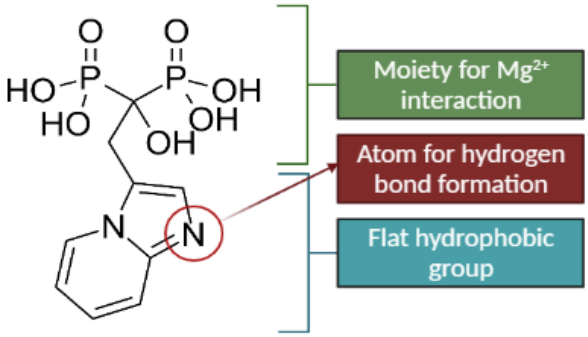



**Figure 22**: Minodronic acid chemical properties responsible for *Saccharomyces cerevisiae* GGPPS inhibition. The phosphoric groups interact with the $Mg^{2+}$ cofactors disturbing correct orientation and function. A flat hydrophobic group is positioned deeper in the catalytic pocket increasing binding energy. Finally, a nitrogen atom forms a hydrogen bond with Thr 170.

This library of chemical inhibitor candidates for *Helicoverpa armigera* GGPPS was initially screened using a molecular docking software to predict the best hits. GOLD, a protein-ligand docking software of Cambridge Crystallographic Data Centre (CCDC) was used for the *in-silico* screening (Figure 23). The chemicals were evaluated based on their given score and manually using protein visualization tool PyMOL. The most promising chemicals of the library were tested using the *in vitro* assay (Figure 24).

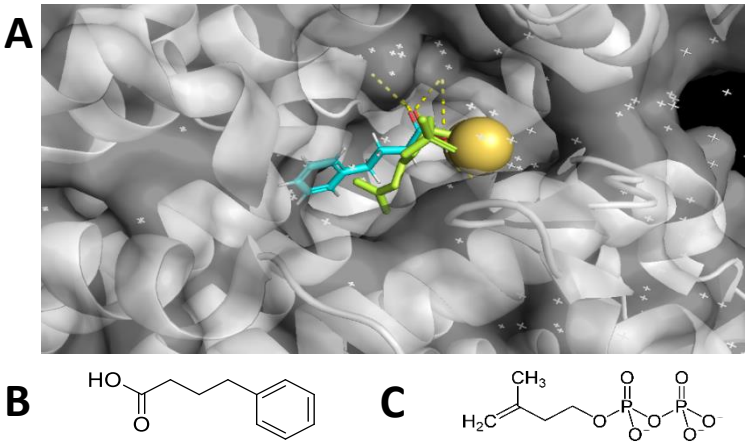



**Figure 23**: A) Crystal structure of *Saccharomyces cerevisiae* GGPPS (PDB code: 2E8W) with the natural substrate IPP (green) and two $Mg^{2+}$ cofactors (yellow) bound. One of the compounds of the initial library (light blue) is docked in the catalytic pocket mimicking the substrate. Docking results are presented in protein visualization software PyMOL. B) Chemical structure of the carboxylic acid tested as a potential inhibitor. C) Chemical structure of the natural substrate IPP.

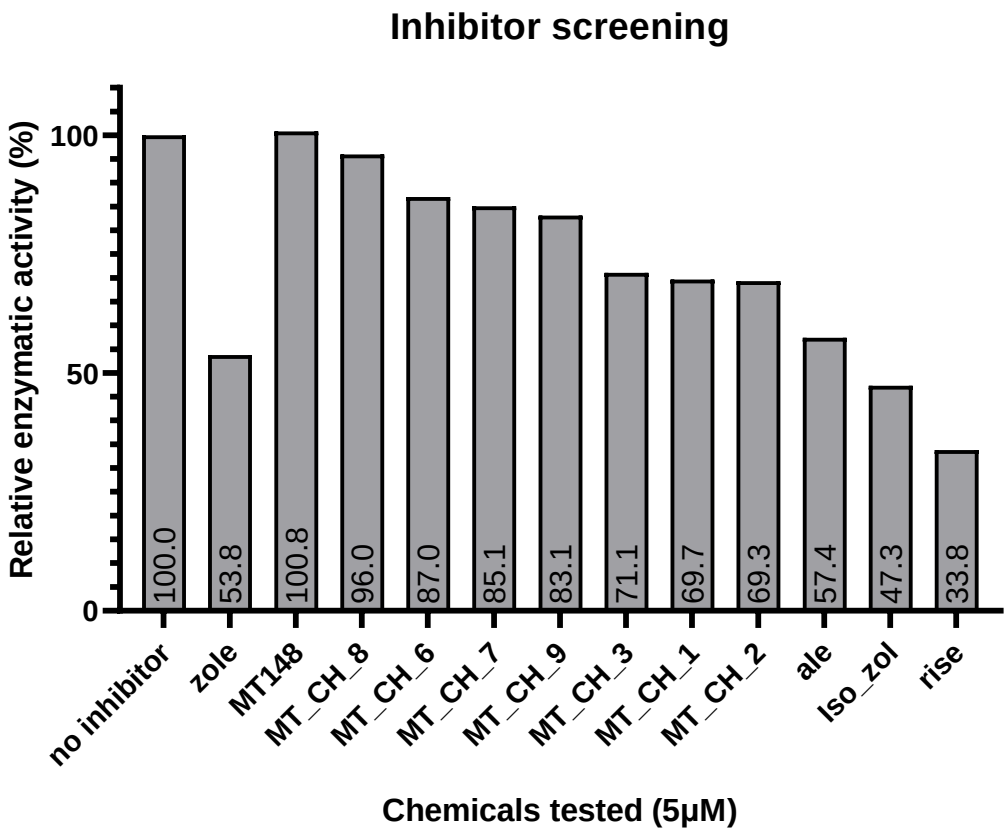
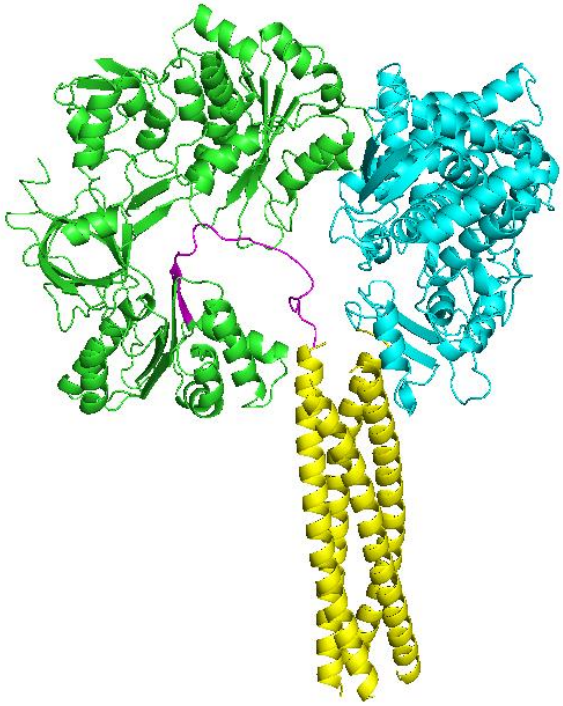



**Figure 24:** Results of the potential inhibitors screening. Relative activity is shown with the no-inhibitor control set at 100%. 5µM of inhibitor is used each time for every chemical. A sample of zoledronic acid (zole) is also included as a positive control. Samples showing at least 30% inhibition are considered interesting candidates for future optimizations.

Results of the screening show that almost all chemicals with similar properties with zoledronic and minodronic acid can inhibit HaGGPPS activity at some level. Interesting cases are Iso-zoledronic acid and riseodronic acid that have higher potency than the already available in literature zoledronic acid.

# CHAPTER 2

## Protein engineering of *Bemisia tabaci* cytochrome P450 for optimal bacterial expression



# 4. Introduction

## 4.1 Cytochrome P450(CYP450) proteins

### 4.1.1 P450 enzymes

Cytochrome P450 proteins are encoded by CYP genes and are present in all organisms (bacteria, protists, plants, fungi, and animals). These proteins are highly diverse as a result of successive gene duplications succeeded by sequence divergence (Sezutsu et al. 2013, Nelson et al. 2018). These enzymes are highly promiscuous as they have been observed to show activity as monooxygenases, oxidases, desaturases, reductases, isomerases, etc. and have been associated with the catalysis of at least 60 distinct chemical reactions (Sezutsu et al. 2013, Nelson et al. 2018, Willingham et al. 2004). The catalytic mechanism always involve heme, which is bound to the enzyme and forms a FeII-CO complex essential for catalysis that has a characteristic ~450 nm absorbance peak (Sezutsu et al. 2013, Nelson et al. 2018, Feyereisen, 2020).

### 4.1.2 Insect P450 enzymes

Cytochrome P450 monooxygenases (P450s) present in insects are associated with pathways producing metabolites essential for their growth and development. In order to feed, herbivorous insects must bypass plant defenses. These molecular defenses often involve a variety of secondary metabolites that are toxic to the insects. Through evolutionary pressure, insects respond to such chemicals by avoiding or metabolizing them. The best-known system for metabolizing toxic chemical products is the system of polysubstrate monooxygenases, with the center component being cytochrome P450. The main mechanism of function is the direct binding of toxic substates to the P450 enzyme. P450s can identify and metabolize (bind) a variety of substrates due to many isoenzymes being present in one organism that have different specificities (Chapman, 1998).

Detoxicative P450s are frequently evolved from those with catabolic function. Even though P450 isoenzymes can identify a variety of toxic chemicals, not all insects can metabolize the plethora of plant toxins. Depending on the evolutionary pressure, herbivorous insects adapt to specific plants establishing specific plant-host relationships (Ortiz de Montellano, 2015). There are also examples of host shifts of insects resulting from different expression levels of certain P450 enzymes or even from mutations leading to new specificities. These genetic differences allow insects to increase their fitness on novel hosts leading to diversification, reproductive isolation and thus initiating speciation. The direct involvement of P450s in this process highlights the importance of their natural roles (Singh et al. 2020).

### 4.1.3 Classification of P450 enzymes based on the redox partner

For a P450 catalytic system to function, the P450 enzyme that contains a heme-iron, the substrate, the redox partner(s) that function as an electron transfer shuttle and the cofactor (NAD(P)H) need to be found in proximity. This is naturally achieved either by membrane docking or by fusion of the components (Li, Zhong, et al. 2020).

Regardless of the number of redox partners or the system's intracellular location, most P450s share a common catalytic mechanism. The P450 enzyme first binds a substrate (RH) by displacing a water molecule. The ferric iron of the P450-substrate complex is then reduced to ferrous iron by one electron transferred via a redox partner. Next, dioxygen binds to the ferrous iron of the complex. Then, the complex is reduced by a second electron and uses a proton from solvent to generate ferric hydroperoxo species. The bond between the oxygens is then cleaved upon the addition of a second proton and releases a molecule of water generating a high-valent porphyrin radical cation tetravalent iron. This highly reactive complex abstracts a hydrogen atom from the substrate forming a hydroxylated product (R-OH) and releasing it. Finally, a molecule of water returns to coordinate with the ferric iron restoring the P450 enzyme to its initial state (Li, Zhong, et al. 2020).

Continuous electron transport to the heme-iron by redox partners is needed to maintain the P450 catalytic cycle. Based on the types of redox partners and the P450-redox partner interaction relationships, P450 systems can be divided into five main classes. The Class I P450 system is present in most bacterial and mitochondrial P450s and has a two-component redox partner system, comprised of an FAD-containing ferredoxin reductase (FdR) and a small iron-sulfur-containing ferredoxin (Fdx). The Class II P450 system employed by eukaryotic organisms has a single-component redox partner, which is a membrane-bound protein containing both an FAD and an FMN domain, termed cytochrome P450 reductase (CPR). This class is very common among insects and the system is usually incorporated into the endoplasmic reticulum. Class III P450 systems have a eukaryotic-like CPR naturally fused to the C terminus of the P450 domain through a flexible linker. Class IV P450 systems have the FMN/Fe2S2-containing reductase domain forming a natural fusion with the P450 domain. Finally, a few P450s can directly interact with their electron donors and are independent of additional redox partner proteins to accomplish the catalytic reactions; these are Class V P450s (Figure 25), (Li, Zhong, et al. 2020).

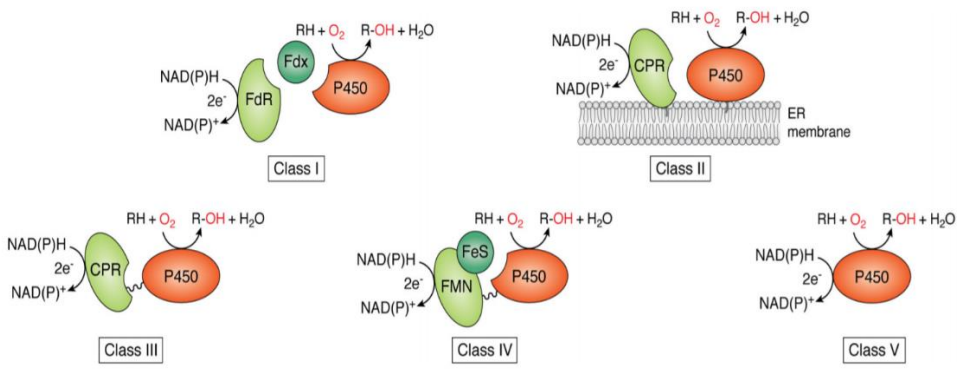



**Figure 25:** Schematic representation of classification of the P450 systems based on redox partner proteins. Figure adapted from Li, Zhong, et al. 2020.

### 4.1.4 Structure-Function relationship

Structural data of P450 enzymes are limited due to problematic behavior in lab conditions. Specifically, there are reported many efficiency problems when expressed, as complex cofactors must be integrated to maintain functional catalysis and participating proteins need to match each other in terms of concentration, stability, affinity, and activity. In addition, many of them, especially those having transmembrane regions, have solubility problems when expressed in most expression systems (Beyer et al. 2016).

In 2014 there were only 54 unique structures deposited in structural data bases. However, by now, there are plenty of known structures to safely state that the overall P450-fold is quite conservative. Even though the overall fold remains the same there are many variations especially in regions that are known to interact with the substrate, as expected. On the contrary, the heme domain is the most conserved area (Ortiz de Montellano, 2015).

For the P450 enzyme to function normally, some protein regions must stay intact. Helices I and L are in contact with the heme and thus having the native fold is necessary. Just prior to the L helix there is the β-bulge segment that contains the Cys ligand. This part of the protein is very rigid in order to both protect the Cys ligand and hold it in place to be within H-bonding distances of two peptide NH group. This amino acid and its orientation are of extreme importance as these H-bonds that it creates aid in regulation of the heme iron redox potential. Another highly conserved region lies in the portion of helix I near the heme domain and is associated with the $O_2$ activation. There, a polar amino acid (usually a Thr but it varies) is involved in local helical distortion such that the side chain OH donates an H-bond to a peptide carbonyl oxygen that would normally be involved in an α-helical H-bond. helix This arrangement is thought to be quite important for the proper delivery of protons to the iron-linked oxygen required for cleavage of the O–O bond, thus generating the active Fe–O hydroxylating species (Figure 26). The domain responsible for substrate specificity contains the B' helix that has to be present as well, although the sequence and orientation varies among proteins.

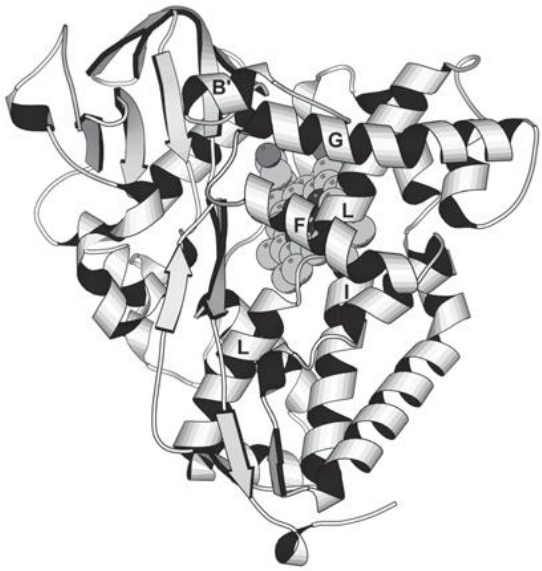



**Figure 26:** The structure of P450cam (PDB: 5CP4) with key helical segments labeled. Adapted from Ortiz de Montellano, 2015.

For eukaryotic transmembrane P450s, the membrane region is important for the enzyme to be docked in a specific orientation. This domain contains 30-50 amino acids (depending on the organism) that form an α-helix to reduce energetic cost of placing a polar backbone in a hydrophobic core and are located in the N-terminal region of the protein. This transmembrane region has nothing to do with the active center and its only function is to target the membrane. Following these transmembrane amino acids there are about 10, usually positively charged, amino acid residues, that connect the transmembrane region with the rest of the enzyme. The positively charged amino acid residues keep the meta-membrane region of the protein from sinking into the membrane (Ortiz de Montellano, 2015).

