6-1-2014

Polyketide Synthase III isolated from uncultured deep-sea Proteobacterium from the Red Sea- functional and evolutionary characterization

Hadeel Mansour El Bardisy

Follow this and additional works at: https://fount.aucegypt.edu/etds

The American University in Cairo

School of Sciences and Engineering

Polyketide Synthase III Isolated from Uncultured Deep-Sea Proteobacterium from the Red Sea – Functional and Evolutionary Characterization

A Thesis Submitted to

The Biotechnology Master's Program

In partial fulfillment of the requirements for

the degree of Master of Science

By Hadeel Mansour El Bardisy

Under the supervision of

Dr. Ahmed Moustafa

And

Dr. Ari José Scattone Ferreira

February / 2014

The American University in Cairo

A Thesis Submitted by

Hadeel Mansour El Bardisy

To the Biotechnology Graduate Program

February / 2014

In partial fulfillment of the requirements for
The degree of Master of Science

Has been approved by

Thesis Committee Supervisor/Chair _____

Affiliation _____

Thesis Committee Reader/Examiner _____

Affiliation _____

Thesis Committee Reader/Examiner _____

Affiliation _____

Thesis Committee Reader/External Examiner _____

Affiliation _____

_____   _____   _____ _____
Dept. Chair/Director   Date   Dean   Date

DEDICATION

*To my icons **Mother** and **Father***

*To my awesome sisters **Wesam** and **Nardine***

*To my lovely husband and cornerstone **Mohammed***

And definitely to my very dear friends who gave me direct or indirect support and craved wonderful memories through this very long journey!

ACKNOWLEDGMENT

ABSTRACT

The American University in Cairo
**Polyketide Synthase type III isolated from Uncultured Deep-Sea Proteobacterium
from the Red Sea – Functional and Evolutionary Characterization**
By Hadeel Mansour El Bardisy
Under the supervision of Dr. Ahmed Moustafa

Natural polyketide products are one of the major secondary metabolites produced among bacteria, fungi, and plants. They vary from flavonoids, pyrones, and stilbenes to phloroglucinols and resorcinols that are involved in important functions as antimicrobial activity, defense mechanisms and pigmentation. They are biosynthesized from acyl-CoA precursors by polyketides synthases (PKSs) that are categorized into 3 types: I, II and III. PKS type III is considered the simplest in its structure. It was believed that PKS type III was exclusively encoded by higher plants until the enduring efforts of bacterial genomes sequencing revealed the presence of more and more PKSs type III among them. There is an urge to investigate novel PKSs type III due to their promising polyketides of great biological and pharmaceutical advantages. This allowed metagenomic approaches to be a valuable tool to explore diverse environments for PKSs type III. Extreme environments as deep sea brine pools could probe unique natural polyketides capable of functioning in such conditions with valuable biotechnological and pharmaceutical applications.

In this study, screening of the Lower Convective Layer (LCL) of Atlantis II (ATII) deep brine pool in the Red Sea was done. It identified sequences belonging to bacterial PKSs type III. A candidate encoding sequence was amplified from the environmental DNA. Functional annotations were assigned to the translated open reading frame including the conserved catalytic triad, domains, motifs and 3D modelling. Preliminary structural analysis showed well-fitted superimposition with the flowering plant *Medicago sativa* PKS type III crystal structure and predicted the interaction of the catalytic triad with the most common substrate malonyl-CoA. Further optimization of heterologous expression is required to investigate this isolated PKS type III functional activity.

In an approach to gain better insights into the enzyme's unresolved evolutionary origin, a comprehensive phylogenetic analysis was conducted. The analysis pinpoints the possible involvement of symbiotic bacterium *Parachlamydia acanthamoebae* in horizontal gene transfer events to eukaryotes. On the other hand, the sequence isolated from ATII brine pool was clustered in a clade with related PKSs type III sequences belonging to alpha-proteobacteria. Environmental assessment of PKSs type III abundance in ATII and nearby Discovery Deep (DD) brine pool revealed the presence of PKSs type III in ATII only, where most sequences were located in the LCL. This could be attributed to the high aromatic content within the brine as possible substrates for the enzyme. Based on these analyses, we could propose ATII microbial community as a unique source for natural polyketides.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS LIST

ACP   Acyl Carrier Protein

AT   Acyl Transferase

ATII   Atlantis II brine pool

BAC   Bacterial Artificial Chromosome

BCA   Bicinchoninic Acid

BLAST   Basic Local Alignment Search Tool

BSA   Bovine Serum Albumin

CDD   Conserved Domains Database

CHS   Chalcone Synthase

CoA   Co enzyme A

CTD   Conductivity, Temperature and Depth

DAPG   2,4 Di-acetyl-phloro-glucinol

DD   Discovery Deep brine pool

DH   β-Hydroxy-acyl Dehydratase

DHPG   Di-hydroxy-phenyl-glycine

DNA   Deoxyribonucleic Acid

DOPE   Discrete Optimized Protein Energy

ER   Enoyl Reductase

Gcs   Germicidin Synthase

HGT   Horizontal Gene Transfer

HMM   Hidden Markov Model

HPQ   Hexa-hydroxy-perylene-quinone

HPLC   High Performance Liquid Chromatography

INP   Interphase Layer

IPTG   Isopropyl β-D-1-thiogalactopyranoside

KAS or KS Keto Acyl Synthase

KR	β-Ketoacyl Reductase
KAUST	King Abdullah University of Science and Technology
LB	Luria-Bertani
LCL	Lower Convective Layer
Mpa	Megapascal
molpdf	Molecular Probability Density Function
NCBI	National Center for Biotechnology Information
Ni-NTA	Nickel Nitrilotriacetic Acid
NT	Nucleotide
OD	Optical Density
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
Pfam	Protein Families
PDB	Protein Data Bank
PKS	Polyketide Synthase
PMSF	Phenylmethylsulfonyl fluoride
PhlD	Phloroglucinol Synthase
rRNA	Ribosomal Ribonucleic Acid
RppA	Red pigment production
SD	Shine Dalgarno
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SrsA	Streptomyces resorcinol synthase
Ta	Temperature annealing
THNS	1, 3, 6, 8-Tetra-hydroxy-naphthalene synthase
UCL	Upper Convective Layer
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

AMINO ACID ABBREVIATIONS

A, Ala   Alanine

C, Cys   Cysteine

D, Asp   Aspartic acid

E, Glu   Glutamic acid

F,Phe   Phenylalanine

G,Gly   Glycine

H,His   Histidine

I, lie   Isoleucine

K,Lys   Lysine

L, Leu   Leucine

M, Met   Methionine

N, Asn   Asparagine

P, Pro   Proline

Q, Gin   Glutamine

R,Arg   Arginine

S, Ser   Serine

T,Thr   Threonine

V,Val   Valine

W,Trp   Tryptophan

Y,Tyr   Tyrosine

CHAPTER 1: LITERATURE REVIEW

Microorganisms play a central role in the climate, geological, geochemical and biological evolution on earth (Xu, 2006). Through diverse metabolic activities, microorganisms are also considered a rich source of natural products (Roessner & Scott, 1996). Natural products like polyketides confer a series of defensive, communicative, antibiotic, antitumor, antifungal, immunosuppressant, other pharmacological and biological advantages (Austin & Noel, 2003; Demain, 1999). The discovery of natural polyketides from diverse environments still remains challenging in terms of screening approaches (Zhao, Yang, & Zeng, 2008). Several efforts have been employed to overcome these challenges as applying metagenomic approaches. In general, metagenomics is considered one of the successful techniques for investigating diverse environments especially extreme ones that exhibit challenging conditions for life survival (Zhao et al., 2008). Extreme environments could provide a novel source for unique enzymes and natural polyketides produced by their microbial community adapted to such extreme habitats (Niehaus, Bertoldo, Kähler, & Antranikian, 1999).

In this chapter, a brief review on the importance of metagenomics as a screening tool for diverse environments will be covered. The common approaches conducted in metagenomic studies are also discussed highlighting the differences between them. A main focus of this review is the Red Sea and its brine pools characteristics especially Atlantis II brine pool (the brine of interest). A deeper insight into the hydrothermal nature of the brine pool promoting itself as a candidate source for natural polyketides is examined. Then, a comprehensive review on PKSs is covered where PKSs type III is explained in details. Finally, an additional insight into bacterial PKS type III is discussed with major focus on its diversity, evolutionary origin and biological importance.

1. Metagenomics: The paradigm shift

Estimation of only 1% of microorganisms being cultured disclose the limitation of common culturable approaches to well study microbial diversity (Schloss & Handelsman, 2005). Shifting to cultural independent approaches was a key to access the unexploited microbial biodiversity. These approaches provided an ample understanding for microbial abundance, entire metabolic pathways with their associated products and novel enzymes discovery (Cowan et al., 2005). **Metagenomics** - or the cultural independent genetic analysis of microbial genomes - greased the wheels to explore various environmental samples from microorganism's aspect.

The idea of cultural-deprived analysis of environmental samples was first proposed by Pace and his colleagues in 1985 (Lane et al., 1985). Later on, Giovannoni *et al* utilized this approach in phylogenetic analysis of the DNA purified from the Sargasso sea picoplankton by using ^{32}P-labeled 16S rRNA hybridization probes directly on the environmental samples (Giovannoni, DeLong, Schmidt, & Pace, 1990). Several studies showed the applicability of direct 16S rRNA genes amplification in reforming bacterial classifications and recovery of novel taxa (Eden, Schmidt, Blakemore, & Pace, 1991; Redburn & Patel, 1993; Schmidt, DeLong, & Pace, 1991).

The main concept behind metagenomics is to isolate the genetic material from the environment of interest. Then, undergoes direct sequencing approaches or being cloned into suitable vectors, constructing metagenomic libraries to be further analyzed (Kennedy, Marchesi, & Dobson, 2008). By 1998, "**Metagenomics**" term became renowned (Handelsman, Rondon, Brady, Clardy, & Goodman, 1998) and paved the path for investigating diverse environments. A glimpse of several metagenomics studies carried in different ecosystems is enumerated in the following table.

Table (1) Metagenomic studies in different ecosystems:

Enviromental Conditions	Location	Study finding	Reference
Low temperature	Ace Lake, Vestfold Hills, Antarctica	Isolation of Cold-adaptive Antarctic GSB (Green Sulphur Bacteria)	(Ng et al., 2010)
High temperature and High pressure	Lost City Hydrothermal Field on the Mid-Atlantic Ridge	Abundance of transposases genes in carbonate chimneys	(Brazelton & Baross, 2009)
High alkaline medium	Kenya soda lakes	Detection of novel cellulases genes	(Rees, Grant, Jones, Grant, & Heaphy, 2003)
Fresh water	Amazon River	Species that dominate the amazon river	(Ghai et al., 2011)
Diverse environment	forest soil from Jindong Valley, Korea	Isolation of novel anti-leukemia agent indirubin	(Lim et al., 2005)
Enriched environment	Human gut	Establishing a large scale human gut microbial gene catalogue for understanding bowl diseases	(Qin et al., 2010)
High acidic medium	Acid mine drainage in Rio Tinto,Spain	Detection of novel nickel resistance genes	(Mirete, Figueras, & González-Pastor, 2007)
Low temperature (Temperate Sea)	Baltic Sea sediment	Detection of novel lipase	(Hårdeman & Sjöling, 2007)
Low temperature and High pressure	East Pacific deep sea sediments	Detection of novel alkane hydroxylases	(M. Xu, Xiao, & Wang, 2008)

1.1 Shotgun sequencing technologies:

Utilization of shotgun sequencing became fundamental within metagenomic studies for better understanding of microbial diversity (Venter et al., 2004) and biological function prediction (DeLong et al., 2006). At first, studies relied on Sanger dideoxynucleotide terminated sequencing technology. However, it was time consuming

and expensive compared to the Next-Generation Sequencing (NGS) platforms such as Roche/454 FLX Pyrosequencer, Illumina Genome Analyzer, and Applied Biosystems SOLiD™ Sequencer (Adams et al., 2009; Mardis, 2008). Pyrosequencing, the massively parallel sequencing technology has been extensively coupled in many recent metagenomic studies (Biddle, Fitz-Gibbon, Schuster, Brenchley, & House, 2008; Poinar et al., 2006; Roesch et al., 2007; Sogin et al., 2006). Although pyrosequencing and other NGS techniques could generate shorter reads (approximately 400 bp read length) than the traditional Sanger sequencing (almost 700 read length) , however more intense magnitude of reads are generated for better coverage of the template(s) (Schuster, 2008). Besides, the expedite approach for direct sequencing by pyrosequencing eliminates unbiased representation of the metagenomic sample with dominance of one species over other and facilitate detection of low abundant microorganisms (Huse et al., 2008) .

The generation of huge metagenomic data helped in developing bioinformatics tools necessary for analyzing this data (Kunin, Copeland, Lapidus, Mavromatis, & Hugenholtz, 2008). The computational analysis includes quality control, reads assembly, gene prediction and annotation, whole genomes assembly, functional diversity and taxonomical classification and comparative genomics studies (Wooley & Ye, 2009).

1.2 Metagenomics for novel discoveries:

One of the main applications of metagenomics is intended to quest promising biomolecules and enzymes (Wooley & Ye, 2009). Exploiting extreme environments are driven with great sympathy as an unvisited source of promising novel biomolecules and enzymes that function under ruthless conditions (Lewin, Wentzel, & Valla, 2013). The extreme conditions vary from pressure, temperature, salinity and radiation to presence of heavy metals; mostly found in hydrothermal vents, brine pools, acid mines, hot springs, glaciers soil, glaciers ice, hypersaline basins and many others (Rothschild & Mancinelli, 2001; Simon & Daniel, 2011), see Table (1). These isolated biomolecules and enzymes

could be considered optimal choices for many industrial and bioprospecting purposes (Lewin et al., 2013).

1.3 Metagenomic library construction:

To identify novel enzymes and biomolecules from metagenomic samples, two targeted approaches are implemented: functional screening and homology screening (Simon & Daniel, 2011). Construction of metagenomic libraries is the basic step to facilitate these approaches (Streit & Schmitz, 2004). Environmental DNA is cloned to an adequate vector system according to the overall goal of screening (Schmeisser, Steele, & Streit, 2007). If the goal is to screen large gene clusters or operons, large insert libraries will be suitable utilizing vectors such as cosmids, fosmids (up to 40 Kb insert) or bacterial artificial chromosomes (BAC) (more than 40 kb insert). Screening for individual genes or small operons (up to 15 kb) requires small insert libraries utilizing vectors such as pUC19, pZero-2 and ZAP phagemid vectors (Steele, Jaeger, Daniel, & Streit, 2009).

In functional screening approach, mining for the catalytic activity of the novel enzymes comes prior identifying the responsible genes. It depends on observing a particular phenotype on the metagenomic cloned strains after employing a specific substrate as color detection or a halo-zone formation (Simon & Daniel, 2011). In such fashion, peculiar enzymes could be discovered with no constraining for a homology to known enzymes. However, it would be still confined to the accuracy of the heterologous host (Simon & Daniel, 2011). Although *Escherichia coli* is considered one of the well-established hosts used in functional screening studies, owing to its exemplified genetics and easy transformation (Steele et al., 2009); however, successful expression is not always the outcome. Lack of adequate regulatory factors for proper transcription and translation, toxicity to the host upon expression, and protein misfolding are common factors that could not be solved by classical *E. coli* expression (Handelsman, Liles, Mann, Riesenfeld, & Goodman, 2002).

In homology-based screening, DNA primers or hybridization probes designed on the conserved regions of known genes are utilized. This eradicates the possibility of novel biocatalysts detection but only new variants of already known functional proteins (Simon & Daniel, 2011). Once positive amplicons are obtained, upstream and downstream walking primers are designed to recover the entire gene (Steele et al., 2009). Being self-autonomous from heterologous host utilization, it overcomes the expression limitations experienced in functional based screening approach (Steele et al., 2009).

2. Marine metagenomics: a superb source for unique biocatalysts

With 71% of earth's surface attributed to oceans, this untapped marine environment comprises an immense source for diverse bacteria (Kennedy et al., 2010). Generally, the oceans host is considered very diverse ecosystems in terms of temperature, salinity, pressure and nutrients. Such environments delineate microbial adaptation and distribution among marine habitats ranging from normal sunlit surfaces to very deep trenches (Kennedy et al., 2008). One of the unique marine environments in oceans is hydrothermal brine pools (Dick et al., 2013). The extreme conditions within these pools mark their microbial habitat as distinct microorganisms with unprecedented adaptation.

2.1 The Red Sea and its hydrothermal brine systems

Red Sea is a semi-isolated slender shaped sea and the inlet of the Indian Ocean. It represents one of the warmest and saltiest marine environments owing to its arid regional location. This is due to rareness of rainfalls, the elevated evaporation rate and absence of major river inflows (Qian et al., 2011). The long deep trench within the bottom of the Red Sea stands behind its exceptional features. This trench resulted from the deviated movement of the Arabian and African tectonic plate layers during volcanic activity lead to Red Sea formation (Antunes, Ngugi, & Stingl, 2011). The leakiness of geothermal

solution within this seafloor trench induced the formation of deep sea brine pools (Ramboz, Oudin, & Thisse, 1988).

Up to date, this trench has been geologically active, comprising twenty-five brine pools (Hartmann, Scholten, Stoffers, & Wehner, 1998; Wang et al., 2011). These brine pools are extremely hydrothermal due to the integration of hot and metalliferous submarine solution with deep sea water (Wang et al., 2013). This integration generates convective currents that form distinctive layers within the brine with different density. Besides being extremely saline and hot, these brines are anoxic and rich in heavy metals deposit (Antunes et al., 2011).