### 4.1.5 P450 mediated insecticide resistance

Insect P450s are implicated in detoxification of exogenous compounds like insecticides (Feyereisen, 2010, Scott, 1999). P450 enzymes can result in resistant insect populations when the gene is upregulated or has mutation(s) that help with compound identification or improved catalysis (Terriere, 1984, Zhu et al. 2008). In fact, multiple P450s in an insect that are co-expressed and or upregulated is a very common resistance mechanism observed in several cases (Zhu, Liu, 2008, Festucci-Buselli et al. 2005, Liu et al. 2011) making the characterization of such enzymes necessary for the field of molecular entomology.

Characterization of such enzymes is achieved with multiple approaches. Synergistic studies is usually the first step where using chemical such as piperonyl butoxide (PBO), S,S,S,-tributyl phosphorotrithioate (DEF), and diethyl maleate (DEM), which are inhibitors of cytochrome P450 monooxygenases, hydrolases, and glutathione S-transferases (GSTs), respectively, the general resistance mechanism is revealed (Cochran, 1994). In case of P450 mediated resistance the responsible gene is then identified by gene expression studies where overexpression of P450 enzymes in resistant strains in contrast to the susceptible is a strong indicator that this P450 is directly involved in resistance (Feyereisen, 2011, Yang, Liu, 2011, Itokawa et al. 2010).

Many attempts have been also made towards heterologous expression of P450 enzymes responsible for resistant phenotypes and In-vitro assays. These studies are less frequent as P450s are challenging to be expressed in expression systems due to their properties (heme binding proteins, high hydrophobicity/membrane docking, redox partner, complex folding). However, there are successful cases where P450 enzymes are heterologous expressed in E. coli systems (Stevenson et al. 2012, Zhang et al. 2010) and baculovirus expression system using Sf9 cells (Tiwari et al 2010, Mao et al. 2011).

In silico 3D P450 homology modeling and insecticide molecular docking is also an approach used to rationalize the detoxification mechanism by prediction tools. Since there are no P450 enzymes from insects currently available in protein structure databases modeling is based on homology models generated. These techniques are also used to evaluate allelic variants (Schuler et al. 2013, Hiratsuka, 2012)

## 4.2 Protein ROP and the RM6 mutant

### 4.2.1 Repressor of primer protein (ROP)

In nature, protein ROP binds to RNA and is involved in the regulation of ColE1 plasmid copy number. The structure was solved via X-ray crystallography and Nuclear Magnetic Resonance (NMR) and it is referenced as an example of a 4-α-helical bundle with antiparallel α-helices. Specifically, it forms a homodimer with each monomer consisted of two α-helices connected with three amino acid residue turn that brings them in an antiparallel position (Amprazi et al., 2014).

Due to ROP having a relatively simple tertiary structure it was used as a model for the study of correlation between the amino acid sequence and the secondary pattern of a protein. That had as a result the design and characterization (structure, function, thermodynamics, and kinetics) of many of its mutants.

### 4.2.2 RM6 mutant

RM6 is the result of a five amino acid deletion from the turn region of the wild type protein ROP (15 DNA base pairs without frame shift). The amino acid residues removed were from Asp30 to Gln34 (DADEQ). This mutant structure was also solved using X-ray crystallography and its properties were studied in detail (Amprazi et al., 2014, Glykos et al., 2006, Lassalle et al., 1998).

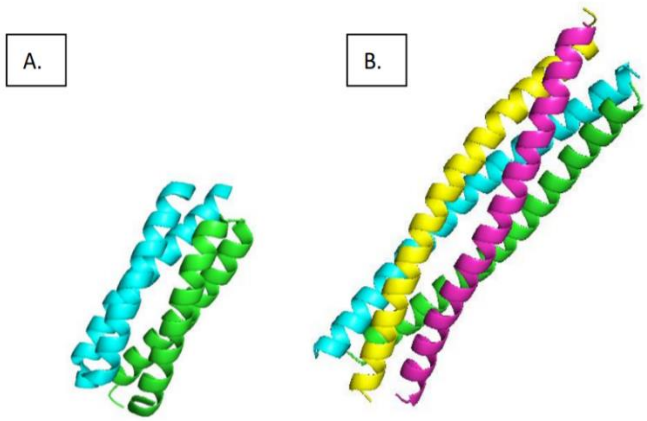



**Figure 27:** (A) The homodimer ROP presented using PyMOL (PDB code: 1ROP). Each monomer has a structure of helix-turn-helix. (B) The homotetrameric mutant, RM6 presented using PyMOL (PDB code: 1QX8).

In contrast with ROP, RM6 is an homotetrameric all-antiparallel four-α-helical bundle. Its molecular surface is a slightly curved cylinder 78 Å in length and 24 Å in diameter (Figure 27). Each helix twists around one another forming a four-stranded left-handed coiled coil. Even though RM6 is a tetramer, the intermolecular interactions are between each of the

opposite chains and for that reason it is best described as a dimer of dimers. The tetramer has a hydrophobic core with no buried cavities. This together with the extent of the buried accessible area and the presence of eight interhelix salt bridges explain its hyper thermostability. Its electrostatic potential is very polar with the center of the molecule being negatively charged and the edges being positively charged, something not observed in the wild type ROP (Glykos et al., 2006).

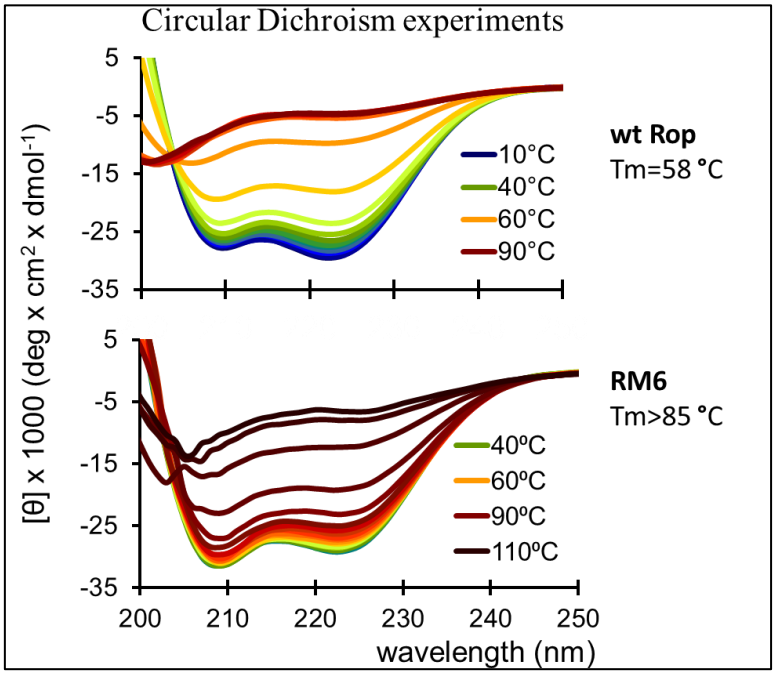



**Figure 28:** Circular Dichroism experiment results for ROP and RM6 mutant. Figure adapted from Amprazi et al., 2014.

Circular dichroism experiment results show that RM6 has an impressive thermostability with a melting point of 92ºC (Figure 28). In addition, size exclusion chromatography shows that RM6 exists in a single tetrameric state (maintaining this property when various buffers were tested) and β-mercaptethanol has no effect on the molecule. Its molecular weight was calculated as 6.5 kDa per monomer (approximately a 26 kDa tetramer) (Amprazi et al., 2014).

### 4.2.3 Applications of scRM6

The physicochemical characteristics of RM6 (thermostable, soluble, easy to express) makes the protein good candidate for protein engineering. RM6 can be fused with proteins that have expression problems because of aggregation, formation of inclusion bodies or inclusion of hydrophobic surfaces/ cavities. In this case, RM6 has the role of a solubility tag and can make the target protein behave better when expressed.

In addition, RM6 can be used for Homo- / Hetero- dimer stability. Protein fusions using RM6 as a scaffold, will have increased stability and/or efficiency, solving problems like high Kd complexes or unwanted homo-dimer tendency (when the study of the hetero-dimer is desired). The fact that the α' helices have an antiparallel orientation plays an important role for the design and fusion of proteins. RM6 forms tetramers with both top and bottom ends having two N-terminal and two C-terminal sites that alternate forming the structure presented below (Figure 29).

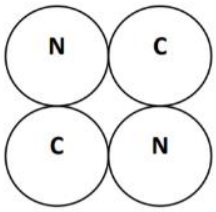



**Figure 29:** Bottom view of RM6 and the retribution of the N- and C-terminal ends.

Another possible application for RM6 is for coupled reactions. It is well known that enzymes that are involved in the same pathway are found fused in some organisms and as single, full-length proteins in other species (Enright et al., 1999). This has been observed in evolution, perhaps the most widely known example being the fusion of tryptophan synthetase α- and β-subunits from bacteria to fungi (Burns et al., 1990). This phenomenon is also observed in the different types of P450 reaction systems as mentioned before (P450 fused with CPR in some cases naturally) (Li, et al., 2019). Specifically, because P450s are not well-behaved proteins when expressed in bacterial expression systems, there have been some efforts of creating fusion to solve that problem (Aalbers et al., 2019). However, the fusion is usually achieved by a short linker between the two enzymes and even though there are cases of fusions being functional, the efficiency is generally low.

## 4.3 The purpose of this work

Functional expression of cytochrome P450s comes with a lot of challenges. To address those challenges, baculovirus expression system generating recombinant proteins in insect cells (Sf9 cell lines) is most commonly used. Crude extracts show activity against model substrates, but the system is time consuming and expensive. Attempts to fuse cytochrome P450 with cytochrome P450 reductase (CPR) have also been made to improve catalytic activity by facilitating electron transfer.

Bacterial expression system is simple and inexpensive compared to other expression system. *Bemisia tabaci* P450 and *Musca domestica* CPR are engineered to explore bacterial expression, and purification, in order to achieve higher expression yield while still maintaining function. For the purpose of this work, both proteins were N-terminally truncated to

potentially resolve solubility issues by removing the membrane association part of the proteins.

In attempt to achieve higher catalytic efficiency, *Bemisia tabaci* P450 and *Musca domestica* CPR were also used to design a fusion model. This theoretical construct involved both proteins to be fused using scRM6 as a scaffold. Because the fusion of two proteins in each of RM6 monomers would result in the placement of four proteins (two P450 and two CPR) in every side of the scaffold, scRM6 is used as a simpler model. This molecule retains the general structure of the four antiparallel α' helices with the difference being that the more flexible ends of the monomers are linked together creating a single monomer (the sequence of the RM6 monomer is repeated four times). The design was based on the idea that the fusion of the two proteins in the N- and C-terminal sites of scRM6 would bring them in proximity for interaction.

# 5. Protocols and results

**<u>Protocol 1:</u> Expression and purification of ΔNBtCYP6CM1**
**(Sonication, no-detergent, $Ni^{2+}$-NTA purification)**

The first purification attempt of the N-terminal truncated BtCYP6CM1 (ΔNBtCYP6CM1) was by a standard protocol to observe expression and purification patterns and to understand limitations. Briefly, the soluble fraction, after cell lysis, was loaded into a $Ni^{2+}$-NTA Qiagen affinity column without the use of detergents.

**Expression**

*E. coli* BL21(DE3) Star cells (selected from expression tests using five different expression strains in small scale) were transformed with 50 ng of pCW_ΔN-BtCYP6cm1-Flag-His plasmid. A single colony was selected and left to grow on 10 mL LB liquid culture, supplemented with ampicillin (amp), for 16 h (preculture).

5 x 200 mL TB liquid cultures (supplemented with 100 µg/mL amp) were inoculated with 5 mL of preculture each (1:50 inoculation). Cultures were incubated at 30°C and monitored for their OD at 600 nm. When OD reached 0.5, IPTG and ALA were added at final concentration of 0,5 mM and 1 mM correspondingly. After induction, incubation was continued at 25°C for 24 h, shaking. One mL culture sample, before and after 24 h induction (-IPTG, +IPTG), was collected in order to check expression. Cultures were centrifuged at 6000 rpm for 25 min at 4°C. Cell pellet was weighed and stored at -20°C.

**Cell lysis**

Cell pellets stored at -20°C were thawed out, using 30 mL ice cold lysis buffer, mixing by pipetting and gentle vortexing. After the samples were completely homogenized, 5 mM β-mercaptethanol, 1 mM EDTA and 2 mM PMSF were added. The sample was then sonicated for 15 min (20'' pulse – 40'' rest). The precipitates were subsequently removed after centrifugation at 18.000 rpm, 4°C for 30 min. Supernatants were kept on ice until they were loaded to a $Ni^{2+}$-NTA Qiagen affinity column.

**Purification via $Ni^{2+}$-NTA affinity chromatography**

$Ni^{2+}$-NTA affinity chromatography was carried out at 4°C. P450 lysates (from previous cell lysis step), supplemented also with 5 mM imidazole, were loaded on 200 µL of pre-equilibrated $Ni^{2+}$-NTA resin (pre-equilibrated with 10 column volumes equilibration buffer) with a flow rate 0.5 mL/min and Flow Through was collected. Then the column was washed with 10 column volumes (10 * 0,2 mL = 2 mL) of Wash 1 and collected, followed by wash with 10 c.v. (2 mL) of Wash 2. Then, the bound proteins were eluted with 5 c.v. (1 mL) of elution buffer and 5 elution fractions, 200 µL each, were collected.

The protein samples ran on a 10% SDS-PAGE. Upon separation the gel was stained with Coomassie Brilliant Blue to estimate protein purity or wet transferred to nitrocellulose membrane for specific detection using an anti-his antibody.

**Buffers used:**

**Lysis buffer**: 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% glycerol

**Equilibration buffer**: 50 mM Tris-HCl pH 75, 200 mM NaCl, 10% glycerol

**Wash 1**: 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% glycerol, 10 mM imidazole, 5mM β-mercaptoethanol

**Wash 2**: 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% glycerol, 20 mM imidazole, 5mM β-mercaptoethanol

**Elution buffer**: 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 10% glycerol, 200 mM imidazole, 5mM β-mercaptoethanol

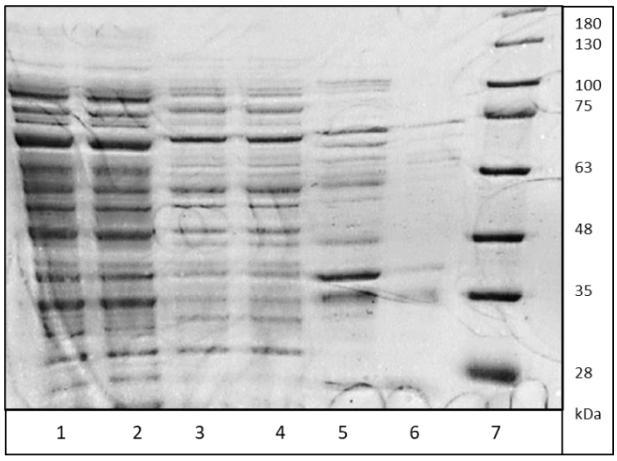



**Figure 30:** Purification attempt of ΔNBtCYP6cm1-his. Total bacterial extracts were loaded in a $Ni^{2+}$-NTA affinity resin. Proteins were separated with SDS-PAGE (10%) and stained with Coomassie Brilliant Blue – Calculated MW of ΔNBtCYP6cm1-his is 58 kDa. Lane 1: total protein loading material, Lane 2: unbound proteins, Lane 3: first wash, Lane 4: second wash, Lane 5: first elution fraction of bound proteins, Lane 6: second elution fraction of bound proteins, Lane 7: molecular weight marker.

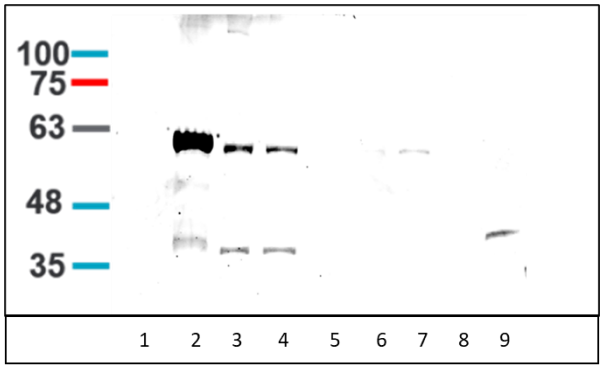



**Figure 31:** Purification attempt of ΔNBtCYP6cm1-his. Total bacterial extracts were loaded in a $Ni^{2+}$-NTA affinity resin. Proteins were separated with SDS-PAGE (10%) and transferred to a nitrocellulose

membarne for immunoblotting. Anti-his antibody was used for specific detection – Calculated MW of ΔNBtCYP6cm1-his is 58 kDa. Lane 1: before induction bacterial extract, Lane 2: after induction bacterial extract, Lane 3: total protein loading material, Lane 4: unbound proteins, Lane 5: first wash, Lane 6: second wash, Lane 7: first elution fraction of bound proteins, Lane 8: second elution fraction of bound proteins, Lane 9: 50 ng of antibody control.

According to the results of this experiment, the protein is expressed in BL21(DE3) Star bacterial cells (verified by western blot using anti-his antibody). However, the protein seems to have a weak affinity to the Nickel resin resulting in the loss of protein to the unbound protein fraction (flowthrough) of the chromatography. Since the protein was not eluted in the final fraction (elution fraction) as expected, more trials and protocol optimization followed.

**Protocol 2: Expression and purification of BtCYP6CM1**

**(Sonication, Triton X-100, $Ni^{2+}$-NTA purification)**

A second purification attempt was done to observe the effect of Triton X-100, a non-ionic detergent, on the purification profile of the protein. Again, the soluble fraction, after cell lysis, was loaded into a Ni2+-NTA Qiagen affinity column in the presence of Triton X-100.

**Expression**

*E. coli* BL21(DE3) Star cells (selected from expression tests using five different expression strains in small scale) were transformed with 50 ng of pCW_ΔN-BtCYP6cm1-Flag-His plasmid. A single colony was selected and left to grow on 10 mL LB liquid culture, supplemented with ampicillin (amp), for 16 h (preculture).

A 200 mL TB liquid culture (supplemented with 100 μg/mL amp) was inoculated with 5mL of preculture (1:50 inoculation). Culture was then incubated at 30$^{o}$C and monitored for its OD at 600 nm. When OD reached 0.5, IPTG and ALA were added at final concentration of 0,5 mM and 1 mM correspondingly. After induction, incubation was continued at 30$^{o}$C for 24 h, shaking. One mL culture sample, before and after 24 h induction (-IPTG, +IPTG), was collected in order to check expression. Culture was centrifuged at 6000 rpm for 25 min at 4$^{o}$C. Cell pellet was weighed and stored at -20$^{o}$C

**Cell lysis**

Cell pellets stored at -20$^{o}$C were thawed out, using 30 mL ice cold lysis buffer, mixing by pipetting and gentle vortexing. After the samples were completely homogenized, 5 mM β-mercaptoethanol, 1 mM EDTA, 2 mM PMSF and 0.4% Triton X-100 were added. The sample was then sonicated for 15 min (20'' pulse – 40'' rest). The precipitates were subsequently removed after centrifugation at 18.000 rpm, 4$^{o}$C for 30 min. Supernatants were kept on ice until they were loaded to a $Ni^{2+}$-NTA Qiagen affinity column.

**Purification via $Ni^{2+}$-NTA affinity chromatography**

$Ni^{2+}$-NTA affinity chromatography was carried out at 4°C. P450 lysates (from previous cell lysis step), supplemented also with 5 mM imidazole, were loaded on 200 µL of pre-equilibrated $Ni^{2+}$-NTA resin (pre-equilibrated with 10 column volumes equilibration buffer) with a flow rate 0,5 mL/min and Flow Through was collected. The column was washed with 10 column volumes (10* 0,2 mL= 2 mL) of Wash 1 and collected. Then, the bound proteins were eluted with 5 c.v. (1 mL) of elution buffer and 5 elution fractions, 200 µL each, were collected.

The protein samples ran on a 10% SDS-PAGE. Upon separation the gel was stained with Coomassie Brilliant Blue to estimate protein purity or wet transferred to nitrocellulose membrane for specific detection using an anti-his antibody.

**Buffers used**

**Lysis buffer**: 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% glycerol

**Equilibration buffer**: 50 mM Tris-HCl pH 75, 200 mM NaCl, 10% glycerol, 0,4% Triton X-100

**Wash 1**: 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% glycerol, 5 mM imidazole, 5mM β-mercaptoethanol, 0,4% Triton X-100

**Elution buffer**: 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% glycerol, 150 mM imidazole, 5 mM β-mercaptoethanol, 0,4% Triton X-100

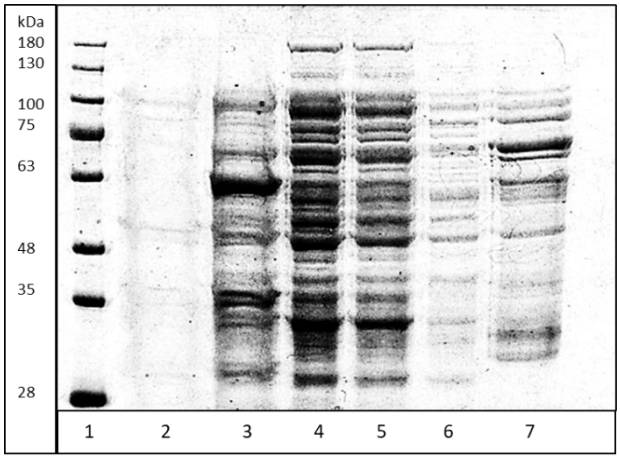



**Figure 32:** Purification attempt of ΔNBtCYP6cm1-his in the presence of Triton X-100. Total bacterial extracts were loaded in a $Ni^{2+}$-NTA affinity resin. Proteins were separated with SDS-PAGE (10%) and stained with Coomassie Brilliant Blue – Calculated MW of ΔNBtCYP6cm1-his is 58 kDa. Lane 1: molecular weight marker, Lane 2: before induction bacterial extract, Lane 3: after induction bacterial extract, Lane 4: total protein loading material, Lane 5: unbound proteins, Lane 6: first wash, Lane 7: second elution fraction of bound proteins

In this experiment, the protein is successfully expressed in the bacterial cells but again it is not eluted in the final purification step. The protein is also not present in high amounts in

the loading fraction indicating that BtCYP6CM1 is not only lost to the flowthrough. These results indicated that the protein is not completely soluble and thus is partially lost when membrane and soluble fractions are separated after cell lysis.

**Protocol 3: Expression and purification of BtCYP6CM1**

**(Sonication, Triton X-100, larger scale, $Ni^{2+}$-NTA purification)**

In order to test protein solubility, the next purification attempt was done using the soluble and membrane fraction after cell lysis. The experiments were performed in parallel using the same cell pellet.

**Expression**

*E. coli* BL21(DE3) Star cells (selected from expression tests using five different expression strains in small scale) were transformed with 50 ng of pCW_ΔN-BtCYP6cm1-Flag-His plasmid. A single colony was selected and left to grow on 10 mL LB liquid culture, supplemented with ampicillin (amp), for 16 h (preculture).

5 x 200 mL TB liquid cultures (supplemented with 100 μg/mL amp) were inoculated with 5 mL of preculture each one (1:50 inoculation). Cultures were incubated at 30°C and monitored for their OD at 600 nm. When OD reached 0.5, IPTG and ALA were added at final concentration of 0.5 mM and 1 mM correspondingly. After induction, incubation was continued at 25°C for 24 h, shaking. One mL culture sample, before and after 24 h induction (-IPTG, +IPTG), was collected in order to check expression. Cultures were centrifuged at 6000 rpm for 25 min at 4°C. Cell pellet was weighed and stored at -20°C.

**Cell lysis**

Cell pellets stored at -20°C were thawed out, using 70mL ice cold lysis buffer, mixing by pipetting and gentle vortexing. After the samples were completely homogenized, 5 mM β-mercaptethanol, 1 mM EDTA and 2 mM PMSF were added. The sample was then sonicated for 15 min (20'' pulse – 40'' rest). Intact cells were removed by a 10-min centrifugation at 4.000 rpm for 10 min. The soluble fraction was centrifuged at 18.000 rpm, 4°C for 30 min. Supernatant was kept on ice until it was loaded to a Ni-NTA Qiagen column. The membrane fraction was resuspended (with homogenizer) in lysis buffer with the addition of 0.5% Triton X-100 and centrifuged at 18.000 rpm, 4°C for 30 min. Again, supernatants were kept on ice until they were loaded to a $Ni^{2+}$-NTA Qiagen affinity column.

**Purification via $Ni^{2+}$-NTA affinity chromatography**

$Ni^{2+}$-NTA affinity chromatography was carried out at 4°C. P450 lysates (from previous cell lysis step), supplemented also with 5 mM imidazole, were loaded on 1 mL of pre-equilibrated $Ni^{2+}$-NTA resin (pre-equilibrated with 10 column volumes equilibration buffer) with a flow rate 0,5 mL/min and Flow Through was collected. The column was washed with 10 column volumes (10* 1 mL = 10 mL) of Wash 1 and collected, followed by wash with 10 c.v.

(10 mL) of Wash 2 (**W2**). Then, the bound proteins were eluted with 5 c.v. (5 mL) of elution buffer and 10 elution fractions, 500 µL each, were collected. The above purification steps were done twice, one for the soluble fraction and one for the membrane fraction.

The protein samples ran on a 10% SDS-PAGE. Upon separation the gel was stained with Coomassie Brilliant Blue to estimate protein purity.

**Buffers used**

**Lysis buffer**: 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% glycerol

**Equilibration buffer**: 50 mM Tris-HCl pH 75, 200 mM NaCl, 10% glycerol, 5 mM β mercaptoethanol

**Wash 1**: 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% glycerol, 20 mM imidazole, 5 mM β mercaptoethanol.

**Wash 2**: 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% glycerol, 40 mM imidazole, 5 mM β mercaptoethanol.

**Elution buffer**: 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% glycerol, 150 mM imidazole, 5 mM β mercaptoethanol.

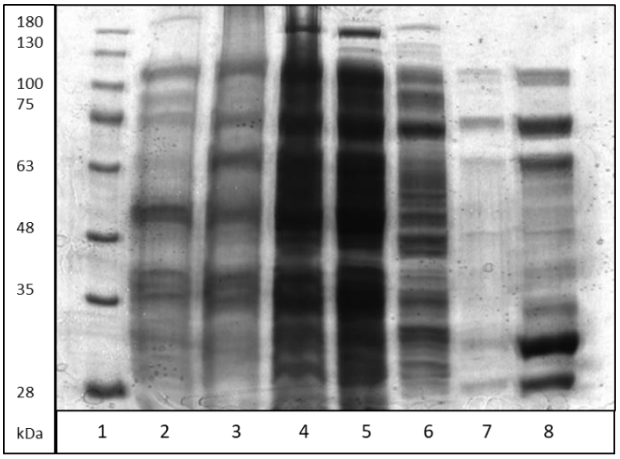



**Figure 33:** Purification attempt of ΔNBtCYP6cm1-his in the soluble fraction. Total bacterial extracts were loaded in a $Ni^{2+}$-NTA affinity resin. Proteins were separated with SDS-PAGE (10%) and stained with Coomassie Brilliant Blue – Calculated MW of ΔNBtCYP6cm1-his is 58 kDa. Lane 1: molecular weight marker, Lane 2: before induction bacterial extract, Lane 3: after induction bacterial extract, Lane 4: total protein loading material, Lane 5: unbound proteins, Lane 6: first wash, Lane 7: second cash, Lane 8: second elution fraction of bound proteins

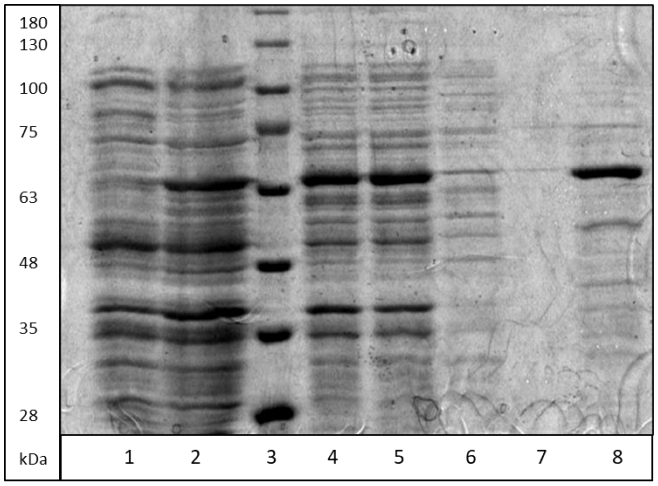



**Figure 34:** Purification attempt of ΔNBtCYP6cm1-his in the membrane fraction. Total bacterial extracts were loaded in a $Ni^{2+}$-NTA affinity resin. Proteins were separated with SDS-PAGE (10%) and stained with Coomassie Brilliant Blue – Calculated MW of ΔNBtCYP6cm1-his is 58 kDa. Lane 1: before induction bacterial extract, Lane 2: after induction bacterial extract, Lane 3: molecular weight marker Lane 4: total protein loading material, Lane 5: unbound proteins, Lane 6: first wash, Lane 7: second wash, Lane 8: second elution fraction of bound proteins

From this trial it is clear that the majority of the protein is not soluble. It is speculated that this happens because of other hydrophobic interactions that still assosiate the protein to the membrane (apart from the trancated helical anchor) or even due to the formation of inclusion bodies.

**Protocol 4: Expression and purification of BtCYP6cm1**

**(Sonication, CHAPS, larger scale, $Ni^{2+}$-NTA purification)**

Next, a protocol using a different non-ionic detergent was also used for potentially better results.

**Expression**

*E. coli* BL21 Star DE3 cells (selected from expression tests using five different expression strains in small scale) were transformed with 50 ng of pCW_ΔN-BtCYP6cm1-Flag-His plasmid. A single colony was selected and left to grow on 10 mL LB liquid culture, supplemented with ampicillin (amp), for 16 h (preculture).

5 x 200 mL TB liquid cultures (supplemented with 100 μg/mL amp) were inoculated with 5 mL of preculture each one (1:50 inoculation). Cultures were incubated at 30°C and monitored for their OD at 600 nm. When OD reached 0.5, IPTG and ALA were added at final concentration of 0.5 mM and 1 mM correspondingly. After induction, incubation was continued at 25°C for 24 h, shaking. One mL culture sample, before and after 24 h induction (-IPTG, +IPTG), was collected in order to check expression. Cultures were centrifuged at 6000 rpm for 25 min at 4°C. Cell pellet was weighed and stored at -20°C.

### Cell lysis

Cell pellets stored at -20$^{o}$C were thawed out, using 70 mL ice cold lysis buffer, mixing by pipetting and gentle vortexing. After the samples were completely homogenized, 5 mM β mercaptethanol, 1 mM EDTA, 2 mM PMSF and 0.25% CHAPS were added. The sample was then sonicated for 15 min (20'' pulse – 40'' rest). The precipitates were subsequently removed after centrifugation at 18.000 rpm, 4$^{o}$C for 30 min. Supernatants were kept on ice until they were loaded to a $Ni^{2+}$-NTA Qiagen affinity column.

### Purification via $Ni^{2+}$-NTA affinity chromatography

$Ni^{2+}$-NTA affinity chromatography was carried out at 4$^{o}$C. P450 lysates (from previous cell lysis step), supplemented also with 5 mM imidazole, were loaded on 400 μL of pre-equilibrated $Ni^{2+}$-NTA resin (pre-equilibrated with 10 column volumes equilibration buffer) with a flow rate 0.5 mL/min and Flowthrough was collected. The column was washed with 10 column volumes (10 * 0.4 mL = 4 mL) of Wash 1 and collected, followed by wash with 10 c.v. (4 mL) of Wash 2. Then, the bound proteins were eluted with 5 c.v. (2 mL) of elution buffer and 10 elution fractions, 200 μL each, were collected.

The protein samples ran on a 10% SDS-PAGE. Upon separation the gel was stained with Coomassie Brilliant Blue to estimate protein purity.