Atlantis II deep is considered one of the largest brines in the Red Sea with a maximum depth of 2194 meter and a 60-km^2 wide, (Figure 1) (Antunes et al., 2011). A gradual increase in Atlantis II (ATII) temperature was reported from 56°C in 1965 to 71°C in 1995. This gradual increase was not observed in an adjacent brine pool called Discovery Deep (DD). A fairly constant change from 44.7°C in 1977 to 50°C in 1995 was only reported within DD (Hartmann et al., 1998; Wang et al., 2011).ATII elevated temperature is due to the presence of fossil hydrothermal veins beneath the pool with temperature range 195°C-310°C pushing hot deep sea water up to the brine (Anschutz & Blanc, 1996). The brine pool is stratified into several distinctive layers exhibiting a stepwise rise in temperature and salinity towards the bottom (Anschutz & Blanc, 1996). Three upper convective layers (UCLs) [Interphase, Upper and Middle convective layers] and a lower convective layer (LCL) were classified that configure this brine pool (Blanc & Anschutz, 1995). The LCL of ATII is characterized by a unique 68.2°C temperature, high 41 Mpa pressure, salinity of 25.7% and pH of 5.3 (Antunes et al., 2011).

Usually the suitable temperature range for aromatic compounds formation is from 60°C to 150°C and pressure exceeding 150 bar (Simoneit, 1993). Thus, both ATII brine's hydrothermal effect and high pressure give a satisfactory environment for aromatic compounds formation from organic debris reduction (Wang et al., 2011). In comparison to

DD, higher percentage of aromatic compounds was found in ATII. This high percentage gives an adaptive privilege of microbial inhabitants to consider these aromatic compounds as nutrient source (Wang et al., 2011). Accordingly, ATII is an excellent environment to screen for enzymes involved in metabolic processes utilizing aromatic compounds as substrates.
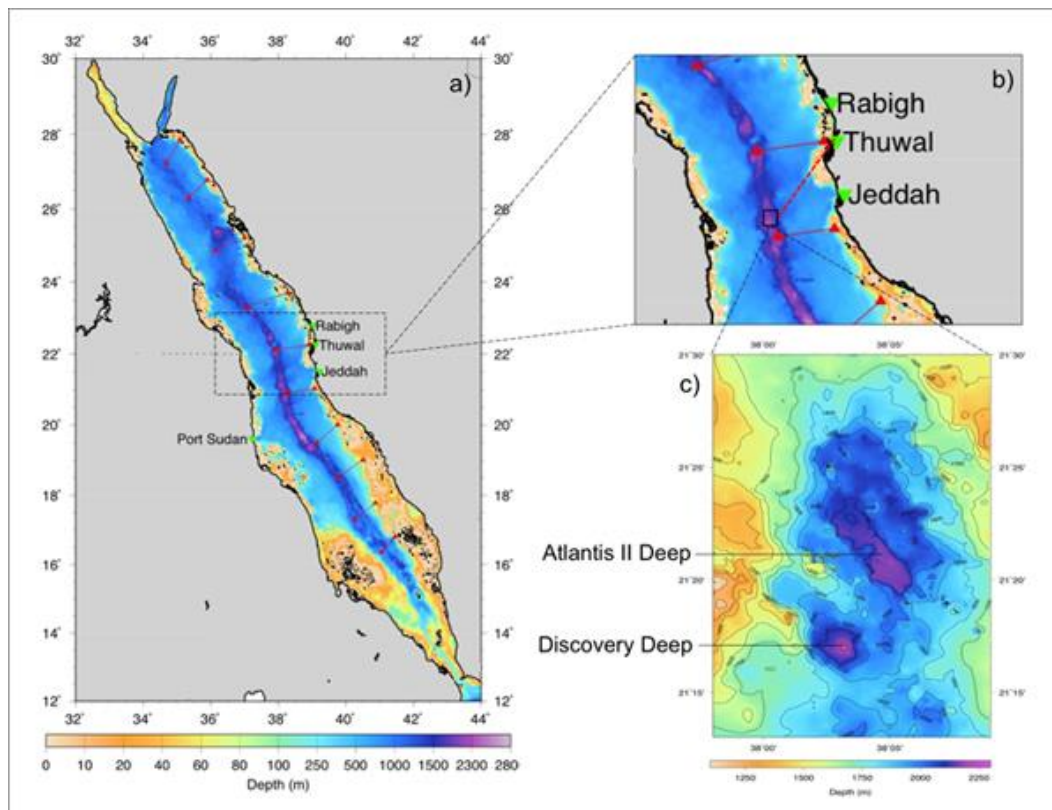
Figure (1) Red Sea and location of Atlantis II Deep brine pool (ATII) and Discovery Deep brine pool (DD) The location of the ATII and DD are adjacent to each other; Intensity of the blue color indicates the depth of the brines. This figure is adapted from http://krse.kaust.edu.sa/spring-2010/mission.html

3. Metagenomics and recovery of natural polyketide enzymes

One of the earliest metagenomics studies done to recover natural polyketide enzymes was conducted by Courtois *et al* (2003). They screened for novel antitumor polyketides in environmental soil. This group was able to isolate twelve novel genes

coding for modular PKS-type I (PKSs types will be discussed shortly) by PCR targeted search (Courtois et al., 2003). This study opened the way for the discovery of novel drugs and natural polyketides by metagenomic approaches (Steele et al., 2009). For instance, another soil metagenomic study succeeded in obtaining 139 (out of 60,000) clones coding for polyketide synthases (Ginolhac et al., 2004). Also, a metagenomic study on agricultural soil isolated novel PKSs related to *Acidobacteria* phylum. Such finding suggested soil *Acidobacteria* could be a novel source for PKSs (Parsley et al., 2011). Another phylogenetic study revealed the diversity of PKSs in the Antarctic sediments. It was considered the first study for PKSs in extreme environments (Zhao et al., 2008).

Other metagenomic polyketide investigations focused on bacterial symbionts in earthbound beetles or marine invertebrates as a powerful source for natural antitumor products (Steele et al., 2009).Through investigations, it was proven that these bacterial symbionts are the actual producers for antitumor natural molecules not the animal invertebrates (Piel et al., 2005). However, these bacterial symbionts were considered a challenge to study before the momentum of metagenomics (Piel et al., 2005). Metagenomic investigations revealed the presence of a putative antitumor pederin PKS enzyme from uncultured bacterial symbiont of *Paederus fuscipes* beetles and marine sponge *Theonella swinhoei* (Piel, Hui, Fusetani, & Matsunaga, 2004). Discodermolide, another promising polyketide antitumor agent was isolated from the microbial community associated with the marine sponge *Discodermia dissoluta* (Schirmer et al., 2005). Several studies in this breadth has been employed (Brady, Simmons, Kim, & Schmidt, 2009; Moffitt & Neilan, 2003). So far, investigating PKSs is mainly focused on PKS type I and type II in metagenomic studies.

4. Polyketides biosynthesis by PKSs:

PKSs are classified into 3 types: I, II, and III according to their different structure, sequences and catalytic mechanisms (Yu, Xu, Zeng, & Zhan, 2012). Generally, the three types catalyze a plausible sequence of decarboxylative condensation reactions initiated by

two carbon acetate units: starter and extender units. These units link with each other forming polyketide chain (Staunton & Weissman, 2001).

The precursors (Starter units) or intermediates (Extender units) are maintained as thioesters conjugates by two different carrier molecules, either coenzyme A (CoA) or acyl carrier protein (ACP). These carrier molecules facilitate loading of the acyl substrates (starter and extender units) to the PKS active site. The thioester bond is formed between the phosphopantetheine moiety within the carrier molecules and the acyl substrate. The phosphopantetheine moiety acts as long arm that permits flexible unconstrained mobility of the acyl substrates to the catalytic active sites of the PKS enzymes. Molecular carrier 'CoA' or 'ACP' are added to acyl substrates by acyltranferase (AT) activity. PKSs prefer CoA or ACP molecular carrier according to its type (Austin, 2005; Staunton & Weissman, 2001).

Typical PKSs comprise the catalytic machinery β-ketoacyl synthase (KS) domain where the clasien condensation reactions take place (Austin & Noel, 2003). Clasien condensation is a reaction between α-carbon of an ester and a carbon of a carbonyl group for another ester, ketone or aldehyde where α- carbon nucleophile attacks the electrophile carbon of the carbonyl group establishing the carbon-carbon bond formation (Jiang, Kim, & Suh, 2008). In PKSs, it is the nucleophile attack of α- carbon of the extender unit on the electrophile carbon of the starter unit (Hopwood & Sherman, 1990).

The KS domain comprises two thiol loci, the CoA or ACP compartment where the extender units are bound and the catalytic cysteine active site compartment where starter units are bound (Hopwood & Sherman, 1990). This catalytic cysteine active site is responsible for the clasien condensation between acetyl unit (C_2-starter) and the malonyl-CoA (C_3-extender unit) resulting in a C_4 diketide intermediate after CO_2 removal. This diketide could either enter another condensation round extending its growing chain forming a triketide, tetraketide, pentaketide and so on or undergoes a series of reduction

reactions. The distal β-keto carbonyl group of the growing chain undergoes three reduction cycles: First, β-ketoacyl reductase (KR) catalyzes the production of a β-hydroxyl group. Second, an unsaturated double bond is formed by β-hydroxy-acyl dehydratase (DH) after water removal. Finally, complete reduction of the β-carbon is finalized by an enoyl reductase (ER). After elongation, the polyketide chain undergoes intermolecular cyclization producing vast mono, poly-cyclic and aromatic products from such simple building acetate blocks (Austin & Noel, 2003; Hopwood & Sherman, 1990; Staunton & Weissman, 2001).

The theme of these reactions highly resemble the biosynthetic pathway of fatty acid synthase (FAS) enzymes for primary metabolism (Austin & Noel, 2003). FAS prefers molecular carrier ACP to CoA so the substrates are maintained as acyl-ACP to be loaded to KS domain. After the formation of diketide, The distal β-keto carbonyl group of the growing chain is first reduced by the three reduction cycles: KR,DH and ER then acyl-ACP bound chain enter several elongation rounds of condensation, dehydration and reduction until the desired length chain is optimized (Austin, 2005; Staunton & Weissman, 2001). Finally, a $C_{14} - C_{18}$ free fatty acid or ester is formed (Austin & Noel, 2003; Hopwood & Sherman, 1990; Staunton & Weissman, 2001). The mechanism of polyketide biosynthesis by PKSs in contrast to FASs is illustrated in Figure (2).

The strong homology between FAS and PKS enzyme's mechanistic pathway and common fold reinforce the fact that all three types of PKSs are prone to be evolved from these distant related enzymes (Austin & Noel, 2003). However unlike FAS, PKSs enzymes are considered "highly programmed" system with an enriched variety of choices during the polyketide chain assembly (Shen, 2000). It could either keep the β-ketone carbon group before entering a new elongation round or be subjected to a pattern of semi or fully reduction modifications. This occurs according to starters and extenders choice, number of required β-reduction and length of the growing polyketide chain (Hranueli et al., 2001). Thus, it is believed that they evolved from FAS pathway via partial gene duplication (Austin, 2005; Hopwood & Sherman, 1990; Shen, 2000).
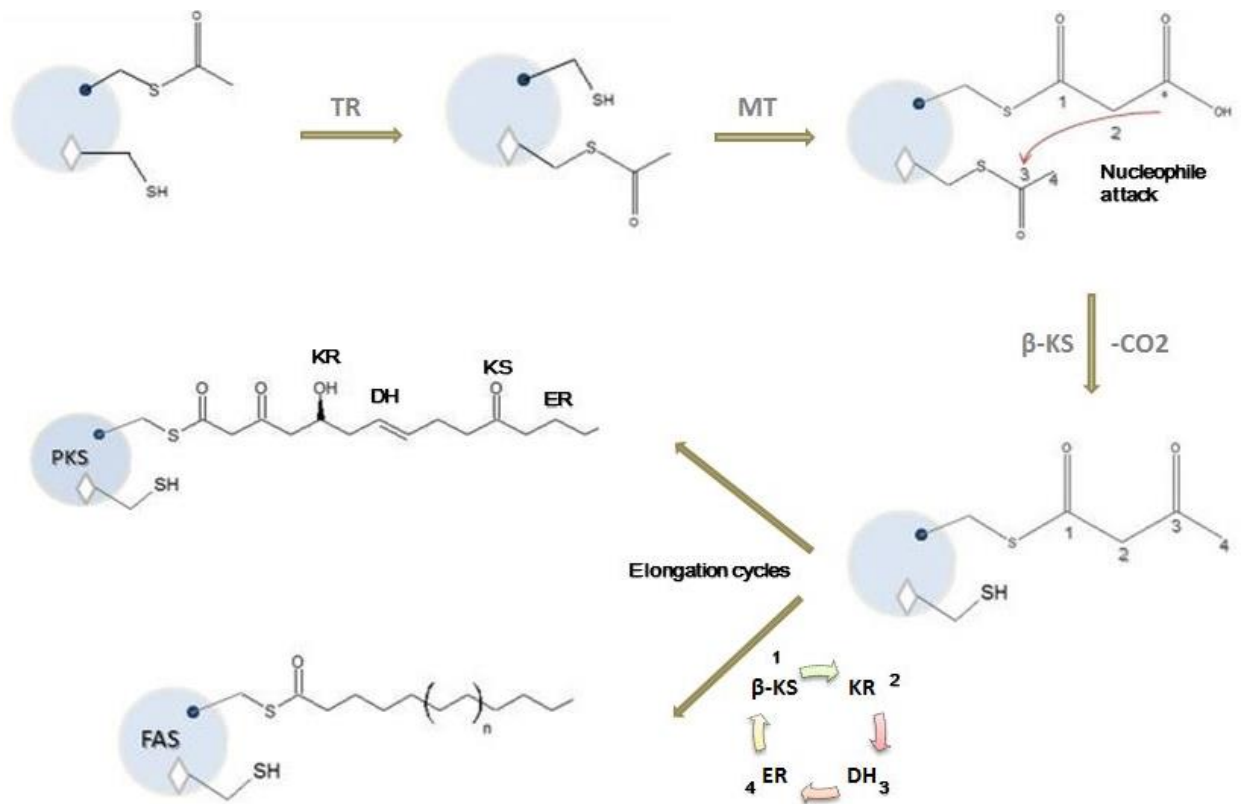
Figure (2) Schematic diagram for the biosynthesis of polyketides by PKSs in contrast to FASs

The blue circle represent PKS or FAS enzyme with two thiol loci where the blue dot is the CoA or ACP attaching site and the diamond is the catalytic cysteine active site within KS domain. Acetyl CoA (C_2-starter) is transferred to catalytic active site by transferases (TR). Malonyl transferase (MT) attaches malonyl CoA (C_3-Exteder) to CoA attaching site. A condensation reaction occurs by nucleophilic attack of α carbon of malonyl extender to electrophile carbonyl group of acetyl starter where a diketide is formed. In PKSs, the polyketide undergoes either elongation (β-KS) or reduction rounds (KR, DH and ER) while in FAS reduction rounds first reduces β-carbonyl group before it enters a new elongation round in an arranged fashion.

5. Polyketide Synthases (PKSs) Type III:

5.1 PKSs type III simple structure:

PKSs type III are main producers of many aromatic metabolites in plants, bacteria and fungi such as flavonoids, pyrones, stilbenes, phloroglucinols and resorcinols that confer beneficial advantages like pigmentation, defensive agent, antimicrobial activity, UV resistance, etc (Yu et al., 2012). They preserve an unsophisticated structure as they form homodimers with two equivalent KS monomers. Within each monomer's sole active site, multifunctional loading, elongation and cyclization reactions are performed in an iterative fashion producing polyketide products (Austin & Noel, 2003). This simple synthase enzyme is quite dissimilar than type I and type II PKSs in genetic organization and architecture, (Figure 3). PKSs type I are large multi-functional polypeptides where each possess a set of modules. Each module acquire from three to six non-iterative (modular) or iterative catalytic domains responsible for a single elongation round (Hranueli et al., 2001). Modular (non-iterative) or iterative PKSs type I is based upon whether the catalytic domains will be re-utilized in the next elongation cycle or not (Donadio, Staver, Mcalpine, Swanson, & Katz, 1991). PKSs type II form a single multi-enzyme complex that is used iteratively for aromatic polyketide production. PKSs type II have a 'minimal PKS' that is highly conservative. It harbors two keto-synthase units (KSα and KSβ) and an acyl carrier protein (ACP). Although KSα and KSβ exploit high sequence similarity, KSβ is responsible for the chain length factor (CLF) for the growing chain (Hertweck, Luzhetsky, Rebets, & Bechthold, 2007). Collectively, the structures of PKSs types I and II are more sophisticated than that of PKSs type III.

Also, PKSs type III are capable of catalyzing complete condensation cycles within its active site in contrast to the other two PKSs sophisticated condensation cycles with various catalytic domains involved. Thus, they are considered mechanistically complex rather than architecturally complex (Austin & Noel, 2003; Katsuyama, 2010). Finally, PKSs type III can utilize acyl CoA substrates unlike acyl ACP substrates used by PKSs type I and II (Staunton & Weissman, 2001).
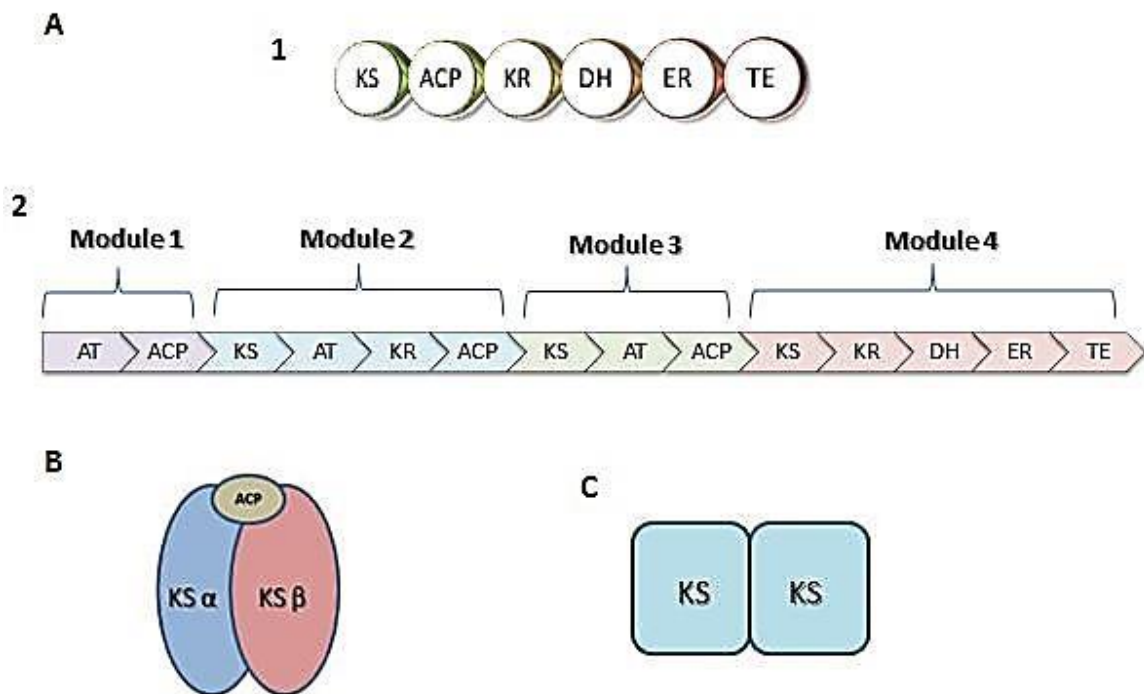
Figure (3) Schematic diagram for the three PKSs types Structure

A1 Iterative PKS type I. **A2** Modular PKS type I (non-iterative). **B** Minimal PKS type II. **C** Homodimer PKS type III.

5.2 Ancestral fatty acid keto-acyl synthases (KASs) type III

It has been hypothesized that PKSs type III retrieved their functionality from structurally related homodimeric fatty acid KASs type III. KASs type III utilize acyl-CoA substrates unlike other KASs in fatty acid metabolism (Austin & Noel, 2003). Both KASs type III and PKSs type III confer a Cys-His-Asn (CHN) catalytic triad essential for activity (Austin & Noel, 2003; Jiang et al., 2008). Analysis of their crystal structures revealed an overall homology despite low sequence similarity suggesting they are evolutionary related (Jiang et al., 2008). Usually they exist as dimers, where each monomer is assembled into five layered α-β-α-β-α core architecture. The β-strands are found antiparallel at the dimer interface and the fully conserved catalytic cysteine is

found at the N-terminal region. The similarity not only includes the overall core architecture, but also position of the catalytic triad. It is located in similar spatial configuration in both KASs type III and PKSs type III (Dawe, Porter, Thornton, & Tabor, 2003). The remaining catalytic loops and residues are allocated at the C-terminal region (Haapalainen, Meriläinen, & Wierenga, 2006). Main differences include the extent of catalytic loops at the C-terminal and number of other active residues involved in biosynthesis (Austin & Noel, 2003).

5.3 Mechanism of PKSs type III catalytic triad

Crystallographic analysis of various PKSs type III revealed the presence of four primary catalytic residues conserved among all known CHS-related enzymes: Cys, His, Asn (Known as the catalytic triad) and Phe (Jez et al., 2001; Jez, Dixon, Bowman, & Ferrer, 1999; Jez, Ferrer, Bowman, Dixon, & Noel, 2000). These important four residues are located at the intersection of CoA binding site (also referred as CoA binding tunnel) and the active site cavity (Jez et al., 2001). These residues play an essential role in acyl-CoA decarboxylation and polyketide formation.

First, the thiolate group of Cys confers a nucleophilic attack on the thioester carbonyl group of the starter acyl-CoA. By this attack, the starter acyl moiety disassociates from CoA loci and transfers to the Cys side chain (Abe & Morita, 2010). The thiolate anion is preserved through an ionic interaction with imidazolium cation of His (Jez & Noel, 2000). The thioester carbonyl is easily accessible to Cys residue through hydrogen bonds with side chains of Asn and His (Jez et al., 2001). This stabilizes the transition phase of the acyl-CoA substrate during the Cys nucleophilic attack (Jeya et al., 2012). The same hydrogen bonds are formed later with intermediate extender units for elongation (Jez et al., 2001).

Second, intermediate extender unit (usually it is malonyl-CoA) orients its carbonyl thioester adjacent to the Cys bound acyl moiety. Asn and His residues together with Phe residue facilitate decarboxylation step of malonyl-CoA (Jez et al., 2001). Phe residue is called the "gate keeper" because it provides proper orientation of the intermediates (Austin & Noel, 2003). Meanwhile, Asn and His residues aid in stabilizing this orientation through hydrogen bonds. This sort of stabilization creates an oxyanion hole that keeps the negative charge during decarboxylation step. This is achieved via maintaining the enol tautomer form of the acetyl anion that in return facilitates CO_2 removal (Jez et al., 2001).

Third, carbonyl thioester of the Cys bound acyl moiety is attacked by the carbanion releasing free Cys thiolate nucleophile. The diketide chain produced rebounds once again to Cys by thiolate nucleophilic attack on carbonyl thioester of diketide. More intermediates incorporate for several elongation rounds producing polyketide chain (Abe & Morita, 2010).

Finally, the polyketide chain could undergo various cyclization patterns. Clasien cyclization occurs between C_1 and C_6. Meanwhile, aldol condensation occurs between C_2 and C_7 (Samappito, Page, Schmidt, De-Eknamkul, & Kutchan, 2003). Lactone cyclization occurs between C_1 and C_5 (Akiyama, Shibuya, Liu, & Ebizuka, 1999). The overall proposed mechanism of PKSs type III catalytic triad is diagrammed in figure 4.

The four catalytic residues are joined together by 3 inter-connective cavities that define active site architecture and capacity: CoA binding tunnel, catalytic triad cavity and cyclization pocket (also called the product binding site as product refers to polyketide chains formed) (Austin & Noel, 2003; Jez et al., 1999).
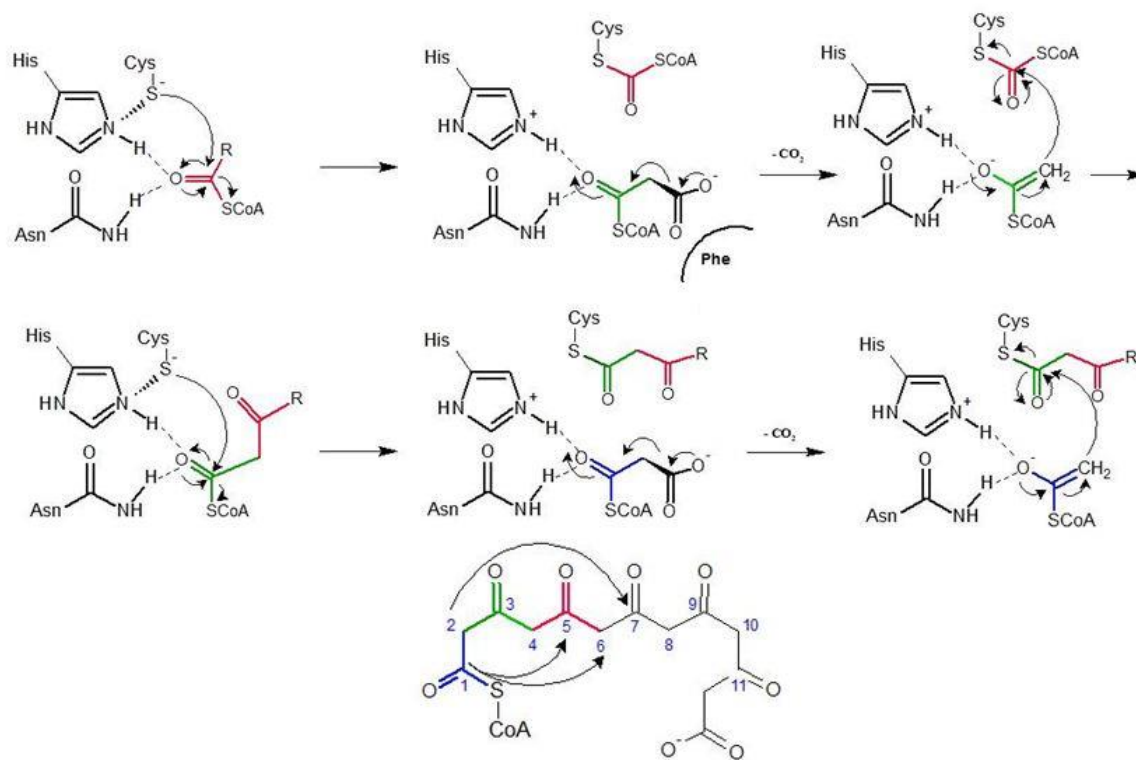
Figure (4) Proposed mechanism of PKSs type III catalytic triad

Initial loading of the starter acyl-CoA to Cys residue is followed by decarboxylation of the extender unit (malonyl-CoA). The elongation pattern of the growing polyketide chain is indicated by the colored carbonyl. The cyclization patterns of the polyketide chain are indicated by the arrows and their corresponding carbon numbering.

The functional diversity between members of PKSs type III family arises from some alterations in residues lining the active site cavity, cyclization pocket and dimer interface. These alterations remarkably governs the volume of active cavity, the substrates selectivity, polyketide chain length and cyclization pattern (Abe & Morita, 2010).

6. Chalcone synthase (CHS) superfamily in plants:

In the plant kingdom, PKSs type III are the main replenisher for a diverse range of secondary metabolite flavonoids including chalcones, stilbenes, phloroglucinols,

resorcinols, benzophenones, biphenyls, chromones, pyrones, and curcuminoids (Abe & Morita, 2010; Austin & Noel, 2003). These metabolites are momentous for plant growth, repository of pigments, responses to environmental stressors such as UV irradiation, defensive mechanisms and symbiosis with microorganisms. Not to mention their beneficial property to humanity as plant-based medicines with anticancer, antioxidant, antiasthamtic, antibiotic activities, etc. (Ververidis et al., 2007). CHS is the first PKS type III early characterized in 1970s by Kreuzaler and Hahlbrock (1972) from parsley *Petroselinum hortense*. Since then, more than 900 PKSs type III genes have been deposited to NCBI database (Abe & Morita, 2010). CHS superfamily was believed to be exclusively ubiquitous in plant kingdom until its recent discovery in bacterial genomes (Moore & Hopke, 2001). The recent discovery raised questions towards the evolutionary history of this enzyme family among both kingdoms (Gross et al., 2006).

7. The PKSs type III in bacteria:

Despite the overall functional similarity, the recently discovered bacterial PKSs type III only share 25 – 50 % identity with each other and with the CHS superfamily in plants (Austin & Noel, 2003). More microbial genome mining revealed that these CHS-like enzymes are more diverse than in plants. Thus, promoting new series of features unknown for plant related enzymes. Also, these microbial PKSs type III produce unique polyketides with promising pharmaceutical applications that trigger interest to investigate these microbial enzymes extensively (Katsuyama & Ohnishi, 2012).

The discovery of the first bacterial PKS III was reported in 1995 by Horinouchi's group where a gene cluster encoding red-brown hexa-hydroxy-perylene quinone (HPQ)-melanin pigment was discovered and identified in *Streptomyces griseus* catalyzed by 1, 3, 6, 8-Tetra-hydroxy-naphthalene synthase (THNS), also known as RppA (red pigment production) (Ueda, Kim, Beppu, & Horinouchi, 1995). Two open reading frames (ORFs) were identified: ORF-1 producing a 109 amino acid RppA and ORF-2 producing 263 amino acids RppB. Heterologous expression of such ORFs conferred red-brown HPQ-melanin pigment formation (Katsuyama & Ohnishi, 2012;

Ueda et al., 1995). Melanin production is catalyzed by the simple homodimer THNS (RppA) from five malonyl CoA precursor units producing THN (Funa et al., 1999). RppA homologues are widely spread in other actinomycetes like *Saccharopolyspora erythrea* (Cortes, Haydock, Roberts, Bevitt, & Leadlay, 1990), *S.lividans* and *S.antibiotics* (Funa, Ohnishi, Ebizuka, & Horinouchi, 2002), *S.toxytricini* NRRL 15445 (Zeng, Decker, & Zhan, 2012) and *S.coelicolor* A3(2) (Izumikawa et al., 2003). Its role involves also the biosynthesis of important metabolites harboring a napthoquinone ring as naphterpin, a terpenoid antioxidative agent (Shin-ya, Furihata, Hayakawa, & Seto, 1990) and napyradiomycin with an antibacterial and antitumor activity (Snyder, Tang, & Gupta, 2009).

Since then, its wide distribution among bacterial genomes has been realized (Katsuyama & Ohnishi, 2012). For instance, Di-hydroxy-phenyl-glycine (DHPG), a synthase found in *Amycolatopsis* from *Actinobacteria* phylum is involved in biosynthesis of balhimycin, an antibiotic from the vancomycin class with powerful resistance to methicillin-resistant *Staphylococcus aureus* (MRSA) (Katsuyama, 2010). Germicidin, a inhibitory autoregulator in spore germination is biosynthesized by Germicidin synthase (Gcs) in *S.coelicolor* (Song et al., 2006;Yu et al., 2012). In *S.griseus*, Streptomyces resorcinol synthase (SrsA) synthesize phenolic lipids, alkyl resorcinols and alkyl pyrones that collectively localize and integrate in the cytoplasmic membrane due to their amphiphilic properties. These lipids providing rigidity to the host membrane where it can survive the presence of β-lactam antibiotics as penicillin G and cephalexin (Funabashi, Funa, & Horinouchi, 2008). Role of PKSs type III in phenolic lipid cell wall (mycolic acid) is also encountered in *Mycobacterium tuberculosis* where three putative PKS 10, PKS 11 and PKS 18 were involved (Cole et al., 1998; Gokhale, Saxena, Chopra, & Mohanty, 2007). Alkyl resorcinol synthase (ArsB and ArsC) present in *Azotobacter vinelandii* are also responsible for the biosynthesis of alkyl resorcinol in the cyst wall (Funa, Ozawa, Hirata, & Horinouchi, 2006). Also, a 2,4 di-acetyl-phloro-glucinol (2,4 DAPG) biosynthesis in *Pseudomonas flourescens* by phloroglucinol synthase (PhlD) is considered a leading biocontrol agent against soil borne fungal pathogens (Bangera &

Thomashow, 1999). It has a suppressive control of *Gaeumannomyces graminis*, a wheat fungal pathogen (Mavrodi et al., 2001). Collectively, microbial PKSs type III generates a broad range of natural products with important pharmaceutical and biological attractions that require further investigations.

As mentioned earlier, some features exist in microbial type III PKSs that are deprived or underrepresented in related CHS superfamily plants. For instance, THNS is capable of utilizing malonyl-CoA as a starter unit not as an extender one as in plants (Katsuyama, 2010). Gcs confers a scrupulous selectivity to ACP-binding starter units unlike plant's selective CoA starter units (Song et al., 2006). Also, incorporating methyl-/ethyl-malonyl-CoA substrates is an unparallel theme for CHS superfamily in plants but seems to be a normal feature in Gcs and SrsA. This points out the unique catalytic capabilities of these enzymes (Funabashi et al., 2008; Song et al., 2006). These unusual features require further investigations for such extremely diverse class of microbial PKSs type III to fully understand their exact role within bacterial community.

8. Study objective:

The aim of this study is to Screen and Isolate a possible bacterial PKS type III from Atlantis II deep brine pool using a metagenomic approach. It was expected that bacterial PKSs type III would be present in ATII due to the abundance of aromatic compounds. This predicts the existence of diverse aromatic degrading pathways that produces CoA tethered metabolic derivatives via CoA ligation (Wang et al., 2011). The CoA tethered metabolic derivatives could possibly be the substrates for bacterial PKSs type III. Isolating possible bacterial PKSs III from such extreme environment would reveal the potential of ATII bacterial community to produce natural polyketides. Also, phylogenetic analysis was done to gain Deeper insights into the evolutionary origin of PKS type III among Prokaryotes and Eukaryotes.

CHAPTER 2: MATERIALS AND METHODS

1. **Collection of brine pool water samples:**

   During the KAUST Red Sea Spring 2010 expedition, water samples were collected from the convective layers of Atlantis II deep and Discovery Deep. Sample collection was performed by utilizing Niskin bottles associated with Conductivity, Temperature and Depth (CTD) meter tool, for better assessment of the physical and chemical properties of the brines. Almost 100 liters of water samples from each layer of the brines were collected and subjected to serial filtration using different sized pores 3 µm, 0.8 µm and 0.1 µm mixed cellulose ester (nitrocellulose/cellulose acetate, Millipore, Ireland) filters, respectively. The filters were preserved in sucrose buffer until DNA extraction. This step was done by KAUST Red Sea Spring 2010 expedition members on the vessel.

2. **Construction of 454 metagenomic database:**

   Environmental DNA was extracted from 0.1 µm filter according to Rusch et al protocol with minor modification (Rusch et al., 2007). A single stranded (ss) DNA library was constructed by nebulization on emulsion beads, then was amplified by emulsion PCR (emPCR). Pyrosequencing of environmental DNA was done by Roche 454 GSFLX genome analyzer using GSFLX Titanium pyrosequencing kit. For assembled 454 metagenomic database, generated reads were collectively assembled to contigs by Newbler® GS assembler version 2.6 with the default options. Using MetaGeneAnnotator (Noguchi, Taniguchi, & Itoh, 2008), potential open reading frames (ORFs) were predicted for the assembled contigs, This step was performed at the Biology Department, American University in Cairo (AUC).