**Buffers used**

**Lysis buffer**: 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% glycerol

**Equilibration buffer**: 50 mM Tris-HCl pH 75, 200 mM NaCl, 10% glycerol, 5 mM β mercaptoethanol, 0,25% CHAPS

**Wash 1**: 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% glycerol, 40 mM imidazole, 5 mM β mercaptoethanol, 0.25% CHAPS

**Wash 2**: 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% glycerol, 60 mM imidazole, 5 mM β mercaptoethanol, 0.25% CHAPS

**Elution buffer**: 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% glycerol, 250 mM imidazole, 5 mM β mercaptoethanol, 0.25% CHAPS

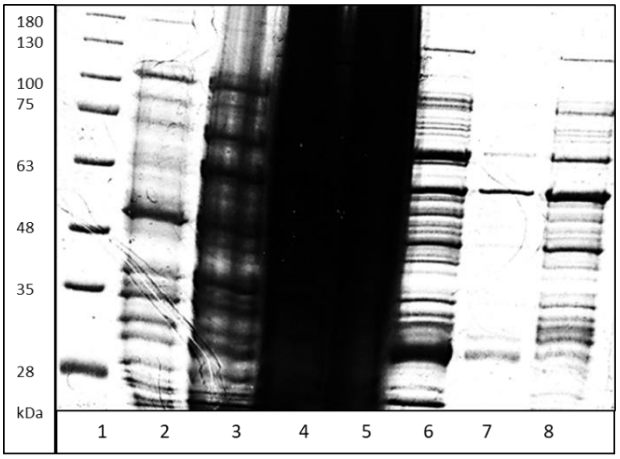



**Figure 35:** Purification attempt of ΔNBtCYP6cm1-his in the presence of CHAPS detergents. Total bacterial extracts were loaded in a $Ni^{2+}$-NTA affinity resin. Proteins were separated with SDS-PAGE (10%) and stained with Coomassie Brilliant Blue – Calculated MW of ΔNBtCYP6cm1-his is 58 kDa. Lane 1: molecular weight marker, Lane 2: before induction bacterial extract, Lane 3: after induction bacterial extract, Lane 4: total protein loading material, Lane 5: unbound proteins, Lane 6: first wash, Lane 7: second wash, Lane 8: second elution fraction of bound proteins

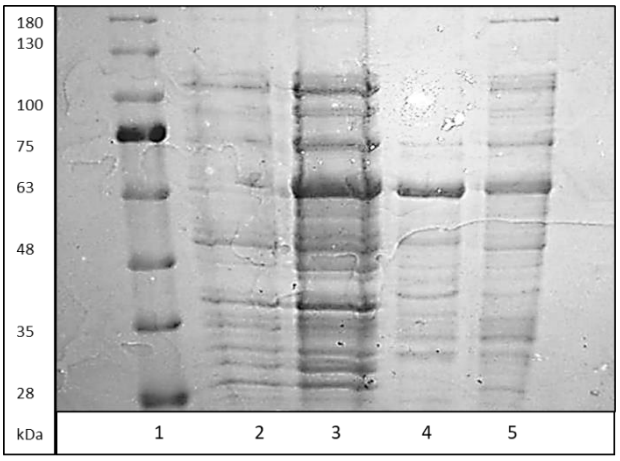



**Figure 36:** Comparison of purification attempts of ΔNBtCYP6cm1-his in the presence of different detergents. Total bacterial extracts were loaded in a $Ni^{2+}$-NTA affinity resin. Proteins were separated with SDS-PAGE (10%) and stained with Coomassie Brilliant Blue – Calculated MW of ΔNBtCYP6cm1-his is 58 kDa. Lane 1: molecular weight marker, Lane 2: before induction bacterial extract, Lane 3: after induction bacterial extract, Lane 4: second elution fraction of bound proteins in the presence of Triton X-100 detergent, Lane 5: second elution fraction of bound proteins in the presence of CHAPS detergent

**Protocol 5: Expression and purification of BtCYP6cm1**

**(Lysozyme, no detergent, $Ni^{2+}$-NTA purification)**

To avoid solubility problems, a softer method for cell lysis was tested. Instead of sonication, lysozyme was used to rapture bacterial cell membranes. With this method

problems like the formation of inclusion bodies could be addressed. In addition, Triton X-100 was not used as solubility needed to be evaluated without the use of detergents.

### Expression

*E. coli* BL21(DE3) Star cells (selected from expression tests using five different expression strains in small scale) were transformed with 50 ng of pCW_ΔN-BtCYP6cm1-Flag-His plasmid. A single colony was selected and left to grow on 10 mL LB liquid culture, supplemented with ampicillin (amp), for 16 h (preculture).

2 x 200 mL TB liquid cultures (supplemented with 100 μg/mL amp) were inoculated with 5 mL of preculture each one (1:50 inoculation). Cultures were incubated at 30$^{o}$C and monitored for their OD at 600 nm. When OD reached 0.5, IPTG and ALA were added at final concentration of 0.5 mM and 1 mM correspondingly. After induction, incubation was continued at 25$^{o}$C for 24 h, shaking. One mL culture sample, before and after 24 h induction (-IPTG, +IPTG), was collected in order to check expression. Cultures were centrifuged at 6000 rpm for 25 min at 4$^{o}$C. Cell pellet was weighed and stored at -20$^{o}$C.

### Cell lysis

Cell pellets stored at -20$^{o}$C were thawed out, using 20 mL ice cold lysis buffer, mixing by pipetting and gentle vortexing. After the samples were completely homogenized, 5 mM β mercaptethanol, 1 mM EDTA, 2 mM PMSF and 1 mg/mL lysozyme were added. The samples were then incubated at 30$^{o}$C shaking at 250 rpm for 1 h.

After the lysozyme treatment, the lysates were left on ice for 30 min to cool down followed by gentle sonication (3 times for 20 sec). After sonication, 50 mM NaCl was added to the samples. The precipitates were subsequently removed after centrifugation at 18.000 rpm, 4$^{o}$C for 30 min. Supernatants were kept on ice until they were loaded to a $Ni^{2+}$-NTA Qiagen affinity column.

### Purification via $Ni^{2+}$-NTA affinity chromatography

$Ni^{2+}$-NTA affinity chromatography was carried out at 4$^{o}$C. P450 lysates (from previous cell lysis step), supplemented also with 5 mM imidazole, were loaded on 200 μL of pre-equilibrated $Ni^{2+}$-NTA resin (pre-equilibrated with 10 column volumes equilibration buffer) with a flow rate 0.5 ml/min and Flowthrough was collected. The column was washed with 10 column volumes (10 * 0.2 mL = 2 mL) of Wash 1 and collected, followed by wash with 10 c.v. (2 mL) of Wash 2. Then, the bound proteins were eluted with 5 c.v. (1 mL) of elution buffer and 5 elution fractions, 200μl each, were collected.

The protein samples ran on a 10% SDS-PAGE. Upon separation the gel was stained with Coomassie Brilliant Blue to estimate protein purity.

**Buffers used**

**Lysis buffer**: 50 mM Tris-HCl pH 7.5, 20% glycerol, 0.1% Triton X-100

**Equilibration buffer**: 50 mM Tris-HCl pH 75, 50 mM NaCl, 20% glycerol, 5 mM β mercaptoethanol

**Wash 1**: 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 20% glycerol, 20 mM imidazole, 5 mM β mercaptoethanol

**Wash 2**: 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 20% glycerol, 40 mM imidazole, 5 mM β mercaptoethanol

**Elution buffer**: 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 20% glycerol, 150 mM imidazole, 5 mM β mercaptoethanol

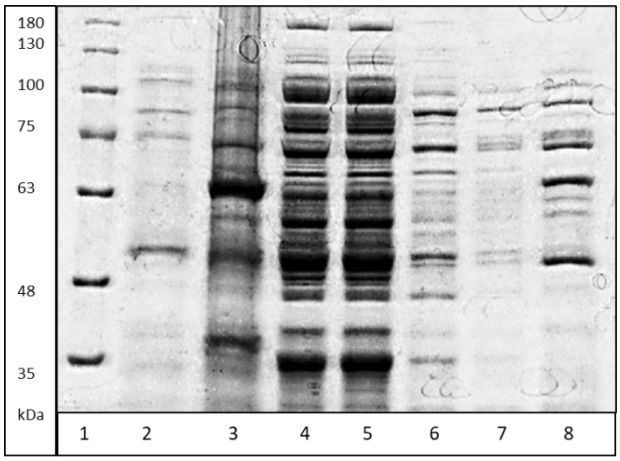



**Figure 37:** Purification attempt of ΔNBtCYP6cm1-his. Total bacterial extracts were loaded in a $Ni^{2+}$-NTA affinity resin. Proteins were separated with SDS-PAGE (10%) and stained with Coomassie Brilliant Blue – Calculated MW of ΔNBtCYP6cm1-his is 58 kDa. Lane 1: molecular weight marker, Lane 2: before induction bacterial extract, Lane 3: after induction bacterial extract, Lane 4: total protein loading material, Lane 5: unbound proteins, Lane 6: first wash, Lane 7: second wash, Lane 8: second elution fraction of bound proteins

In this case, the protein solubility improved to the point that some amount was found at the final elution step. The results are still at a semi-purified level with low amounts of protein eluted in the end, however, the results greatly improved from the initial trials. Since solubility problems were somewhat addressed the next issue was the low affinity to the Nickel resign.

**Protocol 6: Expression and purification of AmCYP9Q3**

**(Lysozyme, no detergent, large scale, ion-exchange, chaperones co-expressed)**

A different chromatography technique, ion exchange, was also evaluated in this work.

### Expression

*E. coli* BL21(DE3) Star cells (selected from expression tests using five different expression strains in small scale) were transformed with 50 ng of pCW_ΔN-Am9Q3-Flag-His plasmid. For this experiment ΔN-CYP9Q3-Flag-His was co-transformed with 50 ng of plasmid vector pTf16 containing the trigger factor (TF) chaperone (see appendix B). A single colony was selected and left to grow on 10 mL LB liquid culture, supplemented with ampicillin (amp), and chloramphenicol for 16 h (preculture).

5 x 200 mL TB liquid cultures (supplemented with 100 μg/mL amp and 34 μg/mL chlor) were inoculated with 5 mL of preculture each one (1:50 inoculation). Cultures were incubated at 30$^{o}$C and monitored for their OD at 600 nm. When OD reached 0.5, IPTG and ALA were added at final concentration of 0.5 mM and 1 mM correspondingly. After induction, incubation was continued at 25$^{o}$C for 24 h, shaking. One mL culture sample, before and after 24 h induction (-IPTG, +IPTG), was collected in order to check expression. Cultures were centrifuged at 6000 rpm for 25 min at 4$^{o}$C. Cell pellet was weighed and stored at -20$^{o}$C.

### Cell lysis

Cell pellets stored at -20$^{o}$C were thawed out, using 40 mL ice cold lysis buffer, mixing by pipetting and gentle vortexing. After the samples were completely homogenized, 5 mM β mercaptoethanol, 1 mM EDTA, 2 mM PMSF and 1 mg/mL lysozyme were added. The samples were then incubated at 30$^{o}$C shaking at 250 rpm for 1 h.

After the lysozyme treatment, the lysates were left on ice for 30 min to cool down followed by gentle sonication (3 times for 20 sec). After sonication, 50 mM NaCl was added to the samples. The precipitates were subsequently removed after centrifugation at 18.000 rpm, 4$^{o}$C for 30 min. Supernatants were kept on ice until they were loaded to a Q or SP SEPHAROSE ion-exchange column.

### Purification via Q or SP Sepharose ion-exchange chromatography

Q or SP Sepharose ion-exchange chromatography was carried out at 25$^{o}$C. P450 lysates (from previous cell lysis step), were diluted until final buffer concentration of 50 mM Tris-HCl pH7.5, 12.5 mM NaCl and 10% glycerol and loaded on 1 mL of pre-equilibrated Q or SP Sepharose resin (pre-equilibrated with 10 column volumes equilibration buffer) with a flow rate 2 mL/min and Flowthrough was collected. The column was washed with 4 column volumes (4 * 1 mL= 4 mL) of Wash 1 and collected. Then, the bound proteins were eluted with 2 c.v. (2 mL) of elution buffer and 5 elution fractions, 400 μL each, were collected.

The protein samples ran on a 10% SDS-PAGE. Upon separation the gel was stained with Coomassie Brilliant Blue to estimate protein purity or wet transferred to nitrocellulose membrane for specific detection using an anti-his antibody.

### Buffers used

**Lysis buffer**: 50 mM Tris-HCl pH 7.5, 10% glycerol, 200mM NaCl, 0.1% Triton X-100

**Equilibration buffer**: 50 mM Tris-HCl pH 7.5, 12.5 mM NaCl, 10% glycerol, 5 mM β mercaptoethanol

**Wash 1**: 50 mM Tris-HCl pH 7.5, 12.5 mM NaCl, 10% glycerol, 5 mM β mercaptoethanol

**Elution buffer**: 50 mM Tris-HCl pH 7.5, 1 M NaCl, 10% glycerol, 5 mM β mercaptoethanol

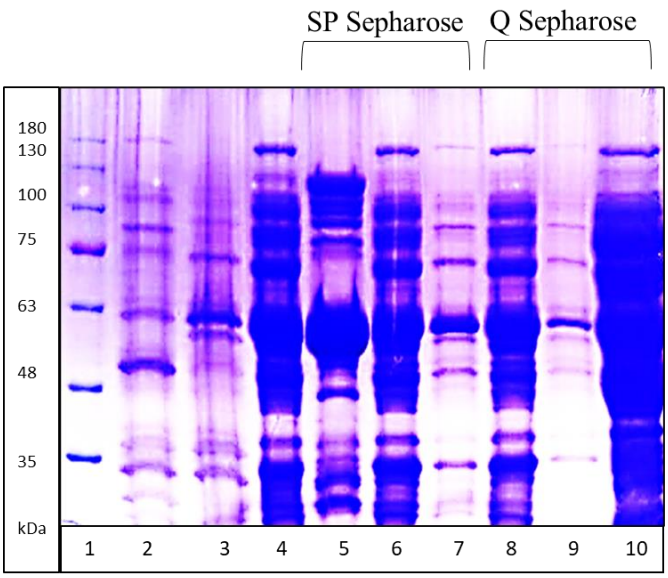



**Figure 38:** Purification attempt of ΔNAmCYP9Q3-his. Total bacterial extracts were loaded in a Q or SP Sepharose resin. Proteins were separated with SDS-PAGE (10%) and stained with Coomassie Brilliant Blue – Calculated MW of ΔNAmCYP9Q3-his is 58 kDa. Lane 1: molecular weight marker, Lane 2: before induction bacterial extract, Lane 3: after induction bacterial extract, Lane 4: total protein loading material, Lane 5: unbound proteins of SP Sepharose resin, Lane 6: first wash of SP Sepharose resin, Lane 7: second elution fraction of bound proteins of SP Sepharose resin, Lane 8: unbound proteins of Q Sepharose resin, Lane 9: first wash of Q Sepharose resin, Lane 10: second elution fraction of bound proteins of Q Sepharose resin

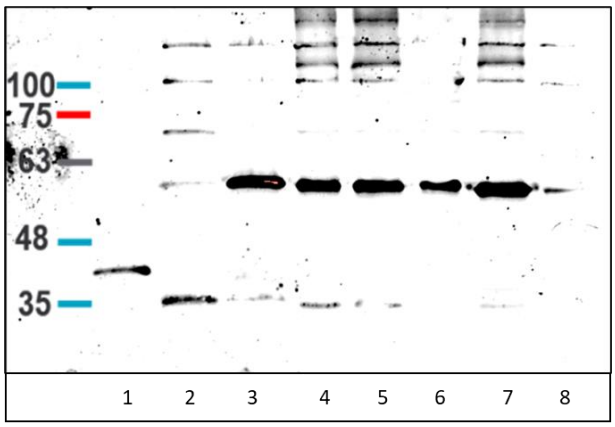



**Figure 39:** Purification attempt of ΔNAmCYP9Q3-his. . Total bacterial extracts were loaded in a Q or SP Sepharose resin. Proteins were separated with SDS-PAGE (10%) and transferred to a nitrocellulose membarne for immunoblotting. Anti-his antibody was used for specific detection – Calculated MW of

ΔNBtCYP6cm1-his is 58 kDa. Lane 1: possitive control of 50ng, Lane 2: before induction bacterial extract, Lane 3: after induction bacterial extract, Lane 4: total protein loading material, Lane 5: unbound proteins of SP Sepharose resin, Lane 6: second elution fraction of bound proteins of SP Sepharose resin, Lane 7: unbound proteins of Q Sepharose resin, Lane 8: second elution fraction of bound proteins of Q Sepharose resin

# 6. Construct design of ΔNBt6CM1 and ΔNMdCPR fusion

## 6.1 Target (theoretical) model

The first step towards the project's goal is to create the DNA construct that will result in the final protein fusion. As mentioned before the desired result would be P450 enzyme from *Bemisia tabaci* right next to CPR from *Musca domestica*, both connected to the scaffold scRM6.

To visualize the final protein product PyMOL was used to create a model by connecting homologous proteins to form a hybrid protein fusion. Specifically, to create the theoretical model the protein sequences of P450 enzyme from *Bemisia tabaci* and CPR from *Musca domestica* were used to create the protein structures via homology modeling. For both cases, Phyre 2 server was used to generate the structures and Mol Probity server was used to evaluate and optimize the models. The two PDB files generated, together with the structure of RM6 (PDB code: 1QX8) were used for the final theoretical model.

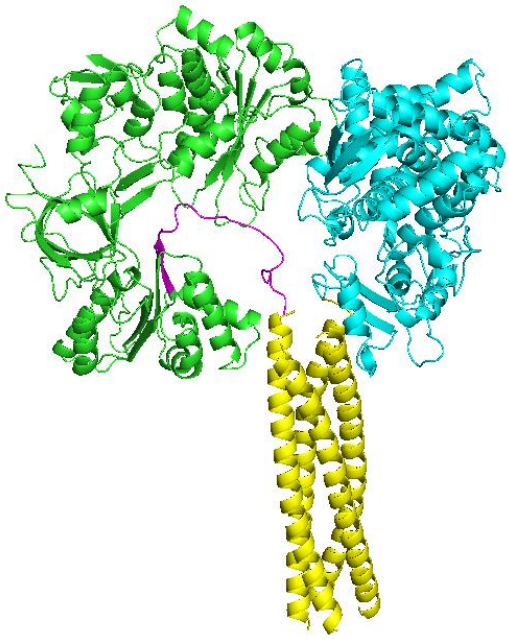


**Figure 40:** Visualization of the desired protein fusion result using PyMOL. scRM6 is represented with yellow, P450 is represented with blue and CPR is represented with green. The pink color corresponds to a linker connecting CPR with scRM6 to add flexibility (the linker looks longer than designed for better visualization purposes).

## 6.2 Transmembrane region removal

Wild type proteins P450 and CPR have a C-terminal transmembrane region as mentioned before. To increase solubility, in the final protein fusion product (CPR-scRM6-

P450), these regions will be removed by not including the DNA sequence that corresponds to the hydrophobic amino-acids in the primers. The decision of what should not be included was made by homology-based modeling.

## 6.3 *Bemisia tabaci* P450

For the P450 enzyme a theoretical model was generated using Phyre2 server and was evaluated using MolProbity server. As mentioned before, the P450s have a very conservative overall fold despite the major differences in amino acid sequence. As shown below the P450 enzyme has a very clear membrane region consisted by 2 α-helices followed by a long loop. Part of the loop will be kept in the final model and will function as a natural linker between P450 enzyme and scRM6 to allow flexibility (Figure 41).

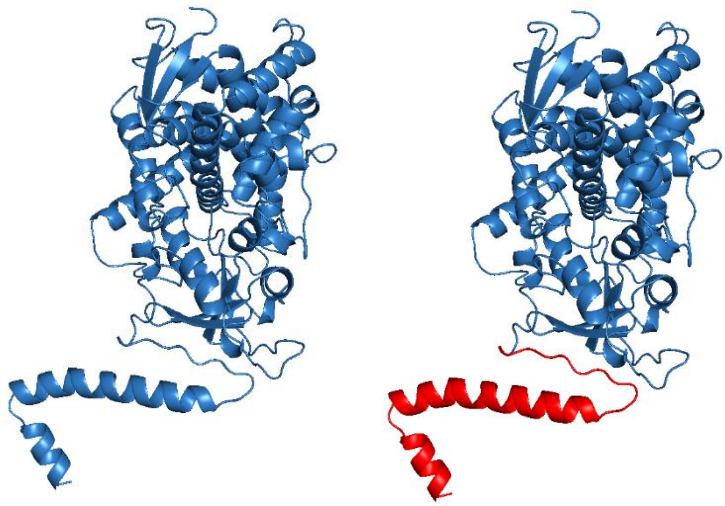


**Figure 41:** Wild type P450 theoretical model (left). Theoretical model of P450 with the part that will be removed in the final protein fusion colored in red (right).

## 6.4 *Musca domestica* CPR

For the CPR enzyme the same method was not as straight forward. The theoretical model was again made using Phyre2 server and was evaluated using MolProbity server. However, because of the lack of structures in protein data bank (PDB) the model lacks the C-terminal region completely (Phyre2 server needs to have a model to base the structure on, since all models did not include the C-terminal domain, the theoretical model was generated without it too) (Figure 42).

Even though the model is not complete, the active site is included. Because of CPR's tight fold shown in the image below, a linker of 3 amino acids will be used to make the N-terminal end (that will be fused with scRM6) more flexible. The amino acids that will be used are NST (all polar) so that they will, in theory, not form a secondary structure (they will form a loop). Note that the DNA sequence that will translate to those amino acids will be AAT AGC ACC.

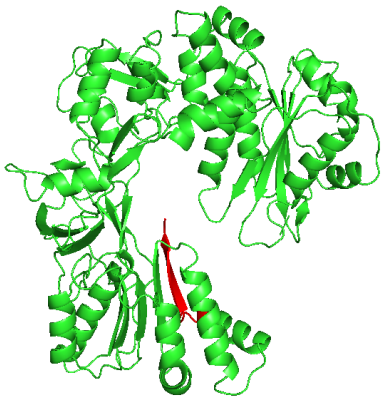


**Figure 42:** Wild type CPR theoretical model. The N-terminal part of CPR (that will be fused with scRM6) is shown in red. Note that the linker is not included in the model.

## 6.5 Overall construct strategy

**Cloning Plan**

1. PCR amplification of scRM6 gene inserting NcoI and SacI restriction sites with primers
2. Digestion of scRM6 gene with NcoI and SacI
3. PCR amplification of P450 gene inserting SacI and XhoI restriction sites with primers
4. Digestion of P450 gene with SacI and XhoI
5. PCR amplification of CPR gene inserting NdeI and NcoI restriction sites with primers
6. Digestion of CPR gene with NdeI and NcoI
7. Digestion of pET26b(+) with NdeI and NcoI
8. Ligation of CPR and pET26b(+)
9. Digestion of the above (pET26b(+)-CPR) with NcoI and SacI
10. Ligation of pET26b(+)-CPR and scRM6
11. Digestion of the above (pET26b(+)-CPR-scRM6) with SacI and XhoI
12. Ligation of the above (pET26b(+)-CPR-scRm6) and P450
13. Send the final constructs for sequencing

# CHAPTER 3

## Discussion and future perspectives

In order to study mechanisms of insecticide resistance as well as investigate new protein targets for the design of new insecticides, *in vitro* techniques were used. In both case studies presented in this work proteins were cloned to a vector, expressed in bacterial cells, and purified via metal affinity or ion exchange chromatography. Isolation of proteins enables better characterization and more accurate results when conducting further experiments. While isolation of *Helicoverpa armigera* GGPPS and *Bamisia tabaci* CYP6CM1 was done to answer different questions, the same techniques were applied by addressing each time the unique challenges of each case.

HaGGPPS was selected as it was evaluated as an interesting insecticide target. Preliminary internal data of our lab (*in vivo* experiments performed in parallel) using the CRISPR/Cas9 system for GGPPS in vivo functional study in *Spodoptera frugiperda,* show that gene knockout results in low survival rate of the injected larvae compared to the control. These findings are very interesting and suggests that disrupting GGPPS function could be a good strategy for lepidopteran pest control. Since the protein is also highly druggable, already being used as a drug target for other organisms and having available crystal structures with several inhibitors bound to GGPPS, new inhibition models for insecticide development is a possible direction.

In this study, HaGGPPS was successfully expressed and purified *in vitro*. A spectrophotometric activity assay was developed and judged ideal as the results were fast and reproducible. Recent studies of *Choristoneura fumiferana* GGPPS activity measured by radio isotopic labelling show low catalytic activity in the presence of DMAPP substrate while the preference of FPP was apparent, something expected from a type III enzyme. Our result support these previous findings with kinetic values obtained from the spectrophotometric assay being similar (results only for DMAPP substrate since FPP was not tested). The spectrophotometric assay was simpler and inexpensive compared to that of radio isotopic labelling, and even though it was not as sensitive, requiring more protein to detect and measure activity, since the protein can be purified in large amounts this in not a problem.

The spectrophotometric assay was also suitable to test inhibition when small chemicals were added to the reaction. This study confirmed that inhibitors of homologous proteins, like zoledronic and minodronic acid, also inhibit HaGGPPS reaching up to 90% inhibition. Using these known inhibitors, more chemicals with similar properties were synthesized and tested. Screening resulted in two compounds, iso-zoledronic acid and riseodronic acid, that performed even better than some already characterized inhibitors. Even though these are promising results, it is still an early stage for drug design. As shown from bioinformatic results, sequences show high similarity between species and active center between pest and pollinators are almost identical with slight amino acid differences being several 8-12 Å away from the natural substrates position. Since selectivity needs to be addressed, chemical hits derived from initial screenings need to be redesigned to be ineffective to non-target species by the addition of moieties that can reach those differences in amino acids or structure and interact differently in targets and non-targets.

Since *in vitro* characterization of HaGGPPS was successful, many options for future perspectives are available. The assay could be further optimized by minimizing the overall volume of reagents needed. Furthermore, enzyme kinetic studies should also be performed using FPP as a substrate, something that could prove a bit challenging because of its high hydrophobicity. More potent inhibitors can be found by screening more chemicals and then optimized to achieve selectivity, and other pharmacokinetical properties to match insecticide

characteristics. Since selectivity is the next step towards insecticide design, GGPPS from non-target species like pollinator *Apis melifera* need to also be isolated *in vitro.* Chemicals, after the second round of optimization could be then tested to observe their effect in different species.

In the second chapter, the case of *Bamisia tabaci* CYP6CM1 was presented. Insect P450s are notorious for their detoxification proteins and are often found up-regulated in resistant strains to various insecticide classes. Since this is a very common resistance strategy many protocols have been developed to characterize P450 enzymes. However, due to their natural properties and challenges, most of these methods can be complicated, expensive and time consuming. In this study, bacterial expression of soluble P450s and purification via metal affinity chromatography was tested as an alternative to other methods currently used. Since the protein complexity and hydrophobicity makes this method extremely challenging, protein engineering was used to address those problems and make the process easier.

N-terminal truncation, even though had a significant effect on solubility, was not sufficient as the protein was found both in the soluble and the membrane fraction. This is speculated to be because of other possible hydrophobic interactions with the membrane achieved by other regions of the P450. However, while some of the protein remained in the hydrophobic, membrane fraction of the bacterial cells, some of the proteins were successfully retrieved from the soluble fraction. These findings are promising, leaving room for even better results after additional protocol optimization.

From early to late trials, poor protein affinity to the $Ni^{2+}$-NTA resign remained unsolved. Purification profiles always suggested that the majority of soluble protein was lost to the unbound protein fraction, with very low amounts finally eluting in the final purification steps accompanied with other non-specifically bound to the resign bacterial proteins. This could mean that the 10X-his-tag added to the N-terminal part of the protein was not properly interacting with the $Ni^{2+}$-NTA resign. It is speculated that this is due to the tag being hidden by the overall structure of the protein. This problem could be solved by the shift of the his-tag to the C-terminal of the P450. Even though *Bamisia tabaci* CYP6CM1 was not successfully purified, some challenges were addressed and solved, and the results were improving. Solving the protein affinity to the resign could be the last step towards the initial goal.

As an alternative protein engineering strategy, a construct of a protein fusion was also designed, though not tested in this work. The concept behind this design was to express a P450 enzyme fused to the CPR redox partner. To keep this system in theory possible for bacterial expression, several protein modifications were made. Both proteins were designed without their N-terminal membrane anchors to increase solubility. In addition, those proteins were docked in a protein scaffold scRM6 to increase solubility and to be in proximity for them to properly function. This strategy, although more complex, could make P450 characterization more straight forward. Since this construct was designed with the thought of easily changing the P450 sequence in the initial vector by digestion, if this strategy worked, different P450 enzymes could be tested this way.

# Appendix A: HaGGPPS

**Protein sequence**

>HaGGPPS as cloned to the pET16b(+) vector

```
MGHHHHHHHH HHSSGHIEGR HMMSKVSTNT GDNTQDEKIL MPYTYIQQVP
GKQIRQKLAS AFNYWLKISD EKLRAVGEIV QMLHNSSLLI DDIQDNSILR
RGIPVAHSIY GVASTINAAN YTMIVALEKT QELGHPMATR VYTEQLLELH
RGQGIEIYWR DNFQCPSEEE YKEMTIKKTG GLFMLAIRLM QLFSENKSDF
SKLSAILGLY FQIRDDYCNL RLQEYTENKS YCEDLTEGKF SFPIIHAIRN
PEGDKQVLHI LRQRTRDLEV KRYCITILER IGSFAYTRET LQSLDEEARR
EVARLGGNPH LEALLDDLLS WRRDKPLENN V*
```

**Primer sequences**

For GGPPS cloning

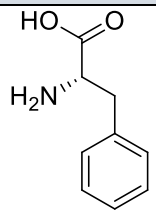
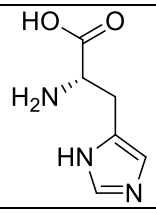
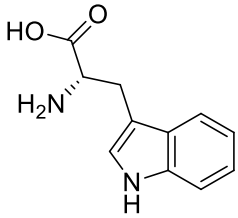
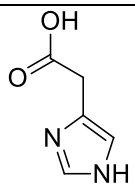
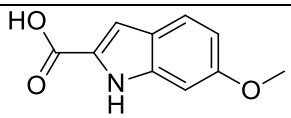
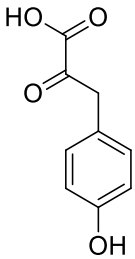
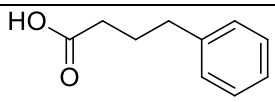
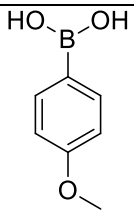
Ha_GGPPS_ORF_NdeI_F: `GTA CCA TAT GAT GTC TAA AGT TAG TAC AAA TAC TGG TGA`

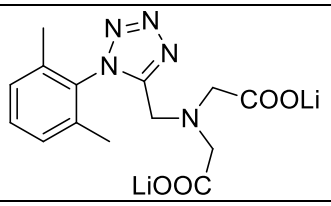
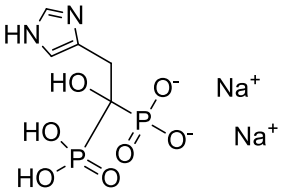
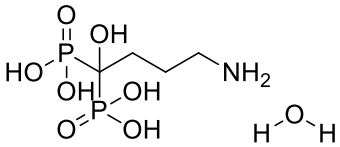
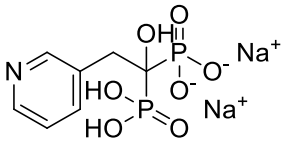
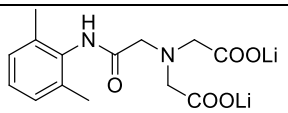
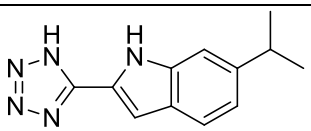
Ha_GGPPS_ORF_BamHI_R: `GTA CGG ATC CTT AAA CGT TGT TCT CGA GGG`

**Table 6:** Genes used for bioinformatic analysis

| ORGANISM | DATABASE | ACCESSION NUMBER | ASSEMBLY |
|---|---|---|---|
| Homo sapiens | UniProt | O95749 | - |
| Saccharomyces cerevisiae | UniProt | Q12051 | - |
| Spodoptera litura | NCBI | LOC111355525 | ASM270686v1 (GCF_002706865.1) |
| Choristoneura fumiferana | UniProt | U3RD44 | - |
| Apis mellifera | NCBI | LOC727189 | Amel_HAv3.1 (GCF_003254395.2) |
| Osmia lignaria | NCBI | LOC117610300 | USDA_OLig_1.0 (GCF_012274295.1) |
| Bombus terrestris | UniProt | I1VX07 | - |

**Table 7:** Chemicals tested in the inhibition screening

| Entry | Structure | Code | MB | Activity (%) |
|---|---|---|---|---|
| 1 |  | MT_CH_1 | 165 | 69.7 |
| 2 |  | MT_CH_2 | 155 | 69.3 |
| 3 |  | MT_CH_3 | 204 | 71.3 |
| 4 |  | MT_CH_6 | 126 | 87 |
| 5 |  | MT_CH_7 | 191 | 85.1 |
| 6 |  | MT_CH_8 | 180 | 85.1 |
| 7 |  | MT_CH_9 | 164 | 83.1 |
| 8 |  | MT_CH_10 | 152 | 161 |

| | | | | |
|---|---|---|---|---|
| 9 |  | MT148 | 319 | 100 |
| 10 |  | Iso_zol | | 47.2 |
| 11 |  | ALE | | 57.4 |
| 12 |  | RISE | | 33.7 |
| 13 |  | MT 52 | | ~78 |
| 14 |  | KN2 | | ~85 |