3. Computational screening of the LCL 454 metagenomic database of the ATII Brine pool:

Using the Hidden Markov model (HMM) from HMMER3 package, the translated protein sequences of LCL ATII database were screened for domains of the N and C terminal of the chalcone synthase. Pfam accessions were retrieved from pfam 27.0 (http://pfam.sanger.ac.uk/) (Punta et al., 2011). The resultant hits with highest similarity domains were selected and assembled to a contig "**Contig1**" by Phred, Phrap, Consed software for trimming, assembly and proper visualization, respectively (Ewing & Green, 1998; Ewing, Hillier, Wendl, & Green, 1998; Gordon, Abajian, & Green, 1998; Gordon, 2002). **"Contig1"** was aligned using the blast tool against the assembled 454 metagenomic database for enhanced annotation of the predicted ORF. The best hit identified was within a large **"Contig2"**, and was chosen for further investigations. The predicted ORF was given the name "**ATII-ChSyn**" and used for sequence similarity search against National Centre for Biotechnology Information (NCBI) non-redundant protein database by using the Basic Local Alignment Search Tool (BLASTX 2.2.28) with expect threshold (E value) set to 0.0001 and the similarity score was calculated by using substitution matrix BLOSUM45 (Altschul et al., 1997).

4. Functional annotation of ATII-ChSyn:

Complete ORF annotation was identified with the aid of Artemis annotation and visualization tool (Rutherford et al., 2000). The regulatory promoter region was predicted by using BPROM (http://linux1.softberry.com); the most accurate region was selected based upon the highest Liner discrimination function (LDF) score. The Shine delgarno "SD" region was predicted manually 5 bps upstream the ORF. The amino acid sequence corresponding to the identified ORF was predicted, and the corresponding molecular mass (in kDa) was hypothetically calculated by Compute pI/Mw tool from ExPASy resources online tool (Bjellqvist et al., 1993) (http://web.expasy.org/compute_pi/).

The Conserved Domains were identified using the conserved domain database (CDD) of the NCBI (Marchler-Bauer et al., 2011) and confirmed by pfam search. Cellular localization of the predicted ORF "ATII-CySyn" was identified by PSORT (Yu et al., 2010) (http://psort.hgc.jp/).

Multiple sequence alignment (MSA) was done with a dataset that includes representative sequences from plant and bacterial PKS III with ATII-ChSyn by Clustal Omega where the conserved domains were identified and aligned in all the submitted sequences (Goujon et al., 2010; Sievers et al., 2011) (http://www.ebi.ac.uk/Tools/msa/clustalo/)

5. Phylogenetic Analysis :

Assortment of bacterial PKS type III was achieved via phylogenetic analysis. A dataset including 85 bacterial, plant, fungi and amoeba PKS type III were acquired via NCBI non redundant protein database using BLASTP 2.2.28. The queries used were *S.griseus* RppA, *Medicago sativa* CHS and ATII-ChSyn. Amoeba sequences were retrieved by excluding all the bacterial PKS type III hits from BLASTP search of *S.griseus* RppA. Phylogenetic analysis was performed using phylogeny.fr web service (Dereeper et al., 2008) (http://www.phylogeny.fr/version2_cgi/index.cgi).

Sequence alignment was done by full mode run MUSCLE (Edgar, 2004a, 2004b). Alignment was curated manually using Jalview version 2.8 (Waterhouse, Procter, Martin, Clamp, & Barton, 2009). Phylogenetic relations were inferred by Maximum likelihood (ML) estimation approach using PhyML version 3.0 program (Guindon et al., 2010). WAG amino acid substitution model was selected with gamma distribution parameter estimated and the number of substitution rate categories was four. The constructed tree confidence was estimated by a bootstrap of 1000 pseudo-replicates. The tree was rooted with bacterial and plant FabH as outgroup. The constructed tree was displayed using the Interactive Tree Of Life (iTOL) version 2.1 online tool (I. Letunic & Bork, 2011; Ivica Letunic & Bork, 2007) (http://itol.embl.de/index.shtml)

6. Comparative homology modelling of ATII-ChSyn :

The prediction of three dimensional (3D) model for isolated ATII-ChSyn protein was generated by MODELLER version 9.12 (Eswar et al., 2007). The template used to generate the model was the crystal structure of chalcone synthase from *Medicago sativa* (PDB ID 1CML, Resolution 1.69Å). This template was retrieved from BLASTP search against PDB database. Comparative modelling predicted the most satisfying model of ATII-ChSyn with the template by structural alignment in regards of spatial restraints and molecular geometry. Structural superimposing and prediction of substrate's binding site in ATII-ChSyn were carried out by Discovery Studio® visualizer 3.5 (Accelrys Software Inc., San Diego, CA). Attempts for Quaternary structure modelling of ATII-ChSyn were achieved by using SWISS-MODEL workspace web tool (Bordoli et al., 2008).

7. Environmental representation of bacterial PKS type III in ATII and DD:

Assessment of the bacterial PKS III abundance in ATII and DD was done by scanning the layers of both brine pools. HMM search was performed against the metagenomic 454 libraries of each layer. The two pfam accessions used were PF00195 [Chalcone and stilbene synthases, N-terminal domain] and PF02797 [Chalcone and stilbene synthases, C-terminal domain]. The reads with highest similarity domains obtained from the two brine pools were used as queries for HMM scan against pfam database. This step (reciprocal HMM) was done by translating these retrieved reads to the 6 possible frames before HMM scan. The correct frame reads were used as queries for a BLASTP against the non-redundant NCBI database to estimate the relative abundance of PKS III among the two brines.

8. Isolation and identification of the putative PKS type III "ATII-ChSyn":

8.1 PCR based screening method:

LCL environmental DNA was screened by using primers designed on one read within the "ATII-ChSyn" ORF's middle region, including the conserved catalytic cysteine (starting from position 306bp to 562bp). After successful amplification, the full ORF was amplified by using a set of primers: F_Orf, R_Orf, R_before_stop and R_downstream1. Primers were designed using primer3web version 4.0.0 online tool (Koressaar & Remm, 2007; Untergasser et al., 2012) (http://bioinfo.ut.ee/primer3/). The sequences of the primers used throughout this study are listed in Table (2) and their position relative to the identified ORF diagrammed in Figure (5). The PCR conditions for all primers used were: initial denaturation at 94°C for 5 minutes, 40 cycles (denaturation at 95°C for 1 minute, annealing at 58°C for 1 minute and extension at 72°C for 1 minute) and a final extension step at 72°C for 7 minutes.
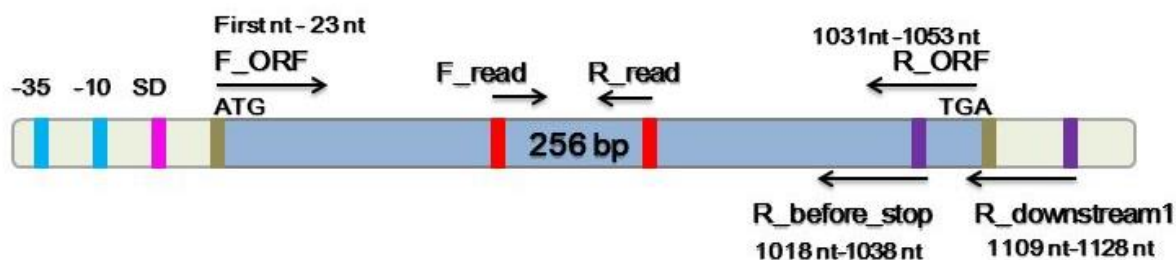
Figure (5) Schematic diagram showing the positions of the primers designed on putative ATII-ChSyn

Blue box represents the full ORF. Regulatory regions such as -35 /-10 and -SD are represented in turquoise and purple boxes, respectively. Brown boxes show the position of the start and stop of the ORF. Red boxes denote the beginning and the end of the middle read containing the catalytic site and amplified by the primers F_read and R_read (256 amplicon size). Purple boxes show the two R_downstream1 and R_before_stop reverse primers. Nucleotide positioning is shown on each primer arrow relative to ORF start codon.

Table (2) Primers used for screening the environmental DNA for partial or total amplification of "ATII-ChSyn" ORF

Location of primers designed	Primer name	Sequence
Within the middle of ORF (Start at position 306 bp till 562 bp). Amplicon size: 256 bp	F_read	gatgtcgtcgtcacggtttc
	R_read	cgacgatgttggcctttgtc
Exact ORF primers (Start and Stop). Amplicon size: 1053bp	F_Orf	ATGGCTCAAGCTACGAAGCTGAT
	R_Orf	TCATGCTGCTCTCTGGAGAGATA
F_Orf + downstream primer1 (36bp after stop codon) Amplicon size:1089bp	F_Orf	ATGGCTCAAGCTACGAAGCTGAT
	R_downstream1	CGATCGTAACCGCAGACAGA
F_Orf + Primer before stop (14 bp) Amplicon size:1038bp	F_Orf	ATGGCTCAAGCTACGAAGCTGAT
	R_before_stop	GAGAGATACACAGCTCACGCT

The table demonstrates the location of the primers relative to ATII-ChSyn ORF, their sequences and names

8.2 Cloning of the amplified "ATII-ChSyn":

Amplified fragments were purified and recovered from electrophoresis gel using the QiAquick® Gel Extraction Kit (Qiagen). Recovered DNA fragment concentration was assessed by Thermo Scientific NanoDrop™ 3300 Fluorospectrometer prior to the ligation reaction. Accordingly, the purified DNA fragment was ligated to p-GEM®-T easy vector (Promega) according to manufacturer's instructions. A volume of 3μl of the ligation reaction was transformed into electrocompetent *Escherichia coli* Top10 cells using the MicroPulser™ Electroporation Device (Bio-Rad) according to manufacturer's instructions. Transformants were plated on LB plates containing 0.1mg/ml ampicillin, 0.5mM IPTG and 40μg/ml X-Gal agar plates. A number of 5 to10 white colonies were carefully picked and colony PCR was done to confirm correct insert orientation. The later was done with the same primers used for amplification (F_Orf / R_downstream_1) and

thermocycler conditions as stated before. Colonies identified to contain the vectors with correct insert size and orientation, were grown overnight in LB-ampicillin at 37ºC. The plasmids were subsequently purified using the PureYield™ Plasmid Miniprep System (Promega) according to manufacturer instructions prior to sequencing.

8.3 Sequencing:

Sequencing was carried out in the Applied Biosystems 3730x DNA Analyzer. Several sequencing reactions were prepared using BigDye® Terminator v3.1 Cycle Sequencing Kit. Primers used for sequencing were F_Orf (T_a 58°C), F_read (T_a 58°C), and R_downstream2 (T_a 58°C), Sp6 promoter reverse primer (T_a 59°C) and T7 promoter forward primer (T_a 51°C). Each Sequencing reaction comprised big dye terminator, 5X Sequencing Buffer, primer (forward or reverse), purified cloned p-GEM-T vectors sample or uncloned amplified fragments (direct sequencing) and distilled sterilized water completed to the desired volume. Thermocycler conditions were initial denaturation 96°C for 2 minutes, 35 cycles (denaturation at 96°C for 45 seconds, annealing at either 58°C/59°C/51°C (according to primer used) for 30 seconds, extension 60°C for 4 minutes). Sequencing reaction was terminated by BigDye® X Terminator™ purification reagents with proper vortex then the samples were loaded to the Applied Biosystems 3730x Analyzer for detecting the sequences. The sequenced reads were assembled by CAP3 sequence assembly program to possible contigs (Huang & Madan, 1999) (http://pbil.univ-lyon1.fr/cap3.php).

8.4 Construction of recombinant ATII-ChSyn expression systems:

8.4.1 ATII-ChSyn / pET SUMO protein expression system (Invitrogen™):

To clone "ATII-ChSyn" into Champion™ pET SUMO expression plasmid (Invitrogen™) [N-terminal Histidine-tagged], the ATII-ChSyn was amplified from purified recombinant p-GEM-T® verified to contain the insert with correct sequence. The amplification was performed using F_Orf and R_downstream1.Thermocycling conditions were: an initial denaturation at 94ºC for 5 minutes, 40 cycles (denaturation at 95ºC for 1

minute, annealing at 58°C for 1 minute and extension at 72°C for 1 minute) and a final extension stage at 72°C for 7 minutes to ensure deoxyadenosine (A) – overhangs for TA cloning. The correct sized amplicon was gel purified using the QiAquick® Gel Extraction Kit (Qiagen) then ligated into Champion™ pET SUMO plasmid **"pET SUMO/ATII-ChSyn"** using the pET SUMO TA Cloning® reagents. The plasmid was transformed into electrocompetent *Escherichia coli* BL21 DE3 by the MicroPulser™ Electroporation Device (Bio-Rad) according to manufacturer's instructions (Transformation in *E.coli* Top10 was done with same procedure and kept as glycerol stocks). Transformants were plated on LB-agar plates containing kanamycin at the concentration of 50μg/ml. The transformants having the insert with the correct orientation were identified by colony PCR using Forward SUMO sequencing primer (AGATTCTTGTACGACGGTATTAG), R_Downstream1 (TGGTATTTCGTCTGGCGAGC). Thermocycling conditions were initial denaturation at 94°C for 5 minutes, 25 cycles of (denaturation at 95°C for 1 minute, annealing at 55°C for 1 minute and extension at 72°C for 1 minute) and a final extension stage at 72°C for 7 minutes. The sequence fidelity of pET SUMO/ATII-ChSyn was confirmed by nucleotide sequencing using the pET SUMO forward and reverse primers.

8.4.2 ATII-ChSyn / pET -28b+ protein expression system (Novagen):

A codon-optimized "ATII-ChSyn" was used for pET -28b+ expression system and was purchased from GenScript,Inc with OptimumGene™ Gene Design. For the C-terminal Histidine-tagged construct, the synthesized gene was designed between the restriction sites *Nco*I (CCATGG) and *Hind*III (AAGCTT). **pET -28b+/ATII-ChSyn** was directly transformed into *Escherichia coli* BL21 DE3 (Transformation in *E.coli* Top10 was done with same procedure and kept as glycerol stocks). Transformants were plated on LB-agar plates containing 50μg/ml kanamycin. Successful transformation was confirmed by colony PCR using the pET -28b+ vector primers: T7 promoter Forward Primer (TAATACGACTCACTATAGGG), T7 Terminator Reverse Primer (GCTAGTTATTGCTCAGCGG). Thermocycling conditions were: initial denaturation at 94°C for 5 minutes, 40 cycles of (denaturation at 95°C for 1 minute, annealing at 54°C for

1 minute and extension at 72°C for 1 minute) and a final extension stage at 72°C for 7 minutes. The sequence of pET -28b+/ATII-ChSyn was confirmed by nucleotide sequencing using the same vector primers mentioned above.

8.4.3 Overexpression of ATII-ChSyn:

The induction of "ATII-ChSyn" protein overexpression in both expression systems was attempted as follows: freshly prepared culture from an overnight growth of *E.coli* BL21 (DE3) harboring either of the expression systems (**pET SUMO/ATII-ChSyn** and **pET -28b+/ATII-ChSyn**) was cultured in LB broth containing 50μg/ml kanamycin at 37°C with shaking at 250 rpm to an optical density at 600 nm (OD$_{600}$) of (0.4 - 0.6). The different induction conditions tested for both expression systems are listed in Table (3). The settled induction was initiated by addition of 100μM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated for an additional one hour at 37°C.

The culture was harvested by centrifugation at 4,629 x g (6,000 rpm), 4°C. The pellet was thawed at 42°C and freezed at -80°C interchangeably for 4 successive times then resuspended in 1000μl of lysis buffer comprising 20mM NaH$_2$PO$_4$, 0.5M NaCl, 20mM imidazole (pH 7.4), 1mM phenylmethylsulfonyl fluoride (PMSF) and 1mg/ml Lysozyme. After 30 minutes of ice incubation, the cells were disrupted with sonication on ice by the SONIFIER$^®$ 150, Branson (15 seconds sonication with 15 seconds pause in between for 6 times). The crude cell lysate was separated from the cell debris by centrifugation for 25 minutes, 15,557 x g (11,000 rpm) at 4°C. The supernatant and pellet were analyzed by 12% sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE). SDS- PAGE gel preparation and buffers used was carried out according to Laemmli's standard protocol (Laemmli, 1970) .

Table (3) Induction Conditions used in both expression systems pET SUMO / pET28b+

Expression Vector	Conditions Tested			
	IPTG	Time	Temp.	Additives to lysis buffer
pET SUMO	1mM	1:30 hr	37°C	None
		3:00 hr		
		5:00 hr		
pET SUMO	0.5mM	1:30 hr	37°C	None
		3:00 hr		
		24 hrs	25°C	
pET SUMO	0.2mM	1:30 hr		None
		3:00 hr	37°C	
		24 hrs	25°C	
pET SUMO	0.1mM	1 hr	37°C	None
pET 28b+	1mM	1hr to 5hr *	37°C	None
pET 28b+	0.5mM	15 min	37°C	None
		30min	15°C	
		1hr		
		1hr to 5hr *	37°C	
pET 28b+	0.2mM	1hr to 5hr *	37°C	None
pET 28b+	0.1mM	15 min	15°C	
		30 min	37°C	
		1 hr to 5hr *		w/o 10%glycerol – 2% Tween 20 -2% Triton X-100
		4hr	15°C	
		6hr		
		10hr		
		12hr		
		13 hr		
		16hr		

		18hr		
		21hr		
pET 28b+	0.05mM	4hr	37°C	w/o 10%glycerol – 2% Tween 20 -2%
		6hr	15°C	Triton X-100
		10hr		
		12hr		
		16hr		
		18hr		
		21hr		
pET 28b+	0.025mM	12hr	37°C	w/o 10%glycerol – 2% Tween 20 -2%
		16hr	15°C	Triton X-100
		18hr		
		21hr		
pET 28b+	0.1mM	1hr	37°C	8 M Urea treatment with agitation (1hr at RT – overnight treatment at RT)

*1hr to 5hr: Induced samples were taken every 1 hour interval till 5 hours. w/o: means the induction samples were treated twice with and without those additives, each additive was added separately. RT: treatment was done at room temperature. All conditions resulted in insoluble protein.