# Appendix B: BtCYP6CM1

**Protein sequence**

>BtCYP6CM1 as cloned to the pCWori vector

```
HMDKFHYWSK RGVPCQSPAQ SIVRTFRLVL RMDSFTDNFY GVYKAFDGHP
YVGSLELTKP ILVVRDPELA RIVLVKSFSS FSGRLKSPDT TLDPLSNHLF
TLNGEKWRQV RHKTATAFST AKLKNMFHSL KDCAREMDAY MERAIGDKGD
VEFDALKVMS NYTLEVIGAC AMGIKCDSIH DEETEFKRFS RDFFRFDARR
MIFTLLDLLH PKLPVLLKWK AVRPEVENFF REAIKEAASL KESEAAARTD
FLQILIDFQK SEKASKTDAG NDTELVFTDN IIGGVIGSFF SAGYEPTAAA
LTFCLYELAR NPQVQAKLHE EILAVKEKLG DDIEYETLKE FKYANQVIDE
TLRLYPASGI LVRTCTEPFK LPDSDVVIEK GTKVFVSSYG LQTDPRYFPE
PEKFDPERFS EENKEKILPG TYLPFGDGPR LCIAMRLALM DVKMMMVRLV
SKYEIHTTPK TPKKITFDTN SFTVQPAEKV WLRFRRRAST PDYKDDDDKH
HHHHH-
```

**Design of the MdCPR-scRM6-BtCYP6CM**

**Primers for MdCPR**

FORWARD PRIMER (27): **CAT ATG GTG AGC ACC ACC GAA AAT AGC**
REVERSE PRIMER (37): **CCA TGG GGT GCT ATT GCT CCA AAC ATC TGC GCT ATA A**

(Black color: the complementary sequence, green color: the extra bases, red color: the restriction site)

**Primers for scRM6**

FORWARD PRIMER (30): **CCA TGG CAT ATG ACT AAG CAA GAG AAG ACC**
REVERSE PRIMER (27): **GAG CTC ACC CTG GAA ATA CAG GTT CTC**

(Black color: the complementary sequence, red color: the restriction site)

**Primers for BtP450**

FORWARD PRIMER (31): **GAG CTC AGC ATT GTT CGT ACC TTT CGT CTG G**
REVERSE PRIMER (24): **CTC GAG CGG GGT GCT TGC ACG ACG**

(Black color: the complementary sequence, red color: the restriction site)

**Final DNA sequence after the gene fusion**

**CATATG**GTGA GCACCACCGA AAATAGCTTT ATCAAAAAAC TGAAAGCAAG
CGGTCGTAGC CTGGTTGTTT TTTATGGTAG TCAGACCGGC ACCGCAGAAG
AATTTGCAGG TCGTCTGGCA AAAGAAGGTC TGCGTTATCG TATGAAAGGT
ATGGTTGCAG ATCCGGAAGA ATGTGATATG GAAGAACTGC TGCAGATGAA
AGATATTCCG AATAGCCTGG CCGTTTTTTG TCTGGCAACC TATGGTGAAG
GTGATCCGAC CGATAATGCA ATGGAATTCT ATGAATGGAT TACCAACGGT
GAAGTGGATC TGACCGGTCT GAATTATGCA GTTTTTGGTC TGGGCAATAA
AACCTATGAG CACTATAACA AAGTGGCCAT CTATGTGGAT AAACGCCTGG
AAGAATTAGG TGCGACCCGT GTTTTTGAAT TAGGTCTGGG TGATGATGAT
GCCAACATTG AAGATGATTT CATCACCTGG AAAGATCGTT TTTGGCCGAG
CGTTTGTGAT TTTTTTGGTA TTGAAGGTAG CGGTGAAGAG GTTCTGATGC
GTCAGTTTCG TCTGCTGGAA CAGCCGGATG TTCAGCCGGA TCGTATTTAT
ACCGGTGAAA TTGCACGTCT GCACAGCATG CAGAATCAGC GTCCGCCTTT
TGATGCAAAA AATCCTTTTC TGGCAAGCGT GATTGTTAAT CGTGAACTGC
ACAAAGGTGG TGGTCGTAGC TGTATGCATA TTGAACTGGA TATCGATGGT
AGCAAAATGC GTTATGATGC CGGTGATCAT ATTGCCATGT ATCCGATTAA
CGATAAGATC CTGGTTGAGA AACTGGGCAA ACTGTGTGAT GCAAATCTGG
ATACCGTTTT TAGCCTGATT AATACCGATA CCGACAGCAG CAAAAAACAT
CCGTTTCCGT GTCCGACCAC CTATCGTACC GCACTGACCC ATTATCTGGA
AATTACCGCA ATTCCGCGTA CGCATATTCT GAAAGAACTG GCAGAATATT
GCAGCGACGA AAAGGATAAA GAATTTCTGC GTAATATGGC CAGCATTACA
CCGGAAGGTA AAGAGAAATA TCAGAACTGG ATTCAGAATA GCAGCCGCAA
CATTGTTCAT ATCCTGGAAG ATATCAAAAG CTGTCGTCCT CCGATTGATC
ATATCTGTGA ACTGCTTCCG CGTCTGCAGC CTCGTTATTA TAGCATTAGC
AGCAGTAGCA AACTGTATCC GACCAATGTT CATATTACCG CAGTTCTGGT
TCAGTATGAA ACCCCGACAG GTCGTGTTAA TAAAGGTGTT GCAACCAGCT
ACATGAAAGA AAAAAATCCG AGCGTGGGTG AAGTTAAAGT TCCGGTTTTT
ATTCGCAAAA GCCAGTTTCG CCTGCCGACC AAAAGCGAAA TTCCGATTAT
TATGGTTGGT CCTGGCACCG GTCTGGCACC GTTTCGTGGT TTTATTCAAG
AACGCCAGTT TCTGCGTGAT GGTGGTAAAG TTGTTGGTGA TACCATTCTG
TATTTTGGCT GCCGCAAAAA AGATGAGGAT TTTATCTATC GCGAAGAACT
GGAACAGTAT GTTCAGAATG GCACCCTGAC ACTGAAAACC GCATTTAGCC
GTGATCAGCA AGAAAAAATC TATGTGACCC ATCTGATTGA ACAGGATGCA
GATCTGATTT GGAAAGTGAT TGGTGAACAG AAAGGCCACT TTTATATCTG
TGGTGATGCC AAAAACATGG CCGTTGATGT TCGTAATATT CTGGTGAAAA
TTCTGAGCAC CAAAGGCAAC ATGAATGAAA GTGATGCAGT GCAGTACATC
AAAAAAATGG AAGCGCAGAA ACGTTATAGC GCAGATGTTT GGAGCAATAG
CACC**CCATGG** CATATGACTA AGCAAGAGAA GACCGCCCTG AATATGGCGC
GTTTTATTCG CTCACAGACC CTGACCCTGC TGGAAAAGCT GAACGAACTG
GCGGACATCT GCGAATCTCT GCACGATCAC GCCGACGAGC TGTACCGTTC
TTGCCTGGCC CGCTTCGGGG ATGACGGCGA AAACCTGACA AAGCAGGAAA
AGACCGCGCT GAACATGGCG CGTTTCATTC GCAGCCAAAC ACTGACACTG
CTGGAAAAAC TGAATGAACT GGCAGACATT TGCGAGTCGC TGCACGATCA
TGCCGACGAA CTGTATCGTA GTTGTCTGGC ACGCTTTGGA GACGACGGAG
AGAACCTGAC TAAGCAAGAA AAGACGGCAC TGAATATGGC TCGCTTTATC
CGTTCGCAAA CACTGACCCT GCTGGAGAAA CTGAACGAAC TGGCTGACAT
TTGTGAATCG CTGCACGATC ACGCAGATGA ACTGTACCGT TCATGTCTGG
CACGTTTTGG GGACGACGGT GAAAATCTGA CCAAACAGGA GAAAACCGCC
CTGAACATGG CCCGCTTTAT TCGTAGCCAA ACTCTGACTC TGCTGGAGAA
GCTGAACGAG CTGGCTGATA TCTGTGAGTC TCTGCACGAT CACGCCGATG

AACTGTATCG TTCCTGTCTG GCACGCTTCG GGGACGATGG AGAAAACCTG
GAGAACCTGT ATTTCCAGGG T**GAGCTC**AGC ATTGTTCGTA CCTTTCGTCT
GGTTCTGCGT ATGGATAGCT TTACCGATAA CTTTTATGGC GTGTATAAAG
CCTTTGATGG TCATCCGTAT GTTGGTAGCC TGGAACTGAC CAAACCGATT
CTGGTTGTTC GTGATCCGGA ACTGGCACGT ATTGTTCTGG TTAAAAGCTT
TAGCAGCTTT AGTGGTCGTC TGAAAAGTCC GGATACCACA CTGGATCCGC
TGAGCAATCA CCTGTTTACC CTGAATGGTG AAAAATGGCG TCAGGTTCGT
CATAAAACCG CAACCGCATT TAGCACCGCA AAACTGAAAA ATATGTTCCA
CAGCCTGAAA GATTGTGCCC GTGAAATGGA TGCATATATG GAACGTGCAA
TTGGCGATAA AGGTGATGTT GAATTTGATG CCCTGAAAGT GATGAGCAAT
TATACCCTGG AAGTTATTGG TGCATGTGCA ATGGGTATTA AATGCGATAG
CATCCATGAT GAAGAAACCG AATTTAAACG CTTCAGCCGT GATTTTTTTC
GTTTTGATGC ACGTCGCATG ATTTTTACCC TGCTGGATCT GCTGCATCCG
AAACTGCCGG TGCTGCTGAA ATGGAAAGCA GTTCGTCCGG AAGTTGAAAA
CTTTTTTCGT GAAGCAATTA AAGAAGCCGC AAGTCTGAAA GAAAGCGAAG
CAGCAGCACG TACCGATTTT CTGCAGATTC TGATTGATTT TCAGAAAAGC
GAGAAAGCCA GCAAAACCGA TGCAGGTAAT GATACCGAAC TGGTGTTTAC
CGACAACATT ATTGGTGGTG TTATCGGCAG CTTTTTTAGC GCAGGTTATG
AACCGACCGC AGCAGCCCTG ACCTTTTGTC TGTATGAGCT GGCACGCAAT
CCGCAGGTTC AGGCAAAACT GCATGAAGAA ATTCTGGCCG TTAAAGAAAA
ACTGGGCGAC GATATTGAAT ATGAAACCCT GAAAGAATTC AAATATGCCA
ACCAGGTTAT TGATGAAACG CTGCGTCTGT ATCCGGCAAG CGGTATTCTG
GTGCGTACCT GTACCGAACC GTTTAAACTG CCTGATAGTG ATGTGGTTAT
TGAGAAGGGC ACCAAAGTTT TTGTTAGCAG CTATGGCCTG CAGACCGATC
CGCGTTATTT TCCGGAACCT GAAAAATTTG ATCCTGAACG CTTTAGCGAA
GAAAACAAAG AAAAAATTCT GCCTGGCACC TATCTGCCGT TTGGTGATGG
TCCGCGTCTG TGTATTGCAA TGCGTCTGGC ACTGATGGAT GTTAAAATGA
TGATGGTGCG TCTGGTGAGC AAATATGAAA TTCATACAAC CCCGAAAACG
CCGAAAAAGA TTACCTTTGA TACCAATTCC TTTACCGTTC AGCCAGCAGA
AAAAGTTTGG CTGCGTTTTC GTCGTCGTGC AAGCACCCCG **CTCGAG**

**Final amino acid sequence after the gene fusion**

5'3' Frame 1

HMVSTTENSFIKKLKASGRSLVVFYGSQTGTAEEFAGRLAKEGLRYRMKGMVADPEECDMEELLQMKDIPNSLAVFCLATYGEGDPTDNAMEFYEWITNGEVDLTGLNYAVFGLGNKTYE
HYNKVAIYVDKRLEELGATRVFELGLGDDDANIEDDFITWKDRFWPSVCDFFGIEGSGEEVLMRQFRLLEQPDVQPDRIYTGEIARLHSMQNQRPPFDAKNPFLASVIVNRELHKGGGRS
CMHIELDIDGSKMRYDAGDHIAMYPINDKILVEKLGKLCDANLDTVFSLINTDTDSSKKHPFPCPTTYRTALTHYLEITAIPRTHILKELAEYCSDEKDKEFLRNMASITPEGKEKYQNW
IQNSSRNIVHILEDIKSCRPPIDHICELLPRLQPRYYSISSSSKLYPTNVHITAVLVQYETPTGRVNKGVATSYMKEKNPSVGEVKVPVFIRKSQFRLPTKSEIPIIMVGPGTGLAPFRG
FIQERQFLRDGGKVVGDTILYFGCRKKDEDFIYREELEQYVQNGTLTLKTAFSRDQQEKIYVTHLIEQDADLIWKVIGEQKGHFYICGDAKNMAVDVRNILVKILSTKGNMNESDAVQYI
KKMEAQKRYSADVWSNSTPWHMTKQEKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCLARFGDDGENLTKQEKTALNMARFIRSQTLTLLEKLNELADICESLHDHADE
LYRSCLARFGDDGENLTKQEKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCLARFGDDGENLTKQEKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCL
ARFGDDGENLENLYFQGELSIVRTFRLVLRMDSFTDNFYGVYKAFDGHPYVGSLELTKPILVVRDPELARIVLVKSFSSFSGRLKSPDTTLDPLSNHLFTLNGEKWRQVRHKTATAFSTA
KLKNMFHSLKDCAREMDAYMERAIGDKGDVEFDALKVMSNYTLEVIGACAMGIKCDSIHDEETEFKRFSRDFFRFDARRMIFTLLDLLHPKLPVLLKWKAVRPEVENFFREAIKEAASLK
ESEAAARTDFLQILIDFQKSEKASKTDAGNDTELVFTDNIIGGVIGSFFSAGYEPTAAALTFCLYELARNPQVQAKLHEEILAVKEKLGDDIEYETLKEFKYANQVIDETLRLYPASGIL
VRTCTEPFKLPDSDVVIEKGTKVFVSSYGLQTDPRYFPEPEKFDPERFSEENKEKILPGTYLPFGDGPRLCIAMRLALMDVKMMMVRLVSKYEIHTTPKTPKKITFDTNSFTVQPAEKVW
LRFRRRASTPPW

**Plasmids expressing members of the chaperone team**

Plasmids overexpressing members of the chaperone team (DnaK, DnaJ,GrpE, GroE, GroES and trigger factor) in various combinations and controlled conditions (dose-dependent L-arabinose and tetracycline addition) were incorporated with the intention of facilitating the folding, aggregation, and stability of the heterologous proteins in the E. coli strains. Plasmid pTf16 was used in this work (Nishihara et al., 1998, Nishihara et al., 2000).

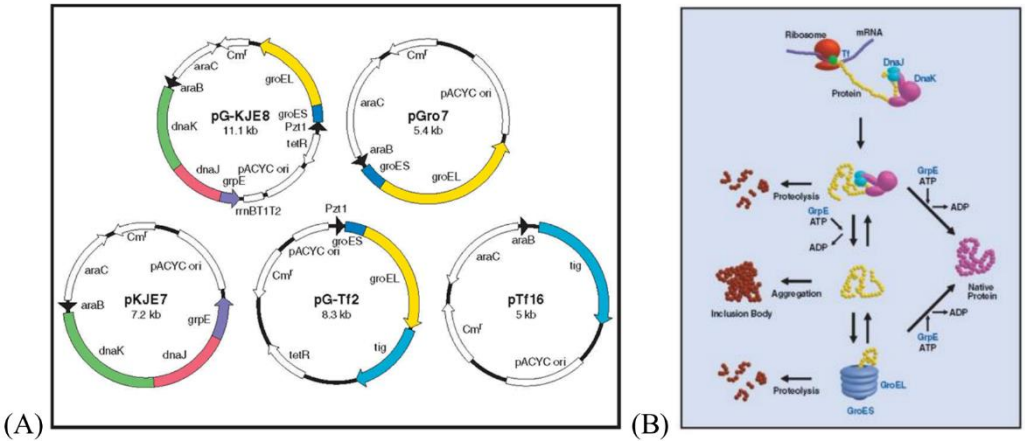



**Figure 43:** (A) The Chaperone Plasmid Set consists of five different plasmids, each of which is designed to express multiple molecular chaperones that function together as a "chaperone team" to enable optimal protein expression and folding and to reduce protein misfolding. Each plasmid carries an origin of replication (ori) derived from pACYC and a chloramphenicol-resistance gene (Cmr) gene, which allows the use of E. coli expression systems containing ColE1-type plasmids that confer ampicillin resistance. The chaperone genes are situated downstream of the araB or Pzt-1 (tet) promoters; as a result, expression of target proteins and chaperones can be individually induced if the target gene is placed under the control of different promoters (e.g., lac). These plasmids also contain the ecessary regulator (araC or tetr) for each promoter. (B) Possible model for chaperone-assisted protein folding in E. coli. Figure adapted from Takara Bio.

## References


Feng, H. Q., K. M. Wu, Y. X. Ni, D. F. Cheng, and Y. Y. Guo. 2005. 'Return migration of Helicoverpa armigera (Lepidoptera: Noctuidae) during autumn in northern China', Bull Entomol Res, 95: 361-70.

Fitt, G P. 1989. 'The Ecology of Heliothis Species in Relation to Agroecosystems', 34: 17-53.

Head, D. J., A. R. McCaffery, and A. Callaghan. 1998. 'Novel mutations in the para-homologous sodium channel gene associated with phenotypic expression of nerve insensitivity resistance to pyrethroids in Heliothine lepidoptera', Insect Mol Biol, 7: 191-6.

Jones, Christopher M., Hazel Parry, Wee Tek Tay, Don R. Reynolds, and Jason W. Chapman. 2019. 'Movement Ecology of Pest Helicoverpa: Implications for Ongoing Spread', 64: 277-95.

Li, G., H. Feng, Y. Gao, K. A. Wyckhuys, and K. Wu. 2010. 'Frequency of Bt resistance alleles in Helicoverpa armigera in the Xinjiang cotton-planting region of China', Environ Entomol, 39: 1698-704.

Reed, W., and C. Pawar. 1982. "Heliothis: a Global Problem." In.

Tay, W. T., M. F. Soria, T. Walsh, D. Thomazoni, P. Silvie, G. T. Behere, C. Anderson, and S. Downes. 2013. 'A brave new world for an old world pest: Helicoverpa armigera (Lepidoptera: Noctuidae) in Brazil', PLoS One, 8: e80134.

Yang Y, Li Y, Wu Y. Current status of insecticide resistance in Helicoverpa armigera after 15 years of Bt cotton planting in China. J Econ Entomol. 2013 Feb;106(1):375-81. doi: 10.1603/ec12286. PMID: 23448054.

Durigan MR, Corrêa AS, Pereira RM, Leite NA, Amado D, de Sousa DR, Omoto C. High frequency of CYP337B3 gene associated with control failures of Helicoverpa armigera with pyrethroid insecticides in Brazil. Pestic Biochem Physiol. 2017 Nov;143:73-80. doi: 10.1016/j.pestbp.2017.09.005. Epub 2017 Sep 15. PMID: 29183613.

Bueno, Regiane C. O. F.; Raetano, Carlos G.; Junior, Joaz Dorneles; Carvalho, Fernando K. Integrated Management of Soybean Pests: The Example of Brazil. Outlooks on Pest Management, Volume 28, Number 4, August 2017, pp. 149-153(5). doi: https://doi.org/10.1564/v28_aug_02

Aline Pomari-Fernandes, Adeney de Freitas Bueno, Daniel Ricardo Sosa-Gómez. Helicoverpa armigera: current status and future perspectives in Brazil. Current Agricultural Science and Technology Vol 21, No 1 (2015). DOI: https://doi.org/10.18539/cast.v21i1.4234

Ahmad M, Rasool B, Ahmad M, Russell DA. Resistance and Synergism of Novel Insecticides in Field Populations of Cotton Bollworm Helicoverpa armigera (Lepidoptera: Noctuidae) in Pakistan. J Econ Entomol. 2019 Mar 21;112(2):859-871. doi: 10.1093/jee/toy409. PMID: 30668865.

Sun X. History and current status of development and use of viral insecticides in China. Viruses. 2015 Jan 20;7(1):306-19. doi: 10.3390/v7010306. PMID: 25609304; PMCID: PMC4306840.

James, C., 2002: Global Review of Commercialized Transgenic Crops: 2001 Feature: Bt Cotton. ISAAA Briefs No. 26. ISAAA, Ithaca, NY.

Barwale, R. B., V. R. Gadwa, U. Zehr, and B. Zehr, 2004: Prospects for Bt cotton technology in India. AgbioForum 7, 23—26.

Dong, H. Z., W. J. Li, W. Tang, and D. M. Zhang, 2005: Increased yield and revenue with a seedling transplanting system for hybrid seed production in Bt cotton. J. Agron. Crop Sci. 191, 116—124.

Gianessi, L. P., and J. E. Carpenter, 1999: Agricultural Biotechnology: Insect Control Benefits. National Center for Food and Agricultural Policy, Washington DC, USA.

Tabashnik, B. E., T. J. Dennehy, M. A. Sims, K. Larkin, G. P. Head, W. J. Moar, and Y. Carriere, 2002: Control of resistant pink bollworm that produces Bacillus thuringiensis toxin Cry2Ab. Appl. Environ. Microbiol. 68, 3790—3794.

Hofte, H., and H. Whitely, 1989: Insecticidal crystal properties of Bacillus thuringiniesis. Microbiol. Rev. 53, 242—255.

Schnepf, E., N. Crickmore, J. Van Rie, D. Lereclus, J. Baum, J. Feitelson, D. R. Zeigler, and D. H. Dean, 1998: Revision of the nomenclature for the Bacillus thuringiensis pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. 62, 775—806.

Greenplate, J., W. Mullins, S. Penn, and K. Embry, 2001: Cry1Ac Bollgard® varieties as influenced by environment, variety and plant age: 1999 gene equivalency field studies. In P. Dugger, and R. Richter, eds. Proceedings of the Beltwide Cotton Conference, Volume 2, pp. 790—793. National Cotton Council of America, Memphis, TN.

Holt, H. E., C. L. Mares, and R. Akhurst, 2002: Determination of the Cryl protein content of Bt transgenic cotton – a technical manual for laboratory use CSIRO Entomol. Tech. Rep. No. 92 pp. 1—13.

Oppenoorth, 1984 Biochemistry of insecticide resistance F.J. Oppenoorth Pestic. Biochem. Physiol., 22 (1984), pp. 183-187

Walsh, T. K., Joussen, N., Tian, K., McGaughran, A., Anderson, C. J., Qiu, X., Ahn, S. J., Bird, L., Pavlidi, N., Vontas, J., Ryu, J., Rasool, A., Barony Macedo, I., Tay, W. T., Zhang, Y., Whitehouse, M., Silvie, P. J., Downes, S., Nemec, L., & Heckel, D. G. (2018). Multiple recombination events between two cytochrome P450 loci contribute to global pyrethroid resistance in Helicoverpa armigera. PloS one, 13(11), e0197760. https://doi.org/10.1371/journal.pone.0197760

Joußen, N., Agnolet, S., Lorenz, S., Schöne, S. E., Ellinger, R., Schneider, B., & Heckel, D. G. (2012). Resistance of Australian Helicoverpa armigera to fenvalerate is due to the chimeric P450 enzyme CYP337B3. Proceedings of the National Academy of Sciences of the United States of America, 109(38), 15206–15211. https://doi.org/10.1073/pnas.1202047109

Tian K, Feng J, Zhu J, Cheng J, Li M, Qiu X. Pyrethrin-resembling pyrethroids are metabolized more readily than heavily modified ones by CYP9As from Helicoverpa armigera. Pestic

Biochem Physiol. 2021 Jul;176:104871. doi: 10.1016/j.pestbp.2021.104871. Epub 2021 May 7. PMID: 34119216.

Teese, M. G., Farnsworth, C. A., Li, Y., Coppin, C. W., Devonshire, A. L., Scott, C., East, P., Russell, R. J., & Oakeshott, J. G. (2013). Heterologous expression and biochemical characterisation of fourteen esterases from Helicoverpa armigera. PloS one, 8(6), e65951. https://doi.org/10.1371/journal.pone.0065951

Sayyed AH, Wright DJ (2006) Genetics and evidence for an esterase-associated mechanism of resistance to indoxacarbin a field population of diamondback moth (Lepidoptera: Plutellidae). Pest Manag Sci 62:1045–1051. https://doi.org/10.1002/ps.1270

Alvi AH, Sayyed AH, Naeem M, Ali M (2012) Field evolved resistance in Helicoverpa armigera (Lepidoptera: Noctuidae) to Bacillus thuringiensis toxin Cry1Ac in Pakistan. PLoS ONE 7:e47309. https://doi.org/10.1371/journal.pone.0047309

Tabashnik B.E., Huang F., Ghimire M.N., Leonard B.R., Siegfried B.D., Rangasamy M., Yang Y., Wu Y., Gahan L.J., Heckel D.G., et al. Efficacy of genetically modified Bt toxins against insects with different genetic mechanisms of resistance. Nat. Biotechnol. 2011;29:1128–1131. doi: 10.1038/nbt.1988.

Tabashnik B.E., Brévault T., Carrière Y. Insect resistance to Bt crops: Lessons from the first billion acres. Nat. Biotechnol. 2013;31:510–521. doi: 10.1038/nbt.2597.

Liliana P.L., Mario S., Alejandra B. Bacillus thuringiensis insecticidal three-domain cry toxins: Mode of action, insect resistance and consequences for crop protection. Fems Microbiol. Rev. 2013;37:3–22.

Bravo A., Likitvivatanavong S., Gill S.S., Soberon M. Bacillus thuringiensis: A story of a successful bioinsecticide. Insect Biochem. Mol. 2011;41:423–431. doi: 10.1016/j.ibmb.2011.02.006.

Schnepf E., Crickmore N., Van Rie J., Lereclus D., Baum J., Feitelson J., Zeigler D.R., Dean D.H. Bacillus thuringiensis and its pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. 1998;62:775–806.

Heckel D.G., Gahan L.J., Baxter S.W., Zhao J.Z., Shelton A.M., Gould F., Tabashnik B.E. The diversity of bt resistance genes in species of lepidoptera. J. Invertebr. Pathol. 2007;95:192–197. doi: 10.1016/j.jip.2007.03.008.

Brévault, T., Heuberger, S., Zhang, M., Ellers-Kirk, C., Ni, X., Masson, L., Li, X., Tabashnik, B. E., & Carrière, Y. (2013). Potential shortfall of pyramided transgenic cotton for insect resistance management. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 5806–5811. https://doi.org/10.1073/pnas.1216719110

Chakroun, M., Banyuls, N., Walsh, T., Downes, S., James, B., & Ferré, J. (2016). Characterization of the resistance to Vip3Aa in Helicoverpa armigera from Australia and the role of midgut processing and receptor binding. Scientific reports, 6, 24311. https://doi.org/10.1038/srep24311

Ling Y, Li ZH, Miranda K, Oldfield E, Moreno SN. The farnesyl-diphosphate/geranylgeranyl-diphosphate synthase of Toxoplasma gondii is a bifunctional enzyme and a molecular

target of bisphosphonates. J Biol Chem. 2007 Oct 19;282(42):30804-16. doi: 10.1074/jbc.M703178200. Epub 2007 Aug 27. PMID: 17724033.

Wiemer AJ, Wiemer DF, Hohl RJ. Geranylgeranyl diphosphate synthase: an emerging therapeutic target. Clin Pharmacol Ther. 2011 Dec;90(6):804-12. doi: 10.1038/clpt.2011.215. Epub 2011 Nov 2. PMID: 22048229.

Dhar MK, Koul A, Kaul S. Farnesyl pyrophosphate synthase: a key enzyme in isoprenoid biosynthetic pathway and potential molecular target for drug development. N Biotechnol. 2013 Jan 25;30(2):114-23. doi: 10.1016/j.nbt.2012.07.001. Epub 2012 Jul 25. PMID: 22842101.

Jonathan P. Day, Vaughn Cleghon, Miles D. Houslay, Shireen-A. Davies; Regulation of a Drosophila melanogaster cGMP-specific phosphodiesterase by prenylation and interaction with a prenyl-binding protein. Biochem J 15 September 2008; 414 (3): 363–374. doi: https://doi.org/10.1042/BJ20080560

Keeling, C. I., Chiu, C. C., Aw, T., Li, M., Henderson, H., Tittiger, C., Weng, H. B., Blomquist, G. J., & Bohlmann, J. (2013). Frontalin pheromone biosynthesis in the mountain pine beetle, Dendroctonus ponderosae, and the role of isoprenyl diphosphate synthases. Proceedings of the National Academy of Sciences of the United States of America, 110(47), 18838–18843. https://doi.org/10.1073/pnas.1316498110

Koyama T, Ogura K (1999) Isopentenyl diphosphate isomerase and prenyltransferases. In: Barton D, Nakanishi K (eds) Comprehensive natural products chemistry, vol 2. Elsevier, Oxford, pp 69–96

Haney, S. L., Varney, M. L., Chhonker, Y. S., Shin, S., Mehla, K., Crawford, A. J., Smith, H. J., Smith, L. M., Murry, D. J., Hollingsworth, M. A., & Holstein, S. A. (2019). Inhibition of geranylgeranyl diphosphate synthase is a novel therapeutic strategy for pancreatic ductal adenocarcinoma. Oncogene, 38(26), 5308–5320. https://doi.org/10.1038/s41388-019-0794-6

Weissenrieder JS, Reilly JE, Neighbors JD, Hohl RJ. Inhibiting geranylgeranyl diphosphate synthesis reduces nuclear androgen receptor signaling and neuroendocrine differentiation in prostate cancer cell models. Prostate. 2019 Jan;79(1):21-30. doi: 10.1002/pros.23707. Epub 2018 Aug 14. PMID: 30106164.

Osborn-Heaford HL, Murthy S, Gu L, Larson-Casey JL, Ryan AJ, Shi L, Glogauer M, Neighbors JD, Hohl R, Carter AB. Targeting the isoprenoid pathway to abrogate progression of pulmonary fibrosis. Free Radic Biol Med. 2015 Sep;86:47-56. doi: 10.1016/j.freeradbiomed.2015.04.031. Epub 2015 May 7. PMID: 25958207; PMCID: PMC4554879.

Dudakovic, A., Tong, H. & Hohl, R.J. Geranylgeranyl diphosphate depletion inhibits breast cancer cell migration. *Invest New Drugs* **29,** 912–920 (2011). https://doi.org/10.1007/s10637-010-9446-y

Botella-Pavía P, Besumbes O, Phillips MA, Carretero-Paulet L, Boronat A, Rodríguez-Concepción M. Regulation of carotenoid biosynthesis in plants: evidence for a key role of hydroxymethylbutenyl diphosphate reductase in controlling the supply of plastidial

isoprenoid precursors. Plant J. 2004 Oct;40(2):188-99. doi: 10.1111/j.1365-313X.2004.02198.x. PMID: 15447646.

Lai, C., McMahon, R., Young, C., Mackay, T. F., & Langley, C. H. (1998). quemao, a Drosophila bristle locus, encodes geranylgeranyl pyrophosphate synthase. Genetics, 149(2), 1051–1061.

Sun ZJ, Li ZX. The terpenoid backbone biosynthesis pathway directly affects the biosynthesis of alarm pheromone in the aphid. Insect Mol Biol. 2018 Dec;27(6):824-834. doi: 10.1111/imb.12521. Epub 2018 Sep 12. PMID: 30039630.

Ding BY, Niu J, Shang F, Yang L, Chang TY, Wang JJ. Characterization of the Geranylgeranyl Diphosphate Synthase Gene in Acyrthosiphon pisum (Hemiptera: Aphididae) and Its Association With Carotenoid Biosynthesis. Front Physiol. 2019 Nov 12;10:1398. doi: 10.3389/fphys.2019.01398. PMID: 31780956; PMCID: PMC6861191.

Chang TH, Guo RT, Ko TP, Wang AH, Liang PH. Crystal structure of type-III geranylgeranyl pyrophosphate synthase from Saccharomyces cerevisiae and the mechanism of product chain length determination. J Biol Chem. 2006 May 26;281(21):14991-5000. doi: 10.1074/jbc.M512886200. Epub 2006 Mar 22. PMID: 16554305.

Kavanagh KL, Dunford JE, Bunkoczi G, Russell RGG, Oppermann U. The crystal structure of human geranylgeranyl pyrophosphate synthase reveals a novel hexameric arrangement and inhibitory product binding. J Biol Chem. 2006 Aug 4;281(31):22004-22012. doi: 10.1074/jbc.M602603200. Epub 2006 May 11. PMID: 16698791.

Barbar A, Couture M, Sen SE, Béliveau C, Nisole A, Bipfubusa M, Cusson M. Cloning, expression and characterization of an insect geranylgeranyl diphosphate synthase from Choristoneura fumiferana. Insect Biochem Mol Biol. 2013 Oct;43(10):947-58. doi: 10.1016/j.ibmb.2013.07.004. Epub 2013 Jul 29. PMID: 23907071.

Barja MV, Rodríguez-Concepción M. A Simple In Vitro Assay to Measure the Activity of Geranylgeranyl Diphosphate Synthase and Other Short-Chain Prenyltransferases. Methods Mol Biol. 2020;2083:27-38. doi: 10.1007/978-1-4939-9952-1_2. PMID: 31745910.