8.5 Protein purification:

Purification was done by affinity chromatography using the Nickel nitrilotriacetic acid (Ni-NTA) purification system (Invitrogen™) under denaturing conditions to purify the His$_6$-tagged protein trapped in the cell debris. Cell pellets were treated with 8 ml Guanidinium lysis buffer (6M Guanidinum Hydrochloride, 20mM Sodium Phosphate, pH 7.8 and 500 mM NaCl). For improved cell lysis, the cell pellets were rocked gently with Guanidinium lysis buffer for 30 minutes at 4°C. Subsequently, the cell lysate was purified under denaturing conditions using 8M urea in all buffers as a protein denaturant and no imidazole was used. All the purification steps were performed at 4°C. The His$_6$-tagged protein was purified in 10 elution fractions and checked by 12% SDS-PAGE. The purified protein concentration was determined by Pierce® BCA Protein Assay (Thermo Scientific). The pooled sample of high concentration was determined

with reference to a standard curve of common protein bovine albumin serum (BSA) absorbance. Absorbance was measured at (A_{595}) 595nm where the pooled sample concentration was calculated from the standard curve. The purified pooled "ATII-ChSyn" protein was of approximate concentration 250μg/ml, (Figure 6). Measurements were carried out with three replicates utilizing the FLUOstar OPTIMA microplate reader (BMG LABTECH).
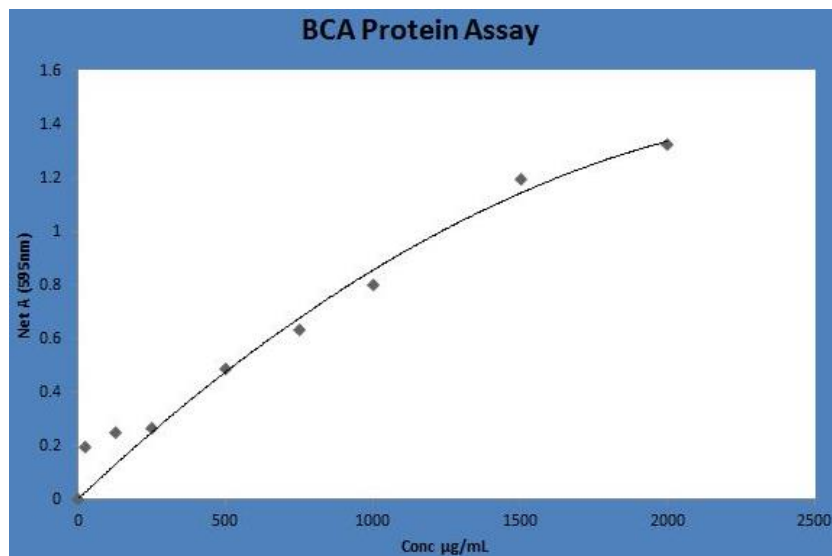
Figure (6) Pierce® BCA Protein Assay standard curve of BSA to determine protein concentration

After plotting the standard curve of BSA, pooled purified "ATII-ChSyn" sample absorbance was plotted. Its concentration was approximately 250μg/ml.

CHAPTER 3: RESULTS AND DISCUSSION

1. **Computational screening of the LCL 454 metagenomic database for PKSs type III:**

By using the keywords "Chalcone Synthase" and "PKS III", two accessions were retrieved from pfam database: PF00195 [Chalcone and stilbene synthases, N-terminal domain] and PF02797 [Chalcone and stilbene synthases, C terminal domain]. Screening of the LCL 454 metagenomic database by using these accessions returned adequate number of reads with high similarity domains. Eighty one and seventy six reads were for N-terminal and C-terminal domains of chalcone synthase, respectively. An assembled contig **"Contig1"** of size 1408bp was generated including 105 assembled reads only. This contig was blasted against the 454 assembled ORFs metagenomic database for enhanced annotation of the predicted ORF. The best hit identified was a large assembled contig **"Contig2"** of size 84,461 bp with 83 possible ORFs. The candidate ORF "ATII-ChSyn" of 1,053 bp size was the best ORF aligned with Contig1, its position was located between 61,132 and 62, 184 within Contig2. It existed as a sole gene not included in an operon or gene cluster.

Protein similarity search against NCBI non-redundant protein database was achieved by using the alignment tool (BLASTX 2.2.28) (Altschul et al., 1997). The parameters were set to use the more permissive BLOSUM45 matrix to retrieve more divergent sequences but with stringent E-value threshold for improved significance. The highest similarity hit was "chalcone & stilbene like synthase domain" protein from *Rhizobium sp.* PDO1-076 (Accession: WP_009109596.1) of Expect (E) value 0.0 with maximum identity 76%, similarity up to 87% and covering 99% of the query length, (Figure 7). Chalcone synthase of *Rhizobium etli* CFN42 (RePKS) (Accession: YP_468285.1) was the highest similar biochemically characterized protein (Jeya et al., 2012) with maximum identity of 61%, similarity up to 77% and covering of 99% of query length. These results might pinpoint some novel aspects of the sequence of the predicted ORF ATII-ChSyn. The significance of the predicted ORF was confident due to

its similarity with several bacterial PKSs type III. Most hits were from the phylum *Proteobacteria*, class *Alphaproteobacteria* showing several order representations: *Rhizobiales* (76% identity), *Rhodospirillales* (72% identity), *Rhodobacterales* (68% identity), *Sphingomonadales* (53% identity) and *Caulobacterales* (45% identity). On the other hand, the *Rhizobiales* (order of the best hits) are found in diverse soil environments as symbionts with plant's nodules for nitrogen fixation (Djordjevic, Gabriel, & Rolfe, 1987). Generally, the phylum *Proteobacteria* has been reported to be present within layers of Atlantis II Brine pool (Qian et al., 2011).


```
chalcone synthase [Rhizobium sp. PDO1-076]
Sequence ID: ref|WP_009109596.1| Length: 350 Number of Matches: 1
▷ See 1 more title(s)

Range 1: 1 to 349 GenPept  Graphics          ▼ Next Match  ▲ Previous Match
Score            Expect Method              Identities    Positives    Gaps        Frame
561 bits(1900)   0.0    Compositional matrix adjust. 265/349(76%) 304/349(87%) 0/349(0%)  +1

Query  1    MAQATKLISIGTAVPSHMIEQRDAAAVAHRAFSDRFKDFDRLAKVFESSGIRRRYAIRPL  180
            M +A  LISIGTAVP+++IEQRDAA  AH+AF+ RF DFDRLAKVFESSGIRRRYA+RP+
Sbjct  1    MNEAVHLISIGTAVPANIIEQRDAARTAHKAFASRFGDFDRLAKVFESSGIRRRYAVRPI  60

Query  181  EWYLEPLGWPERNAAYIEGACSLFVSAAQQALEKADLRGEdvdvvvtvsstGIATPSLEA  360
            +WYL PL WP RNAAYI+GAC LFV AA  ALE A L G  VD V+TVSSTG+ATPSLEA
Sbjct  61   DWYLTPLDWPARNAAYIDGACDLFVVAAANALEIAGLDGSKVDTVITVSSTGVATPSLEA  120

Query  361  RVAHQMGFREDIERVPVFGLGCAGGVSGFSIASRFAAARPGKIVLLVAVELCTLAFRLDQ  540
            RVA ++GFR DIERVPVFGLGCAGGVSG+SIA R AA+RPG IVL+VAVE CTL+FR+DQ
Sbjct  121  RVAKRLGFRADIERVPVFGLGCAGGVSGLSIAGRMAASRPGSIVLVVAVETCTLSFRMDQ  180

Query  541  LTKANIVATALFGDGAAACILKAGDAGIANIEMSGQHTWPDTLDIMGWSVDKQGFGVIFD  720
            +TKANI+ATALFGDGA+ACIL+   GIA +EMSGQH WPD+L+IMGWSVD +GFGVIFD
Sbjct  181  VTKANIIATALFGDGASACILRTDAPGIATVEMSGQHIWPDSLEIMGWSVDNEGFGVIFD  240

Query  721  RAIPPFAEERIAPAISGILNRADLAMADIDRFACHPGGTKVIAALETALSLGQGSLDHER  900
            RAIPPFAEE IA A+SGIL R+  ++ADIDR ACHPGG KVI ALE AL+L +G+L+HER
Sbjct  241  RAIPPFAEENIAAAVSGILARSGHSLADIDRVACHPGGAKVIDALENALTLSRGTLNHER  300

Query  901  DVLSDYGNMSAPTALFVLDRVVQAGLPSRTLLTALGPGFSVSCVSLQRA  1047
            DVL+DYGNMSAPTALFVLDR+ QAGLPSRTLLT+LGPGF VSCVSL+RA
Sbjct  301  DVLADYGNMSAPTALFVLDRLAQAGLPSRTLLTSLGPGFCVSCVSLKRA  349
```

Figure (7) Alignment of ATII-ChSyn with best hit

Alignment with the best similarity hit Chalcone & Stilbene like synthase domain protein from *Rhizobium sp. PDO1-076* showed maximum identity 76%, the highest among all

sequences results from BLASTX. Similarity is up to 87% and covering 99% of the query length.

2. Functional annotation of predicted "ATII-ChSyn" ORF :

Artemis annotation and visualization tool facilitated complete ORF annotation. The predicted ORF encodes for a 350 amino acid length protein of molecular weight 37.23 KDa. The regulatory promoter region was predicted by using BPROM online tool. Based upon the highest Liner discrimination function (LDF) score obtained [which equaled 1.97], the positions of the -10 [GGCTACGT] and -35 [TCGACA] regions were predicted. Two transcription factor (TF) binding sites were predicted as well; rpoD19 [TCCTGCTA] at position -67 and lrp [TTTAGTGT] at position -53 prior the start codon. The Shine Dalgarno (Ribosomal Binding site) region [AAGGAA] was predicted manually 5 bps upstream the ATG start codon owing to its purine-rich content as shown in Figure (8).

CAGCACGTTTTCTGGATACCTCGACAGGAACTTTCTTCCCTGAGGCTACGTTCCTGCTA
GGGAATTTTAGTGTTTGCGCACGCTGCTCAGGGGCATGCGATTGATATCAAGGAATGTC
atggctcaagctacgaagctgatctcgatcggaacagctgttccatcgcatatgatcgaa
 M A Q A T K L I S I G T A V P S H M I E
cagcgagatgcagccgccgtagcgcatcgcgcattttcggatcggttcaaggattttgat
 Q R D A A A V A H R A F S D R F K D F D
cgtttggcaaaggtgttcgagagctcgggcattcgtaggcgctatgcgatcaggcctctt
 R L A K V F E S S G I R R R Y A I R P L
gagtggtatcttgaaccgctgggctggccggaacgaaatgccgcctatatcgaaggggca
 E W Y L E P L G W P E R N A A Y I E G A
tgcagtttgttcgtcagtgcggctcagcaggcgctggagaaagcggaccttcgaggtgaa
 C S L F V S A A Q Q A L E K A D L R G E
gatgtcgatgtcgtcgtcacggtttcctcaacaggtatagccacgcctagtctcgaagcg
 D V D V V V T V S S T G I A T P S L E A
cgcgtcgctcatcaaatggggtttcgtgaagacatcgaaagggtgcctgtcttcggcctt
 R V A H Q M G F R E D I E R V P V F G L
ggctgtgccggcggtgtctcaggattctccattgcatctcgatttgcagcagcgcggccc
 G C A G G V S G F S I A S R F A A A R P
ggaaaaattgtcctgctggttgcagttgaactgtgcacgctcgcgttccgactggatcaa
 G K I V L L V A V E L C T L A F R L D Q
ttgacaaaggccaacatcgtcgcaactgcgctgttcggggatggggcggccgcctgcatt
 L T K A N I V A T A L F G D G A A A C I
ctcaaggcgggtgatgcaggtatcgcaaacatcgaaatgtccggccagcatacttggcct
 L K A G D A G I A N I E M S G Q H T W P
gatacgctcgacatcatgggttggagcgttgataagcaggggttcggcgttattttttgat
 D T L D I M G W S V D K Q G F G V I F D
cgagcaataccgccatttgctgaggaacgcattgcaccggctatatcgggtattttaaac
 R A I P P F A E R I A P A I S G I L N
cgtgctgatcttgccatggcagatatcgacaggtttgcctgtcatccaggtggtaccaag
 R A D L A M A D I D R F A C H P G G T K
gtgattgctgctctggagacagcactgtcgcttggtcaaggcagtttggaccatgagcgt
 V I A A L E T A L S L G Q G S L D H E R
gacgtgttgtccgattatggaaacatgtcggctccaaccgcacttttttgtattggaccgt
 D V L S D Y G N M S A P T A L F V L D R
gtcgtccaagccggcttaccgtcacgtaccctccttactgcactgggaccgggtttttagc
 V V Q A G L P S R T L L T A L G P G F S
gtgagctgtgtatctctccagagagcagcatga
 V S C V S L Q R A A -

🟩 -35   🟦 -10   ⬜ Transcription factors: rpoD19 and lrp   🟨 SD

Figure (8) Complete "ATII-ChSyn" protein sequence with its regulatory region

Predicted regulatory region (-35 / - 10 / Transcription factors / Shine Delgarno 'SD') are highlighted. The Red 'M' indicates the beginning of ORF.

Conserved domains within the ORF were identified by NCBI CDD search tool. The specific hit was chalcone synthase (CHS-like) domain (cd00831), also the presence of nonspecific hit Ketoacyl synthase (KAS-III) (cd00830) indicates the high similarity between the PKSs type III and fatty acid KASs type III. Also, N and C terminal chalcone synthase pfam domains were identified. Four conserved features were detected: 1.Active site: CHN catalytic triad (Cys_{142} His_{275} Asn_{308}), 2.Malonyl CoA binding site (also called the binding tunnel), 3.Product binding site (cyclization pocket) and 4.Dimer interface, (Figure 9). These findings reinforce that the predicted ATII-ChSyn is related to PKS III enzyme family.
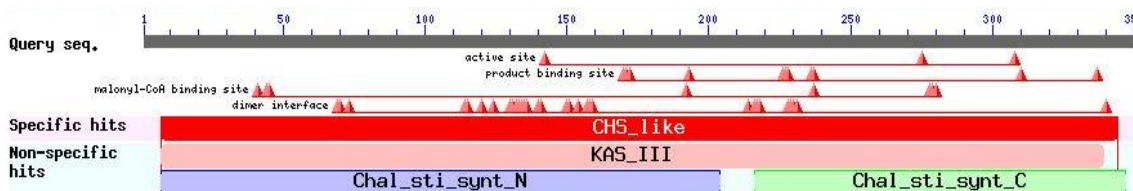
Figure (9) The Conserved Domains identified within the predicted ATII-ChSyn

NCBI CDD search tool detected CHS_like domain (cd00831) as the specific hit. Four features characteristic to the PKS type III enzymes were identified including the conserved catalytic triad CHN essential for activity.

The multiple sequence alignment of ATII-ChSyn with previously characterized bacterial and plant PKS type III representatives showed the conserved catalytic triad **CHN** as a characteristic signature among them, (Figure 10). The other three features escorting the active site were identified as well. Within the malonyl CoA binding site, the conserved phenylalanine [F-215 of *Medicago sativa* CHS (MCHS) or F-192 of ATII-ChSyn] known as the 'gate keeper' (Austin & Noel, 2003) and a conserved glycine [G-306 of MCHS or G-278 of ATIIChSyn] were detected. Generally glycine (**G**), as a non-bulk residue, retains the active site cavity among PKS type III enzymes (Katsuyama, Miyazono, Tanokura, Ohnishi, & Horinouchi, 2011). Product binding site (cyclization pocket) shows a conserved proline (**P**) [Pro-375 of MCHS or Pro-337 of ATII-ChSyn]

that plays an essential role in stabilizing the PKS III enzyme fold; thus contributes to the enzyme stability and activity (Austin & Noel, 2003). This **P** residue is located within a conserved **GFGPG** loop that controls the stereochemistry of cyclization within active site (Suh et al., 2000). In ATII-ChSyn, this loop was **<u>AL</u>GPG**. Replacement of glycine by alanine was previously observed among several bacterial sequences, (Figure 10). Aromatic phenylalanine was replaced by aliphatic leucine in ATII-ChSyn. Replacement of less bulk residues or vice versa requires more investigations to understand its impact on cyclization. Another important conserved residue glutamate (**E**) [E-192 of MCHS or E-170 of ATIIChSyn] was observed. It appears to form a network of hydrogen bonds with serine and threonine residues collectively with a water molecule nearby catalytic cysteine (Abe & Morita, 2010). This network also has an impact on cyclization modes done by the enzymes. The presence of all these essential conserved residues within ATII-ChSyn protein sequence supports it as a member of PKS III type enzymes.