Schedler, David. (2006). Drug Discovery: A History (Sneader, Walter). Journal of Chemical Education - J CHEM EDUC. 83. 10.1021/ed083p215.1.

Cavasotto, C.N.; Phatak, S.S. Homology modeling in drug discovery: current trends and applications. Drug Discov. Today 2009, 14, 676–683. doi: 10.1016/j.drudis.2009.04.006.

Ivanov, A.S.; Veselovsky, A.V.; Dubanov, A.V.; Skvortsov, V.S. Bioinformatics platform development: from gene to lead compound. Methods Mol. Biol. 2006, 316, 389–431. doi: 10.1385/1-59259-964-8:389.

Wang, Y.; Shaikh, S.A.; Tajkhorshid, E. Exploring transmembrane diffusion pathways with molecular dynamics. Physiology 2010, 25, 142–154 12. doi: 10.1152/physiol.00046.2009.

J.W. Behaviour of small solutes and large drugs in a lipid bilayer from computer simulations. Biochim. Biophys. Acta 2005, 1718, 1–21. doi: 10.1016/j.bbamem.2005.07.009.

Wunberg, T.et al.(2006) Improving the hit-to-lead process: data-driven assessmentof drug-like and lead-like screening hits.Drug Discov. Today11, 175–180 doi: 10.1016/S1359-6446(05)03700-1.

Annis DA, Nazef N, Chuang CC, Scott MP, Nash HM. A general technique to rank protein-ligand binding affinities and determine allosteric versus direct binding site competition in compound mixtures. J Am Chem Soc. 2004 Dec 1;126(47):15495-503. doi: 10.1021/ja048365x. PMID: 15563178.

Ferreira LLG, Andricopulo AD. ADMET modeling approaches in drug discovery. Drug Discov Today. 2019 May;24(5):1157-1165. doi: 10.1016/j.drudis.2019.03.015. Epub 2019 Mar 16. PMID: 30890362.

Guo, Rey-Ting & Cao, Rong & Liang, po-huang & Ko, Tzu & Chang, Tao-Hsin & Hudock, Michael & Jeng, Wen-Yih & Chen, Cammy & Zhang, Yonghui & Song, Yongcheng & Kuo, Chih-Jung & Yin, Fenglin & Oldfield, Eric & Wang, Andrew. (2007). Bisphosphonates target multiple sites in both cis- and trans- prenyltransferases. Proceedings of the National Academy of Sciences of the United States of America. 104. 10022-7. doi: 10.1073/pnas.0702254104.

Pastor M, Cruciani G. A novel strategy for improving ligand selectivity in receptor-based drug design. J Med Chem. 1995 Nov 10;38(23):4637-47. doi: 10.1021/jm00023a003. PMID: 7473591.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013 Apr;30(4):772-80. doi: 10.1093/molbev/mst010. Epub 2013 Jan 16. PMID: 23329690; PMCID: PMC3603318.

Martinez X, Krone M, Alharbi N, Rose AS, Laramee RS, O'Donoghue S, Baaden M, Chavent M. Molecular Graphics: Bridging Structural Biologists and Computer Scientists. Structure. 2019 Nov 5;27(11):1617-1623. doi: 10.1016/j.str.2019.09.001. Epub 2019 Sep 26. PMID: 31564470.

Merk D, Grisoni F, Friedrich L, Gelzinyte E, Schneider G. Computer-Assisted Discovery of Retinoid X Receptor Modulating Natural Products and Isofunctional Mimetics. J Med Chem. 2018 Jun 28;61(12):5442-5447. doi: 10.1021/acs.jmedchem.8b00494. Epub 2018 Jun 14. PMID: 29901398.

R. F. Chapman, The insects, Structure and Function. 4th edition, 1998

Sezutsu H, Le Goff G, Feyereisen R. Origins of P450 diversity. Philos Trans R Soc Lond B Biol Sci. 2013 Jan 6;368(1612):20120428. doi: 10.1098/rstb.2012.0428. PMID: 23297351; PMCID: PMC3538418.

Nelson, D. R. & Nebert, D. W. Cytochrome P450 ( CYP ) Gene Superfamily. in eLS (ed. John Wiley & Sons Ltd) 1–19 (John Wiley & Sons, Ltd, 2018). doi:10.1002/9780470015902.a0005667.pub3.

Willingham, A. T. & Keil, T. A tissue specific cytochrome P450 required for the structure and function of Drosophila sensory organs. Mech. Dev. 121, 1289–1297 (2004).

Feyereisen, R. Origin and evolution of the CYP4G subfamily in insects, cytochrome P450 enzymes involved in cuticular hydrocarbon synthesis. Mol. Phylogenet. Evol. 143, 106695 (2020)

Paul R. Ortiz de Montellano, Cytochrome P450, Structure, Mechanism, and Biochemistry, 4th edition, 2015.

Singh, Kumar Saurabh, et al. "The Genetic Architecture of a Host Shift: An Adaptive Walk Protected an Aphid and Its Endosymbiont from Plant Chemical Defenses." Science Advances, vol. 6, no. 19, 2020, p. eaba1070. Crossref, doi:10.1126/sciadv.aba1070.

Li Z, Jiang Y, Guengerich FP, Ma L, Li S, Zhang W. Engineering cytochrome P450 enzyme systems for biomedical and biotechnological applications. J Biol Chem. 2020 Jan 17;295(3):833-849. doi: 10.1074/jbc.REV119.008758. Epub 2019 Dec 6. PMID: 31811088; PMCID: PMC6970918.

Beyer, Nina, et al. "P450BM3 Fused to Phosphite Dehydrogenase Allows Phosphite-Driven Selective Oxidations." Applied Microbiology and Biotechnology, vol. 101, no. 6, 2016, pp. 2319–31. Crossref, doi:10.1007/s00253-016-7993-7.

Feyereisen R. Arthropod CYPomes illustrate the tempo and mode in P450 evolution. Biochim Biophys Acta. 2011 Jan;1814(1):19-28. doi: 10.1016/j.bbapap.2010.06.012. Epub 2010 Jun 25. PMID: 20601227.

Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol. 1999 Sep;29(9):757-77. doi: 10.1016/s0965-1748(99)00038-7. PMID: 10510498.

Terriere LC. Induction of detoxication enzymes in insects. Annu Rev Entomol. 1984;29:71-88. doi: 10.1146/annurev.en.29.010184.000443. PMID: 6362551.

Zhu F, Feng JN, Zhang L, Liu N. Characterization of two novel cytochrome P450 genes in insecticide-resistant house-flies. Insect Mol Biol. 2008 Feb;17(1):27-37. doi: 10.1111/j.1365-2583.2008.00777.x. PMID: 18237282.

Zhu F, Liu N. Differential expression of CYP6A5 and CYP6A5v2 in pyrethroid-resistant house flies, Musca domestica. Arch Insect Biochem Physiol. 2008 Mar;67(3):107-19. doi: 10.1002/arch.20225. PMID: 18163524.

Festucci-Buselli RA, Carvalho-Dias AS, de Oliveira-Andrade M, Caixeta-Nunes C, Li HM, Stuart JJ, Muir W, Scharf ME, Pittendrigh BR. Expression of Cyp6g1 and Cyp12d1 in DDT resistant and susceptible strains of Drosophila melanogaster. Insect Mol Biol. 2005 Jan;14(1):69-77. doi: 10.1111/j.1365-2583.2005.00532.x. PMID: 15663776.

Liu N, Li T, Reid WR, Yang T, Zhang L. Multiple Cytochrome P450 genes: their constitutive overexpression and permethrin induction in insecticide resistant mosquitoes, Culex quinquefasciatus. PLoS One. 2011;6(8):e23403. doi: 10.1371/journal.pone.0023403. Epub 2011 Aug 12. PMID: 21858101; PMCID: PMC3155542.

Cochran DG. Effects of three synergists on pyrethroid resistance in the German cockroach (Dictyoptera: Blattellidae). J Econ Entomol. 1994 Aug;87(4):879-84. doi: 10.1093/jee/87.4.879. PMID: 8071470.

Yang T, Liu N. Genome analysis of cytochrome P450s and their expression profiles in insecticide resistant mosquitoes, Culex quinquefasciatus. PLoS One.

2011;6(12):e29418. doi: 10.1371/journal.pone.0029418. Epub 2011 Dec 29. PMID: 22242119; PMCID: PMC3248432.

Itokawa K, Komagata O, Kasai S, Okamura Y, Masada M, Tomita T. Genomic structures of Cyp9m10 in pyrethroid resistant and susceptible strains of Culex quinquefasciatus. Insect Biochem Mol Biol. 2010 Sep;40(9):631-40. doi: 10.1016/j.ibmb.2010.06.001. Epub 2010 Jun 18. PMID: 20600899.

Stevenson BJ, Pignatelli P, Nikou D, Paine MJ. Pinpointing P450s associated with pyrethroid metabolism in the dengue vector, Aedes aegypti: developing new tools to combat insecticide resistance. PLoS Negl Trop Dis. 2012;6(3):e1595. doi: 10.1371/journal.pntd.0001595. Epub 2012 Mar 27. PMID: 22479665; PMCID: PMC3313934.

Zhang L, Liu X, Wang C, Liu X, Cheng G, Wu Y. Expression, purification and direct eletrochemistry of cytochrome P450 6A1 from the house fly, Musca domestica. Protein Expr Purif. 2010 May;71(1):74-8. doi: 10.1016/j.pep.2009.12.008. Epub 2009 Dec 21. PMID: 20026277.

Tiwari P, Saini S, Upmanyu S, Benjamin B, Tandon R, Saini KS, Sahdev S. Enhanced expression of recombinant proteins utilizing a modified baculovirus expression vector. Mol Biotechnol. 2010 Sep;46(1):80-9. doi: 10.1007/s12033-010-9284-3. PMID: 20424933.

Mao W, Schuler MA, Berenbaum MR. CYP9Q-mediated detoxification of acaricides in the honey bee (Apis mellifera). Proc Natl Acad Sci U S A. 2011 Aug 2;108(31):12657-62. doi: 10.1073/pnas.1109535108. Epub 2011 Jul 20. PMID: 21775671; PMCID: PMC3150950.

Schuler MA, Berenbaum MR. Structure and function of cytochrome P450S in insect adaptation to natural and synthetic toxins: insights gained from molecular modeling. J Chem Ecol. 2013 Sep;39(9):1232-45. doi: 10.1007/s10886-013-0335-7. Epub 2013 Sep 14. PMID: 24036972.

Hiratsuka M. In vitro assessment of the allelic variants of cytochrome P450. Drug Metab Pharmacokinet. 2012;27(1):68-84. doi: 10.2133/dmpk.dmpk-11-rv-090. Epub 2011 Nov 1. PMID: 22041138.

D.R. Jones, Plant viruses transmitted by whiteflies, Eur. J. Plant Pathol. 109 (2003) 195–219.

I. Denholm, M. Cahill, T.J. Dennehy, A.R. Horowitz, Challenges with managing insecticide resistance in agricultural pests, exemplified by the whitefly Bemisia tabaci, Philos. Trans. R. Soc. Lond. B. Biol. Sci. 353 (1998) 1757–1767.

R. Horowitz, I. Denholm, S. Morin, Resistance to insecticides in the TYLCV vector, Bemisia tabaci, in: H. Czosnek (Ed.), Tomato Yellow Leaf Curl Virus Disease, Springer, Netherlands, 2007, pp. 305–325.

R. Nauen, M. Ghanim, I. Ishaaya, Whitefly special issue organized in two parts, Pest Manag. Sci. 70 (2014) 1438–1439.

I. Karunker, J. Benting, B. Lueke, T. Ponge, R. Nauen, E. Roditakis, et al., Overexpression of cytochrome P450 CYP6CM1 is associated with high resistance to imidacloprid in the B

and Q biotypes of Bemisia tabaci (Hemiptera: Aleyrodidae), Insect Biochem. Mol. Biol. 38 (2008) 634–644

I. Karunker, E. Morou, D. Nikou, R. Nauen, R. Sertchook, B.J. Stevenson, et al., Structural model and functional characterization of the Bemisia tabaci CYP6CM1vQ, a cytochrome P450 associated with high levels of imidacloprid resistance, Insect Biochem. Mol. Biol. 39 (2009) 697–706.

Hamada A, Wahl GD, Nesterov A, Nakao T, Kawashima M, Banba S. Differential metabolism of imidacloprid and dinotefuran by Bemisia tabaci CYP6CM1 variants. Pestic Biochem Physiol. 2019 Sep;159:27-33. doi: 10.1016/j.pestbp.2019.05.011. Epub 2019 May 20. PMID: 31400781.

Jeschke P, Nauen R, Gutbrod O, Beck ME, Matthiesen S, Haas M, Velten R. Flupyradifurone (Sivanto™) and its novel butenolide pharmacophore: Structural considerations. Pestic Biochem Physiol. 2015 Jun;121:31-8. doi: 10.1016/j.pestbp.2014.10.011. Epub 2014 Oct 24. PMID: 26047109.

Iyanagi, Takashi. “Structure and Function of NADPH-Cytochrome P450 Reductase and Nitric Oxide Synthase Reductase Domain.” *Biochemical and Biophysical Research Communications*, vol. 338, no. 1, 2005, pp. 520–28. *Crossref*, doi:10.1016/j.bbrc.2005.08.043.

Leferink NG, Pudney CR, Brenner S, Heyes DJ, Eady RR, Samar Hasnain S, Hay S, Rigby SE, Scrutton NS. Gating mechanisms for biological electron transfer: integrating structure with biophysics reveals the nature of redox control in cytochrome P450 reductase and copper-dependent nitrite reductase. FEBS Lett. 2012 Mar 9;586(5):578-84. doi: 10.1016/j.febslet.2011.07.003. Epub 2011 Jul 14. PMID: 21762695.

Ellis J, Gutierrez A, Barsukov IL, Huang WC, Grossmann JG, Roberts GCK. Domain motion in cytochrome P450 reductase: conformational equilibria revealed by NMR and small-angle x-ray scattering. J Biol Chem. 2009 Dec 25;284(52):36628-36637. doi: 10.1074/jbc.M109.054304. Epub 2009 Oct 26. PMID: 19858215; PMCID: PMC2794777.

Murataliev MB, Ariño A, Guzov VM, Feyereisen R. Kinetic mechanism of cytochrome P450 reductase from the house fly (Musca domestica). Insect Biochem Mol Biol. 1999 Mar;29(3):233-42. doi: 10.1016/s0965-1748(98)00131-3. PMID: 10319436.

Amprazi M, Kotsifaki D, Providaki M, Kapetaniou EG, Fellas G, Kyriazidis I, Pérez J, Kokkinidis M. Structural plasticity of 4-α-helical bundles exemplified by the puzzle-like molecular assembly of the Rop protein. Proc Natl Acad Sci U S A. 2014 Jul 29;111(30):11049-54. doi: 10.1073/pnas.1322065111. Epub 2014 Jul 14. PMID: 25024213; PMCID: PMC4121800.

Glykos NM, Papanikolau Y, Vlassi M, Kotsifaki D, Cesareni G, Kokkinidis M. Loopless Rop: structure and dynamics of an engineered homotetrameric variant of the repressor of primer protein. Biochemistry. 2006 Sep 12;45(36):10905-19. doi: 10.1021/bi060833n. PMID: 16953576.

Lassalle MW, Hinz HJ, Wenzel H, Vlassi M, Kokkinidis M, Cesareni G. Dimer-to-tetramer transformation: loop excision dramatically alters structure and stability of the ROP

four alpha-helix bundle protein. J Mol Biol. 1998 Jun 19;279(4):987-1000. doi: 10.1006/jmbi.1998.1776. PMID: 9642076.

Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA. Protein interaction maps for complete genomes based on gene fusion events. Nature. 1999 Nov 4;402(6757):86-90. doi: 10.1038/47056. PMID: 10573422.

Burns DM, Horn V, Paluh J, Yanofsky C. Evolution of the tryptophan synthetase of fungi. Analysis of experimentally fused Escherichia coli tryptophan synthetase alpha and beta chains. J Biol Chem. 1990 Feb 5;265(4):2060-9. PMID: 2404978.

Aalbers FS, Fraaije MW. Enzyme Fusions in Biocatalysis: Coupling Reactions by Pairing Enzymes. Chembiochem. 2019 Jan 2;20(1):20-28. doi: 10.1002/cbic.201800394. Epub 2018 Oct 2. PMID: 30178909; PMCID: PMC6563810.

Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. & Yura, T. Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE and GroELGroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2, inEscherichia coli. Appl. Environ. Microbiol. 64, 1694–1699 (1998). 53 79.

Nishihara, K., Kanemori, M., Yanagi, H. & Yura, T. Overexpression of Trigger Factor Prevents Aggregation of Recombinant Proteins in Escherichia coli. Appl. Environ. Microbiol. 66, 884–889 (2000).