Remarkable functional diversity among PKS type III enzymes is exhibited through substitutions of non-conserved residues. For instance, residues in dimer interface regions show certain diversity, (Figure 10). Also, the amino acid residues lining the active site cavity may differ from one enzyme to another, controlling the specificity of substrates and products towards each enzyme (Abe & Morita, 2010). Most of these residues are located within the product binding site, (Figure 10). For instance, in *Medicago sativa* CHS*,* the residues lining active cavity include T-132, S-133, T-194, T-197, G-256, F-265 and S-338 (Abe & Morita, 2010). The replacement of threonine 197 by less bulk alanine 175 in RePKS enlarges its active cavity. This could explains the uniqueness RePKS in condensing six malonyl-CoA to produce heptaketide pyrone (Jeya et al., 2012). Interestingly, RePKS is the characterized protein that shows highest sequence similarity to ATII-ChSyn. Thus, ATII-ChSyn could exhibit similar condensation reactions with malonyl-CoA. However, such observation can't be proven unless a complete characterization of ATII-ChSyn is employed.

```
                                                133
PCS        -KKFDHICKK--  -EPSLNDRQDI--  --SCGVDMPSADFQCAKLLGLHANVNKY
STS        -DKFKRICER--  -VPSLDARQAM--  --TTTPDLPGADFEVAKLLGLHPSVKRV
2PS        -EKFKRICEK--  -APSLNARQDL--  --TAGVDMPGADYQLVKLLGLSPSVKRY
MCHS       -EKFQRMCDK--  -APSLDARQDM--  --TSGVDMPGADYQLTKLLGLRPYVKRY
GCHS       -EKFQRMCDK--  -APSLDARQDL--  --TSGVDMPGADYQLTKLLGLRPSVKRL
RppA       -DLVLRLIQN--  -HPGFEVRNQV--  --CTGFMMPSLTAWIINSMGFRPETRQL
PhlD       -ALAKRMIQN--  -HTGFTHRSIV--  --CTGFMMPSLTAHLINDLGLRTSTVQL
PKS18      -ERIPRVYQK--  -PATIRDRMHL--  --STGFIAPGVDVAIVKELGLSPSISRV
PKS10      -DIVRQLHAS--  -LTDFGEANKI--  --VTGLAVPSLDARIAGRLGLRADVRRV
PKS11      -EIIRRLHAA--  -LTDFGDANEI--  --VTGVAVPSLDARIAGRLGLRPDVRRM
ATIIChSyn  -DRLAKVFES--  -PLGWPERNAA--  --STGIATPSLEARVAHQMGFREDIERV
2RePKS     -RHLARVFDS--  -PHGWQDRMEA--  --STGFTTPSLDAQMARRMGFRADIERV
              ::             .                *.       :*:         :

                               194 197                       256
PCS        CIYMQGCY-  -MRYAKDLAENNR-  -AELTIMM-  -LEGD-  -CTKQTV--  VIHL
STS        GVFQHGCF-  -LRMAKDLAENNR-  -SETTAVT-  -LEGD-  -WTAQTV--  AIGG
2PS        MLYQQGCA-  -LRLAKDLAENNK-  -SEITAIL-  -LEGD-  -STDQTI--  AMKL
MCHS       MMYQQGCF-  -LRLAKDLAENNK-  -SEVTAVT-  -LEGD-  -WTAQTI--  AIDG
GCHS       MMYQQGCF-  -LRLAKDLAENNK-  -SEITAVT-  -LEGD-  -SAAQTI--  AIDG
RppA       PIAQLGCA-  -INRAHDFCVAYP-  -CEFCSLC-  -LEGD-  -RNGSHL--  WISY
PhlD       PIAQLGCV-  -INRANDFASRAP-  -LEFSSLC-  -LEGD-  -KTGSYF--  YIKY
PKS18      VVNFMGCA-  -LGTATNYVRAHP-  -IELCSVN-  -LEGD-  -SSFSQL--  GIVL
PKS10      PLFGLGCV-  -VARLHDYLRGAP-  -VELCSLT-  -LFAD-  -DSRSHL--  TMGY
PKS11      PLFGLGCV-  -VARLRDYLRGAP-  -VELCSLT-  -LEGD-  -DSRSSL--  IMGW
ATIIChSyn  PVFGLGCA-  -FSIASRFAAARP-  -VELCTLA-  -LEGD-  -MSGQHT--  IMGW
RePKS      PVFGLGCA-  -FAIASRLARSRP-  -IELCTLA-  -LEGD-  -STGEHL--  IMGW
              :          .               *          . .*            :

           265                                            338
PCS        HLRETGMMFYLSK--  --IPHPGGRAILDQVEAKLKLRPEK--  -DYGNMVS--
STS        KVREVGLTFQLKG--  --VVHPGGRAILDRVEAKLNLDPTK--  -EYGNMSS--
2PS        HLREGGLTFQLHR--  --MVHPGGRAILDQVERKLNLKEDK--  -EYGNLIS--
MCHS       HLVEAGLTFHLLK--  --IAHPGGPAILDQVEEKLGLKPEK--  -EYGNMSS--
GCHS       HLKEVGLTFHLLK--  --IAHPGGPAILDQVEFKLGLREEK--  -EYGNMSS--
RppA       AVRDTGFHFQLDK--  --IVHAGGPRILDDLCHFLDLPPEM--  -ERGNIAS--
PhlD       DVKDSGFHFTLDK--  --IFHTGGRKILDELVLQLDLEPGR--  -EAGNIAT--
PKS18      GVNHNGITCELSE--  --AIHPGGPKIIEQSVRSLGISAEL--  -RFGNMLS--
PKS10      DVGSAGFELVLSR--  --VTHPGGPKIINAITETLDLSPQA--  -EIGNLSS--
PKS11      DVGSHGLRLRLSP--  --VSHPGGPKVIDAVATSLALPPEA--  -EIGNLSS--
ATIIChSyn  SVDKQGFGVIFDR--  --ACHPGGTKVIAALETALSLGQGS--  -DYGNMSA--
RePKS      KIDDGGFGIVLAQ--  --ICHPGGTKVLAAMESALSLTPGA--  -DYGNMSS—
              :       *:    :      *   **     ::          ** :

PCS        --LLGFGPGITV--
STS        --LFGFGPGLTI--
2PS        --LFGFGPGMTV--
MCHS       --LFGFGPGLTI--
GCHS       --LFGFGPGLTV--
RppA       --IAGFGPGITA--
PhlD       --LAAFGPGFTA--
PKS18      --AFAFGPGVTV--
PKS10      --MIAMGPGFCS--
PKS11      --MLAMGPGFCT--
ATIIChSyn  --LTALGPGFSV--
RePKS      --MIAMGPGFSA--
              .:*** .
```

39

Figure (10) Multiple sequence alignment of various PKSIII with "ATII-ChSyn"

ATII-ChSyn was aligned with RePKS from *Rhizobium etli*; RppA from *Streptomyces griseus*; PKS18 from *Mycobacterium tuberculosis*; MCHS from *Medicago sativa*; stilbene synthases-STS from *Pinus sylvestris* ; GCHS from *Gerbera hybrid*; 2-pyrone synthase (2PS) from *Gerbera hybrid*; pentakeide chromone synthase (PCS) from *Aloe Arborescens;* PKS10 from *Mycobacterium tuberculosis*; PKS11 from *Mycobacterium tuberculosis;* PhID from *Pseudomonas flourescens.* Conserved domain features: Malonyl-CoA binding site residues labelled in blue, dimer interface residues labelled in grey, Product binding site residues labeled in yellow and conserved catalytic triad labeled black (*). Underlined red residues are the amino acids lining active site cavity, numbering above is for MCHS. Conserved F and G in malonyl-COA binding site are labelled (*). Conserved E and P in product binding site are labelled (*). GFGPG cyclization loop top lined.

3. Phylogenetic analysis of bacterial PKSs type III

A comprehensive phylogenetic analysis was done in order to classify ATII-ChSyn protein among bacterial PKS III and to gain further insights about the evolutionary relations between bacterial, fungal and plants PKS III. The dataset used included most bacterial and fungal sequences available in NCBI database with representatives from plants and ameoba. The phylogentic tree was constructed by maximum likelihood estimation for better robustness. As shown in Figure (11), the tree topology is classified into an outgroup clade of FabH (KAS type III) sequences, several clades confined to bacterial sequences, single clades for ameoba, fungi and plants sequences, respectively.

Bacterial sequences are distrbuted among four clades, where clade A is related to actinobactrial DpgA synthases that plays a role in biosynthesis of the antibiotic vancomycin (Katsuyama & Ohnishi, 2012). Clade B is related to actinobacterial RppA synthases. Clade C is subdivided to two clades, where one is related mostly to *Mycobacteria* spp. while the other is divided into bacilli, α- proteobacteria and *Deinococcus*. ATII-ChSyn enzyme is located in a sub-branch within clade C clusterd with α- proteobacteria *Rhizobium sp*. PDO1−076 CHS (best hit), *Rhizobium etli* CFN 42 CHS (best characterized hit) and *Rhodobacter sphaeroides* 2.4.1 CHS. Clade D includes

miscellaneous strians from several phyla as actinobacteria, ß- / δ- proteobacteria, cyanobacteria, *Deinococcus thermus*, planctomycetes and bacteroidetes. Overall, the four bacterial clades indicate the diversity among bacterial PKS III, in contrast to the homogenous characeristic clades of eukaryotes PKS III placed at the top of the tree. Concerning the retreived plant sequences, they are assigned to different PKS III families as chalcone synthase CHS, stilbene synthase STS, benzophenone synthases BPS, benzolacetine synthases BAS, acridone synthases ACS and resverstrol synthases RS but are still very closely related.

FabH is a KAS type III enzymes invloved in fatty acid metabolism. It is considered highly similar to PKS III enzymes owing to their high resemblence in strucutre and mechanism (Austin & Noel, 2003). In this analysis, The FabH sequences included bacteria, archaea and plant FabHs. Its clade distrubution in the tree as an outgroup indicates that they are the common ancestors to the highly diverse PKS type III enzymes. Their position in close distance with bacterial clades also suggests that PKS type III was orignially evoloved in bacteria long time before the plants.

The evolutonary origin of PKS type III enzymes among plants and bacteria is yet considered vague (Gross et al., 2006). Two hypothesis are claimed in regards of their evolutionary origin. The first hypothesis suggests that PKS type III enzymes was recently acquired to bacteria via horizontal gene transfer (HGT) events from plants. This was based upon the recent discovery of only few bacterial PKS type III in a striking contrast to its high abundance in plant kingdom (Austin & Noel, 2003). The second hypothesis suggests that higher plants acquired PKS type III via HGT events from ancient eubacteria where it was then lost during prokaryotic evolution (Gross et al., 2006; Moore & Hopke, 2001). But with the augmented number of bacterial genome sequenced, these hypothses were seriously challenged (Gross et al., 2006). In this phylogentic analysis, it was found that: first, plants PKS type III enzymes are evolutionary distant from the common ansector FabH outgroup, (Figure 11). This indicates that bacterial PKS

type III has higher sequences similarity than the plant ones, thus disputing the hypothesis of HGT events from plants. Second, plant symbiotic bacterial PKS type III like the *Rhizobium* species in clade C are distant from plants, which indicates that they are not involved in HGT events to plants. Third, although clade D comprises some sequences related to cyanobacterial phyla (*Synechococcus sp.* WH 8102 CHS and *Prochlorococcus marinus str.* MIT 9313 CHS), they are not the key players for HGT events to plants. *Parachlamydia acanthamoebae* UV−7, a bacterium clustered with ameoba sequences *Acanthamoeba castellanii str.* Neff and *Dictyostelium discoideum* AX4 CHS could be the main key player in this riddle. Interestingly, *Parachlamydia acanthamoebae* is associated with the free living *Acanthamoebae spp.* and a caustive agent for lower respiratory tract pneumonia (Fukumoto et al., 2010). The protein involved in polyketide III synthesis has not been well studied yet but was identified through the pan-genome sequencing of *Chlamydia* families (Collingro et al., 2011).

The free living *Acanthamoebae spp.* is considered a wide range host for bacteria and viruses where possible evolutionary events could occur. In a recent study, the whole genome sequencing of *Acanthamoeba castellanii str.* Neff revealed that HGT has great impact on its genes (Clarke et al., 2013). This could support the hypothesis of HGT from *Parachlamydia acanthamoebae* to *Acanthamoeba castellanii str.* Neff in aquiring PKS type III. Both are clustered with *Dictyostelium discoideum* AX4 CHS domain that has been previously characterized (Austin et al., 2006; Ghosh et al., 2008). This phylogentic analysis disagree with Gross *et al* study that predicted the role of cyanobacteria species in HGT events to plants (Gross et al., 2006). These findings are yet considered preliminay where further investigation is required for better insight to the PKS type III evolution history.

Across the tree, the general distrubution of proteobacteria and actinobacteria phyla indicates some evolutionary events in bacterial PKS type III enzymes. Clade A and B that comprise actinobacterial sequences are located near the FabH outgroup. This could

suggest that actinobacteria were the first to acquire PKS type III in bacteria. Within clade B, *Sorangium cellulosum* SCcoCHS1 and *Pseudomonas fluorescens* PhlD from proteobacteria were clustered with actinobacteria. It predicts the HGT events from actinobacteria to protoebacteria. The same observation was noticed within the sub-branch of clade C, where proteobacteria *Magnetospirillum magnetotacticum* MS−1 CHS and *Myxococcus xanthus* CHS were also clustred with *Mycobacterium,* an actinobacterial strain.
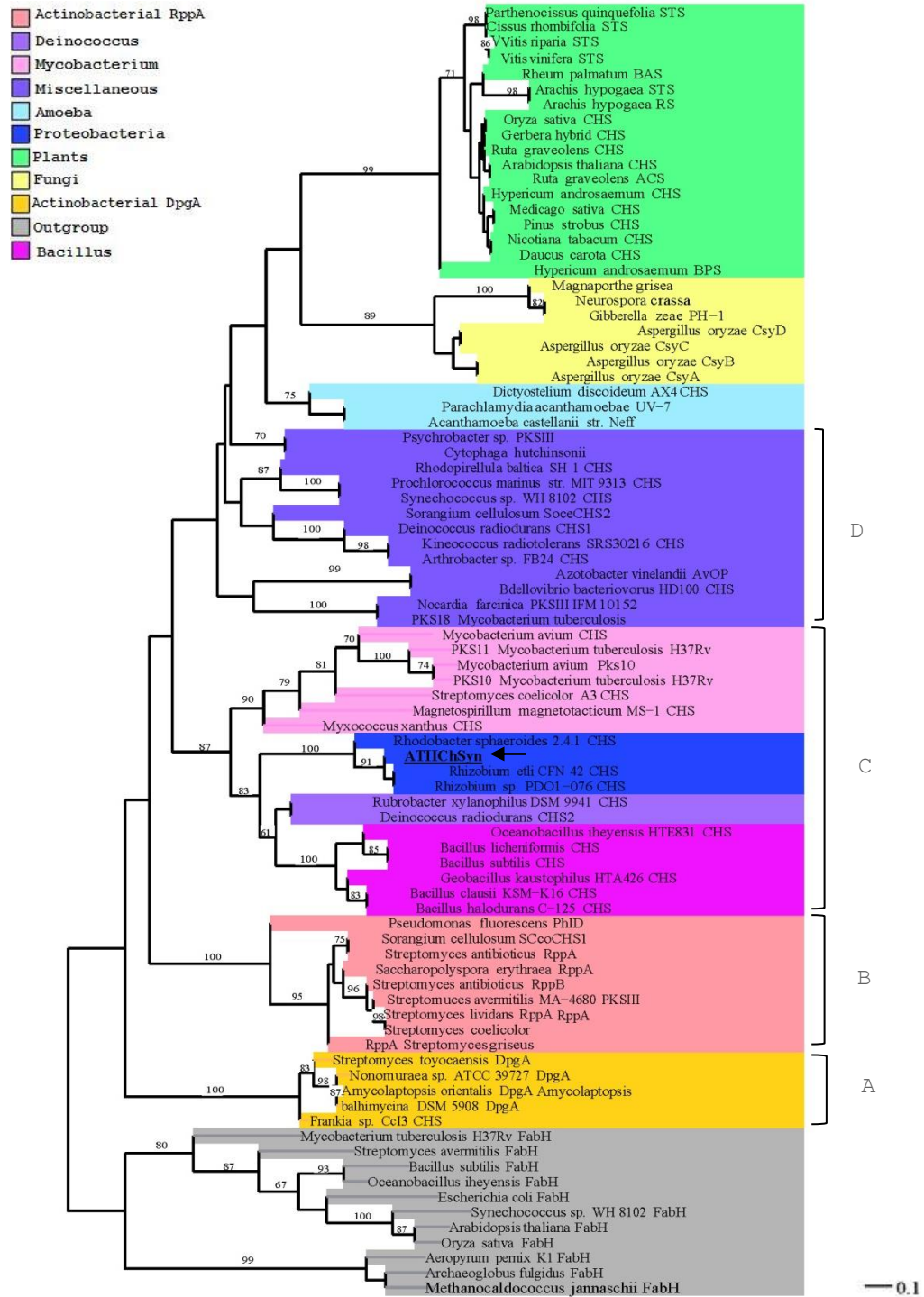
Figure (11) Phylogenetic tree for bacterial PKS III

Sequences from bacteria, amoeba, fungi and plants were used to construct the tree. The tree was inferred by Maximum likelihood estimation using (PhyML) after MUSCLE alignment. The tree was rooted with FabH sequences from bacteria, archea and plants as outgroup. Branch numbering indicate bootstrap values, above 60% values are only shown. Each clade is given a certain color that corresponds to the color label box.

4. **Comparative homology modelling of ATII-ChSyn:**

A three dimensional model structure of ATII-ChSyn was generated to evaluate the structural basis and predict the possibility of substrate interaction within its active cavity. This was done by selecting the crystal structure *Medicago sativa* CHS complexed with malonyl CoA as the template (PDB ID 1CML, Resolution 1.69Å). It was retrieved from BLASTP search against PDB database using ATII-ChSyn as query. This template showed a sequence similarity of 43%, sequence identity of 24% and a coverage of 98% of ATII-ChSyn protein length. Although there were other proteins showing higher sequence similarity and identity, none were co-crystalized with malonyl CoA, so they were discarded. The three dimensional model of ATII-ChSyn was built using MODELLER 9.12 by structural alignment where five models were generated. The best model was selected based upon satisfying molecular probability density function (molpdf). Simply, it calculates the probability spatial restraints of the built model in comparison to the template's. Also satisfying Discrete Optimized Protein Energy (DOPE) was considered. DOPE is the pairwise atomistic statistical potential that measures the energy of the predicted models according to the number of iterations in each generated model. Higher iteration numbers in one model ensures its reliability. Finally, a GA341 score (1.0) which indicates the reliability of the model was calculated. Usually, a model is considered reliable when its score is above a pre-specified cutoff (0.7). This indicates that at least 30% of C-alpha atoms of the model superimpose correctly and about 95% probability to perform a correct fold (Melo, Sánchez, & Sali, 2002).

Superimposition was visualized by Discovery Studio® visualizer 3.5, where ATII-ChSyn was well fitted with the template and malonyl CoA substrate, (Figure 12). The dimer interface, malonyl CoA binding site were detected. The prediction of the candidate active site interacting with malonyl CoA revealed the catalytic triad Cys 142, His 275, Asn 308 and the Phe 192 involvement. The residues were perfectly superimposed with the corresponding catalytic triad in *Medicago sativa* CHS, (Figure 13). This structural superimposition and substrate interaction strongly support ATII-ChSyn as a PKS type III enzyme.

Attempts were made to detect the quaternary structure of ATII-ChSyn as by SWISS-MODEL workspace web tool. Classically, PKS type III have been reported to exist as homodimer (Katsuyama & Ohnishi, 2012). In an automated mode without any assigned template, the predicted ATII-ChSyn model was generated based on PKS 11 of *Mycobacterium tuberculosis* (PDB ID 4JAP Resolution 1.83Å). This template showed the highest similarity among all PKS III crystal structures identified to date. Its sequence identity with ATII-ChSyn was up to 33% while the similarity was 52%. However, no quaternary structure was predicted. This could be attributed to the relatively low sequence identity between ATII-ChSyn and the template. Usually for quaternary structure prediction, sequence identity should be at least 50% which was not achievable in this case (Arnold, Bordoli, Kopp, & Schwede, 2006).

A.
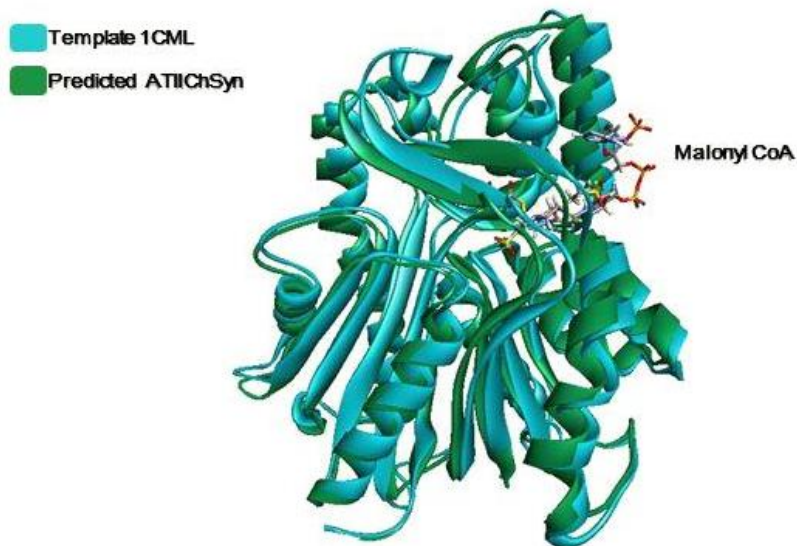
B.
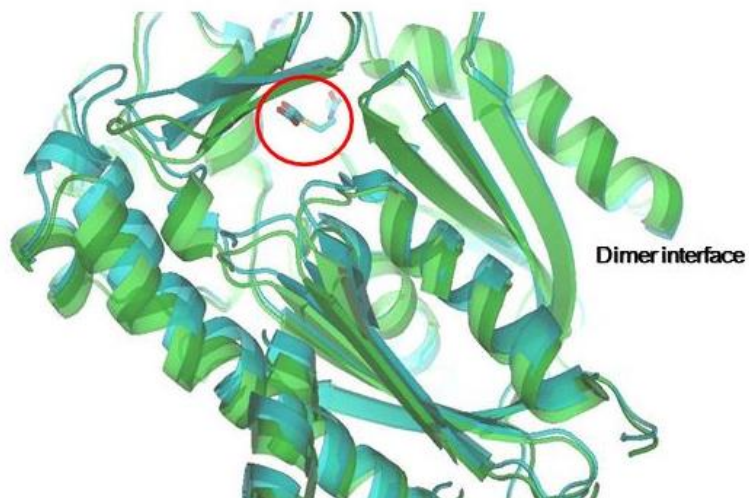
Figure (12) Structural superimposition of ATII-ChSyn with *Medicago sativa* CHS

A. The overall superimposition of predicted ATIIChSyn model with the template *Medicago sativa* CHS co-crystalized with malonyl CoA. B. A closer image showing the antiparallel β-sheets in the dimer interface. The red circle labels the binding site for malonyl CoA.
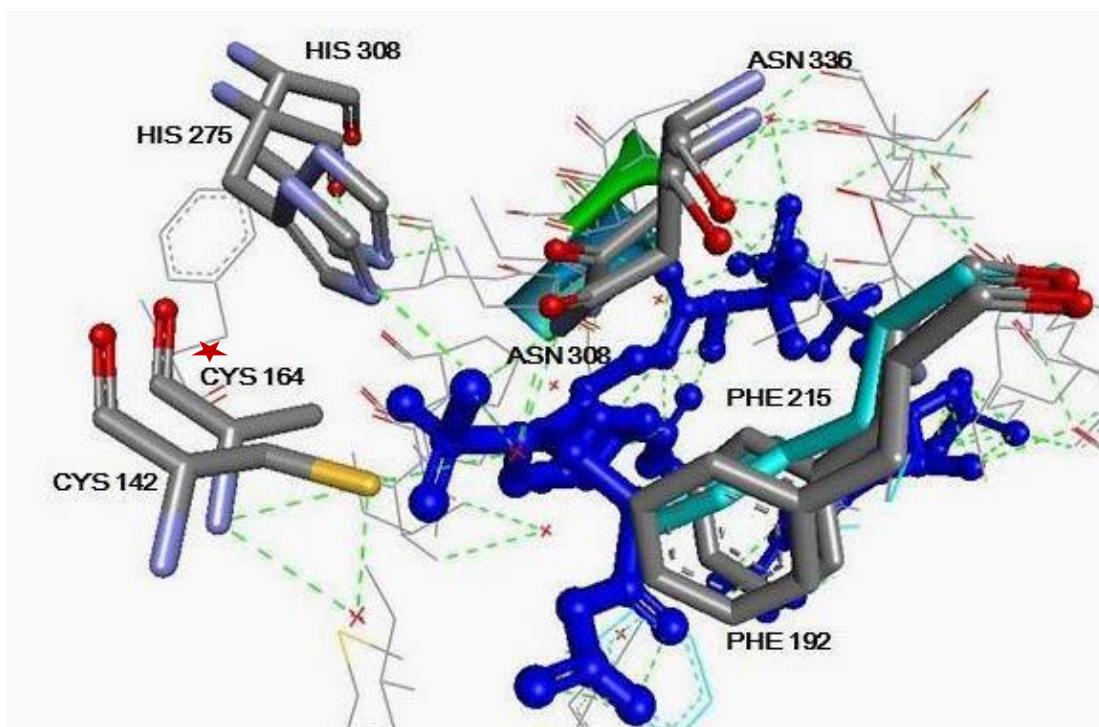
Figure (13) Predicted binding sites of ATII-ChSyn with malonyl CoA substrate

Superimposition of ATII-ChSyn with *Medicago sativa* CHS template shows the interaction of the catalytic triad of both proteins with the malonyl CoA (shown in as blue ball and stick model). The green dots represent hydrogen bonds interacting with malonyl CoA. The residues of interests are displayed as sticks while other residues within the active cavity are displayed as lines.

5. Environmental representation of bacterial PKS type III in ATII over DD:

A number of reads from each brine pool layers were retrieved by HMM search. The fidelity of the domains were confirmed by HMM scan against pfam database in a reciprocal HMM approach. Total reads obtained were 25, 58 and 134 in INP, UCL and LCL of ATII, respectively. In contrast, the dearth of returned reads was observed in DD with only three reads in INP of DD. The number of computationally identified reads was increasing towards the deepest layers, where LCL comprised the highest number of reads.

These results reveal the impact of the elevated temperature and aromatic content of ATII layers on microbial metabolic pathways. Utilization of metabolites as substrates from aromatic degradation could possibly be the logic behind the abundance of PKS type III in ATII over DD.

A taxonomic classification at the class level was done to evaluate bacterial PKS type III distribution among brine layers (reads from DD were excluded due to scarcity), (Figure 14). BLASTP against the non-redundant NCBI database was done where the parameters were set to retrieve the best hit only for each read. The three layers were dominated with *Alphaproteobacteria* PKS type III, 65%, 36% and 78% in INP, UCL and LCL, respectively with minor representation of *Gammaproteobacteria* PKS type III, being less than 10% in all layers. Classes *Bacilli* and *Actinobacteria* PKS type IIIs showed the highest abundance within UCL, at percentages of 24% and 29%, respectively. Minor representation was belonging to unclassified species, *Verrucomicrobiace*, *Chlamdyia* and *Clostridia* PKS type IIIs. Within *Alphaproteobacteria* class, the largest reads number belonged to *Rhizobiaceae* family, with the highest percentage in LCL.

Interestingly, members of the *Rhizobiaceae* family showed the ability to grow and utilize hydroaromatic and aromatic compounds as a sole carbon sole in several studies (Frassinetti et al., 1998; Gajendiran & Mahadevan, 1990; Parke, Rynne, & Glenn, 1991; Poonthrigpun et al., 2006; Yessica et al., 2013). However, the overall low number of

reads obtained in the three layers of ATII could not provide a significant estimate of PKS type III representation in the environment.
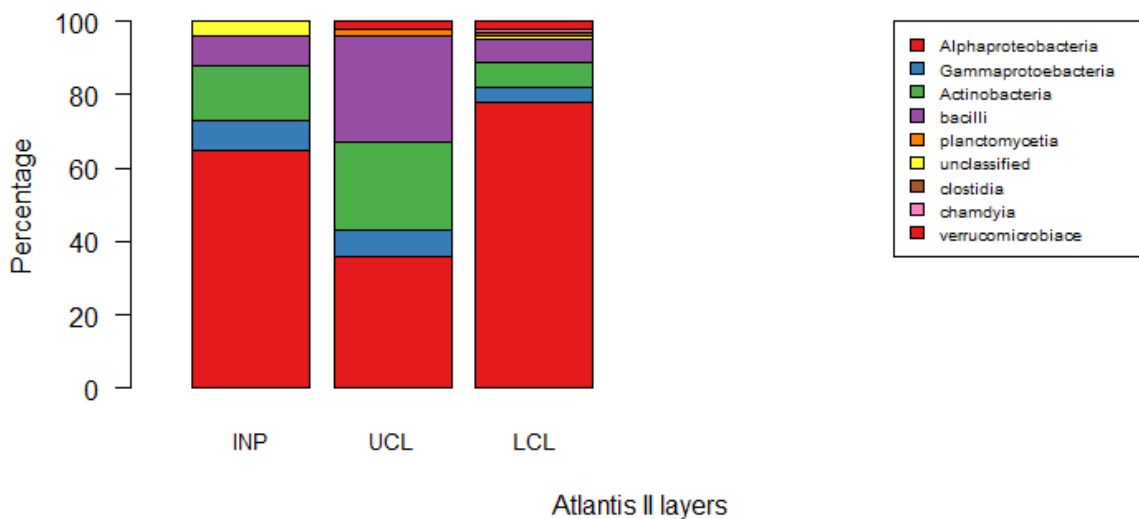
Figure (14) Environmental assessment of bacterial PKS type III in ATII layers
Retrieved reads were classified at class level in each layer of ATII brine pool.

6. Isolation and identification of the putative PKS type III enzyme from ATII brine pool:

6.1 Screening ATII deep brine pool environmental DNA:

To confirm the previous computational findings related to the presence of putative "ATII-ChSyn" ORF in ATII, screening the LCL environmental DNA was done using specific primers. The latter were designed on one read covering the middle region of ORF (including the catalytic cysteine), a positive 256 amplicon was obtained, (Figure 15). Second, three sets of primers were used to obtain the complete ORF including one forward primer at the ORF ATG start, F_Orf and three different reverse primers: R_Orf, R_before_stop and R_downstream1, see Figure (5) and Table (2). Positive amplicons of 1038bp and 1128bp were obtained using the primers pairs [F-Orf + R_before_stop] and [F_Orf + R_downstream1], respectively (Figure 16). No amplicon could be identified with the primer pair [F_Orf + R_Orf]. The positive amplicons were cloned into the p-GEM-T®. The proper orientation of the inserts in several clones was confirmed by colony PCR; see supplementary data, Figure (S1).
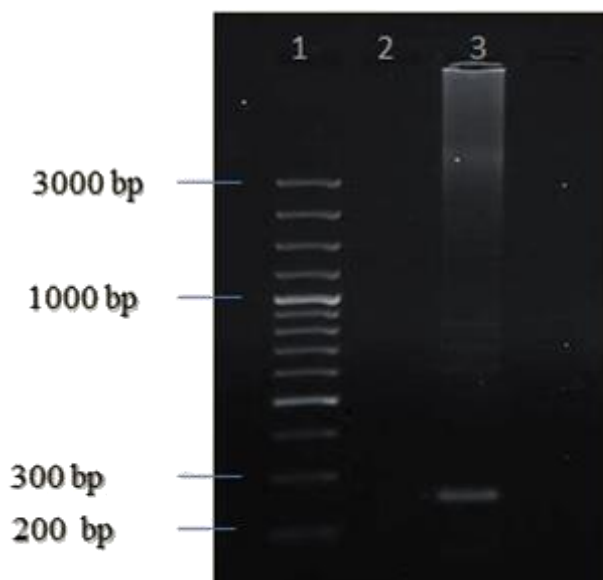
Figure (15) PCR screening of LCL environmental DNA for single read of ATIIChSyn lane 1: GeneRuler™ 100bp Plus DNA ladder (Thermo Scientific), lane 2: negative control and lane 3: 256 bp positive amplicon using F_read and R_read primers.
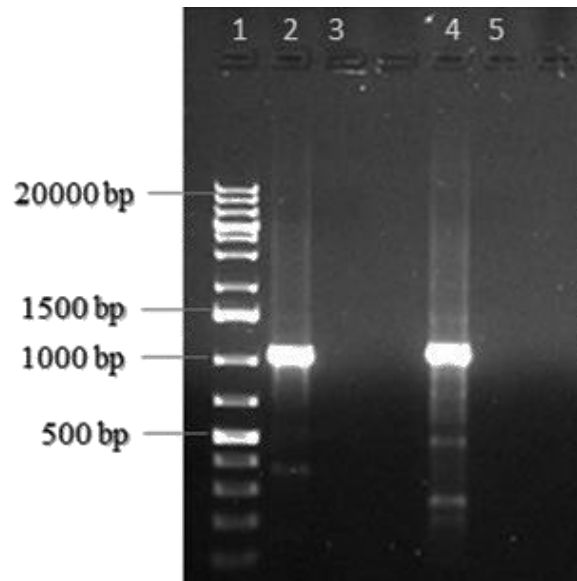
Figure (16) ATII-ChSyn whole gene amplification from the LCL environmental DNA lane 1: GeneRuler™ 1 kb Plus DNA ladder (Thermo Scientific), lane 2: 1038bp positive amplicon obtained using [F-Orf+R_before_stop], lane 3: negative control, lane 4: 1128 bp positive amplicon, [F_Orf+R_downstream1] and lane 5: negative control.

6.2 Sequencing of cloned "ATII-ChSyn" inserts:

Four samples of purified p-GEM-T® clones were sequenced by the dideoxy chain termination method. They were chosen from the 1128 bp amplicon to retrieve the whole ORF. The purified plasmids were checked for proper DNA concentration prior sequencing, see supplementary data Figure (S2). Sample p-GEM-T®/clone2 showed the best sequencing results. Its sequencing was done using these five primers: F_Orf (T_a 58°C), F_read (T_a 58°C), R_downstream1 (T_a 58°C), and p-GEM-T® plasmid's primers: Sp6 promoter reverse primer (T_a 59°C) and T7 promoter forward primer (T_a 51°C). These primers were used to ensure confident coverage of the ORF. The sequencing reads were

assembled by CAP3 giving 1834 bp contig. Alignment results with the computational "ATII-ChSyn" showed 100% identity starting from nt 379 to nt 1431.

7. Expression of ATII-ChSyn in two expression systems: pET-SUMO and pET - 28b+:

Generally, PKS type III are cytosolic enzymes which don't require specific treatment to be functionally active (Pitel, 2009). The cytosolic localization of ATII-ChSyn was predicted according to PSORT prediction results. Also, *Escherichia coli* strain is considered one of the most common systems used for expression of PKSs. Besides its easy culture conditions and rapid growth rate, most importantly it is void from intrinsic polyketide pathways that would create a 'noisy' background upon recombinant expression (Gao, Wang, & Tang, 2010). Accordingly, two expression systems were utilized in this study.

As described previously, **Champion™ pET SUMO** expression plasmid was used to express ATII-ChSyn as N-terminal-6-Histidine-tagged-SUMO fusion recombinant protein. The recombinant pET SUMO/ATII-ChSyn had a predicted theoretical molecular mass of 51 kDa with 473 amino acid residues while native ATII-ChSyn was predicted to be 37.23 kDa. This apparent increase of molecular mass is due to the presence of 11kDa SUMO fused protein and 2kDa 6-Histidine tag. Theoretically, the SUMO protein aids in increasing the expression and solubility of the recombinant protein (Malakhov et al., 2004). The ORF was amplified from p-GEM®-T/clone2 to create A-overhangs then ligated to pET SUMO plasmid and the recombinant plasmids were transformed into *E.coli* BL21 (DE3), see supplementary data Figure (S3) and (S4) for ORF amplification from p-GEM®-T/clone2 and colony PCR of pET SUMO inserts, respectively. As described in Table (2), induction conditions ranged from IPTG concentrations of 1mM to 0.1mM and temperatures ranging between 25°C to 37°C for different time intervals. Induction for 1hour at 37°C with different IPTG concentrations (0.1, 0.2, 0.5 and 1mM IPTG) showed good induction pattern, (Figure 17). However,

using the least IPTG concentration 0.1mM showed high yield of protein trapped in inclusion bodies after cell lysis, (Figure 18).
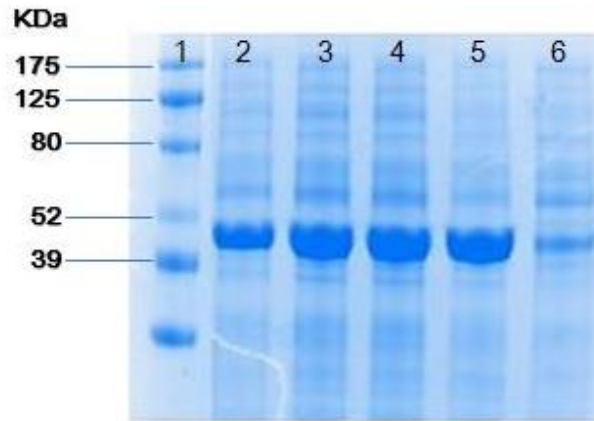
Figure (17) Analysis for "pET SUMO /ATII-ChSyn" expression after IPTG induction on 12% SDS-PAGE at 37°C for 1 hour lane 1: ProSieve® Colour Protein Marker (Lonza), lane 2: 0.1mM IPTG induction, lane 3: 0.2 mM IPTG induction, lane 4 : 0.5 mM IPTG induction, lane 5: 1mM Induction andlane 6: Uninduced

Figure (18) Localization of the expressed pET SUMO /ATII-ChSyn after cell lysis, 0.1mM IPTG for 1 hour at 37°C on 12 % SDS-PAGE lane 1: ProSieve® Colour Protein Marker (Lonza), lane 2: Supernatant and lane 3: cell debris (pellet).

Another expression system was used to avoid the large sized SUMO fusion protein and N-terminal histidine tagged. Since the conserved catalytic domain is located close to the N-terminal region, the conservation of a free N-terminal could be a better approach for recombinant expression of ATII-ChSyn.

The **pET -28b+** expression system with a codon-optimized ATII-ChSyn sequence was utilized. This synthesized version comprises a codon usage tailored for *E.coli*. This optimization step was done to avoid improper translation due to lack of suitable tRNAs for the heterologous protein, see supplementary data (section B). ATII-ChSyn insert was designed between *Nco*I and *Hind*III to be C-terminal-6-Histidine-tagged protein. The recombinant pET -28b+ / ATII-ChSyn had a predicted theoretical weight of 38.75 kDa with 363 amino acid residues due to the histidine tag. The plasmid was first transformed into *E.coli* Top10 for maintenance, then extracted and transformed into E.coli BL21 (DE3), see supplementary data Figure (S5) and (S6) for plasmid purification from *E.coli* Top 10 and colony PCR in *E.coli* BL21 (DE3), respectively. All induction attempts for expression are listed in Table (2). Induction for 1 to 5 hours at 37°C with 0.1mM IPTG indicates a good expression pattern. However, cell lysis after 1 hour induction revealed the presence of most ATII-ChSyn in inclusion bodies even in uninduced sample but with low pattern compared to the induced, (Figure 19).
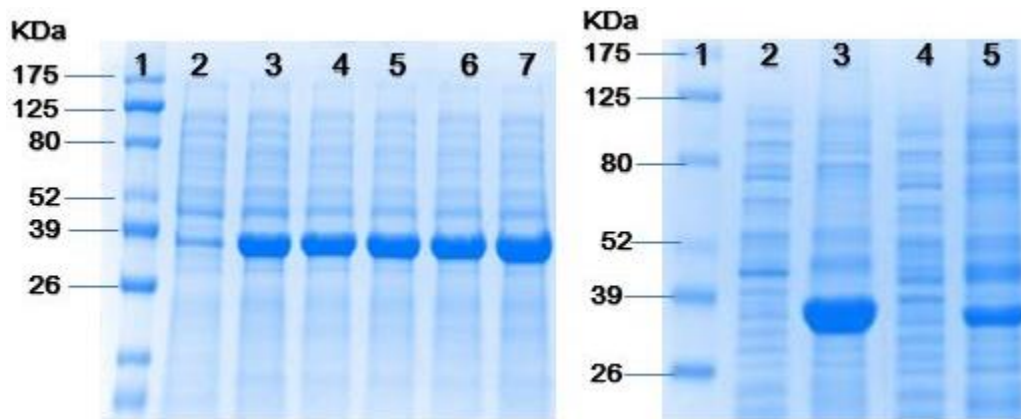
Figure (19) Analysis for "pET -28b+ /ATII-ChSyn" expression after 0.1 mM IPTG induction on 12% SDS-PAGE at 37°C for 1 to 5 hours and the localization of the expressed pET -28b+ /ATIIChSyn after cell lysis, 0.1mM IPTG for 1 hour at 37°C
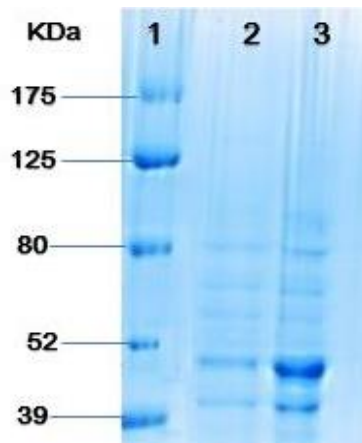
Left panel: lane 1: ProSieve® Colour Protein Marker (Lonza), lane 2: uninduced, lane 3: 0.1 mM IPTG induction after 1 hour, lane 4 :0.1 mM IPTG induction after 2 hours, lane 5: 0.1 mM IPTG induction after 3 hours, lane 6: 0.1 mM IPTG induction after 4 hours and lane 7: 0.1 mM IPTG induction after 5 hours. Right panel: lane 1: ProSieve® Colour Protein Marker (Lonza), ;lane 2: 0.1 mM IPTG induced supernantant compartment, lane 3: Induced cell debris compartment showing protein as inclusion bodies, lane 4 :uninduced supernatant compartment, lane 5: Uninduced cell debris compartment showing protein as inclusion bodies.

Induction attempts were done with lower IPTG concentrations 0.05 and 0.025 mM IPTG but a similar induction pattern was observed at 37°C for 0.05 mM IPTG induction (data not shown). Almost no detecable induction was obsereved compared to uninduced for 0.05 mM IPTG at 15°C and 0.025mM IPTG at 37°C, 15°C (data not shown), see Table (2). This indicates that 0.1mM IPTG is the most suitable concentration for induction. No improved solubility was detected after using a range of solubizing and mild detergent agents in cell lysis buffer as glycerol, tween 20, triton X-100, see suplementary data Figure (S7). Treatment with mild denaturant urea (8M) showed no improved results as well, see suplementary data Figure (S8). A final attempt was done by using IPTG at a concentration of 0.1mM at 15°C and 37°C for 15 min, 30 min, 40 min and 1 hour. Suprisingly, ATII-ChSyn protein appeared in inclusion bodies even after 15 min induction at 15°C, Figure (20). Accordingly, the ATII-ChSyn was inferred to be moslty localized in inclusion bodies and to require denaturing conditions for purification. The pET -28b+ / ATII-ChSyn was chosen for purification.
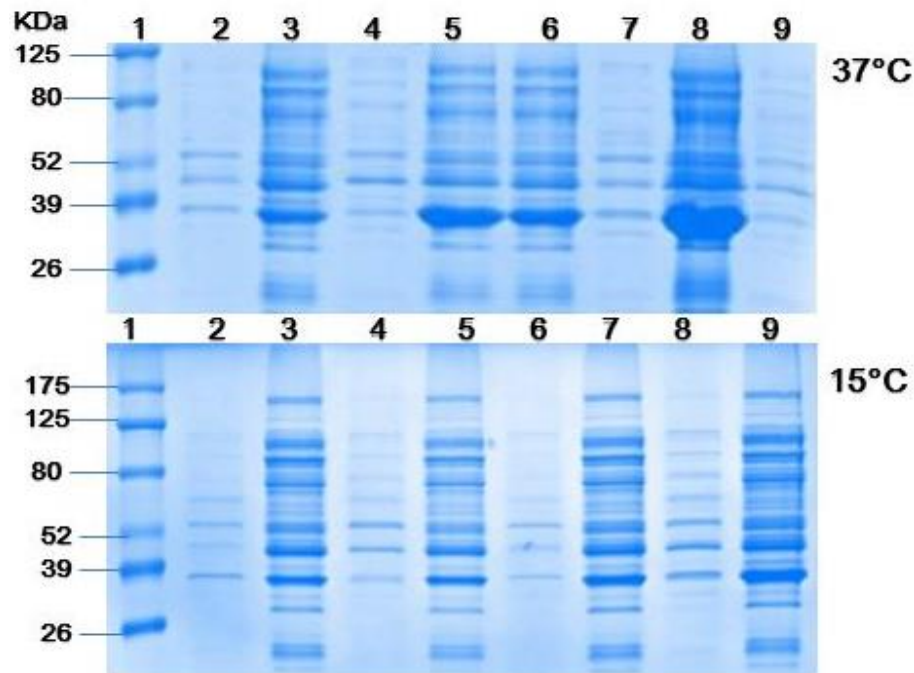
Figure (20) Localization of the expressed pET -28b+ /ATII-ChSyn after cell lysis, 0.1mM IPTG for 1 hour at 37°C and 15°C on 12 % SDS-PAGE 37°C

lane 1: ProSieve® Colour Protein Marker (Lonza), lane 2,4,7 and 9: Supernatant compartment after 15,30,40, 60 minutes induction respectively, lane 3,5,6 and 8: Cell debris compartment after 15,30,40, 60 minutes induction respectively. 15°C: lane 1: ProSieve® Colour Protein Marker (Lonza), lane 2,4,6 and 8: Supernatant compartment after 15,30,40, 60 minutes induction respectively, lane 3,5,7 and 9: Cell debris compartment after 15,30,40, 60 minutes induction respectively. In both temperatures, protein was trapped in inclusion bodies.

The selection of pET expression systems was based upon several studies reporting succesful soluble PKS type III protein utilizing pET vectors (Jeya et al., 2012; Saxena, Yadav, Mohanty, & Gokhale, 2003; Zeng et al., 2012). However, the very high levels of protein expressed in these systems seem to be inappropiate for obtaining soluble ATII-ChSyn recombinant protein even in uninduced conditions,(Figure 19). This might be attributed to the leakiness of the strong promoter P_{T7} even in the repressed state. Less attenuated expression plasmids could improve the solubility of the recombinant protein

with weaker promoter and lower plasmid copy number. Also, shifting to a more stringent host could control the expression level.

8. Purification of recombinant "ATII-ChSyn" from the pET28b+-ATII-ChSyn construct:

Purification of the recombinant "ATII-ChSyn" was done by Ni-NTA purification system (Invitrogen) using the denaturation conditions. The later were chosen to enable recovering of the C-His$_6$-tagged ATII-ChSyn protein trapped in the cell debris. Treatment was done with the strong denaturant guanidinium chloride. Purification under denaturing conditions depends mainly on the gradual decrease of pH gradient to elute the C-His$_6$-tagged ATII-ChSyn. Successful purification was achieved by using 8M urea in all buffers used with no imidazole added starting from pH 7.8 in binding buffer to final pH 4 in the elution buffer. Imidazole was avoided in the buffers to avoid protein precipitation at pH 4. Through 10 elution fraction, 500μl each, purified ATII-ChSyn was obtained with clear absence in the flowthroughs, (Figure 21). Elution fractions with high band intensity were pooled to determine their concentration. Further investigation is required to examine the catalytic features of ATII-ChSyn. Enzymatic assays and product characterization would reveal the functional capability of ATII-ChSyn.
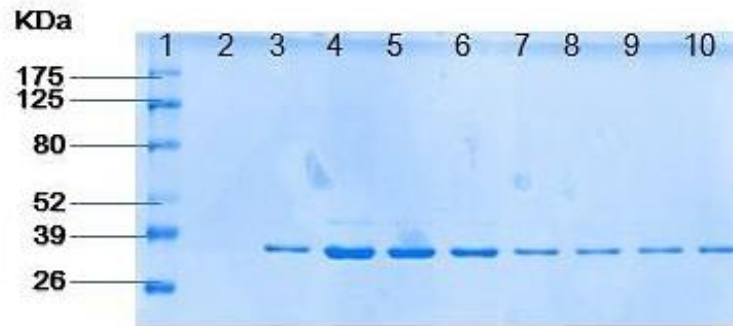
Figure (21) Purification of recombinant "ATII-ChSyn" from pET28b+/ATII-ChSyn under denaturative conditions

lane 1: ProSieve® Colour Protein Marker (Lonza), lane2: Column flowthrough showing no band, lane (3) to (10) elution fractions of purified ATII-ChSyn showing protein elution where lane (4), (5), (6) has the highest band intensity.

CHAPTER 4: CONCLUSIONS AND PERSPECTIVES

Investigations for the biosynthesis of natural polyketides with important biological and pharmaceutical benefits have been augmented over the last 50 years (Staunton & Weissman, 2001). Accordingly, it became apparent that mining for the fascinating machinery of PKSs in extreme environments could introduce novel drug candidates with unique properties. In this study, a putative PKS type III (ATII-ChSyn) was identified, amplified and sequenced from the LCL of ATII brine pool, Red Sea. Multiple sequence alignment with previously characterized PKSs type III in plants and bacteria showed the presence of the conserved catalytic triad and other important residues crucial for activity. Preliminary homology modelling probed an overall conserved fold of ATII-ChSyn structure. It also predicted a possible interaction of catalytic triad with malonyl-CoA substrate within its active site cavity. These observations provide evidence that ATII-ChSyn is probably a member of PKSs type III enzymes. However, more structural prediction of other important residues lining the active cavity, cyclization pocket and dimer interface would reveal the unprecedented biosynthetic machinery of this putative enzyme.

Attempts for expressing the isolated ATII-ChSyn in heterologous host were done. However, optimization is required to avoid insoluble recombinant protein trapped in inclusion bodies. Nevertheless, an enzymatic assay for ATII-ChSyn is required to characterize its catalytic machinery in terms of substrate specifity, functional capabilities and product identification. This could be detected with the aid of instrumental analysis as high performance liquid chromatography (HPLC). A wide range of acyl-CoA substrates should be tested including aromatic substrates as benzoyl-CoA and phenyl-CoA where analysis should give insights to the correspondent product of each substrate. Parallel efforts should be exploited to evaluate the enzyme pH, salinity and thermostable characteristics in response to its extreme LCL habitat. Aside from characterization, size exclusion chromatography and native PAGE will facilitate the elucidation of ATII-ChSyn structure as a homodimer.

Interestingly, environmental assessment disclosed the presence of diverse PKSs type III in ATII layers. In contrary, the adjacent brine pool DD was deprived from PKSs type III. The presence of high aromatic content in ATII could possibly be the justification behind the enzyme's existence. Microbial community within the brine could degrade these aromatic compounds via CoA ligation. Thus, producing CoA tethered metabolites that could be utilized as substrates for PKS type III. The analysis showed that the majority of PKSs type III were located in LCL due to its highest temperature among the other layers within ATII brine pool. As mentioned above, the high temperature and the high pressure within LCL most probably induces the production of aromatic compounds from the organic debris of sea floor. However, the environmental assessment across the layers could not provide a proper significance due to low number of reads especially in the INP layer.

On another aspect, a comprehensive phylogenetic analysis was conducted in this study for PKSs type III. Sequences included bacterial, fungal, amoeba and plants PKSs type III. The sequence of ATII-ChSyn investigated in this study was clustered with other PKSs type III sequences affiliated to *Alphaproteobacteria* class. On the other hand, some insights into the unresolved evolutionary origin of PKS type III were investigated as well. According to this phylogentic analysis, PKSs type III were first gained in prokaryotes from the ancestoral KASs type III class before being aquired to eukaryotes. Preliminary findings showed possible involvement of the bacterium *Parachlamydia acanthamoebae* in evoutionary events from prokaryotes to eukaryotes. A possible HGT of PKSs type III from symbiotic bacteruim *Parachlamydia acanthamoebae* to *Acanthamoeba castellanii str.* amoeba host could be the case. However, more investigations is required to gain further insights into the evolutionary origin of PKSs type III among prokaryotes and eukaryotes as most of bacterial PKSs type III are yet not characterized particularly the sequence of the bacterium *Parachlamydia acanthamoebae* that is associated with the eukaryotes sequences.

Last but not least, this metagenomic study proposes the ATII brine pool as an unvisited warehouse for natural polyketide molecules biosynthesized by PKSs type III machinery. Undoubtedly, a lot of efforts are yet to be done for an exhaustive analysis. Collectively, these efforts -if accomplished- could reveal exciting and unique polyketide products.

REFERENCES

Abe, I., & Morita, H. (2010). Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Natural Product Reports*, *27*(6), 809.

Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., … Boonham, N. (2009). Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Molecular Plant Pathology*, *10*(4), 537–545.

Akiyama, T., Shibuya, M., Liu, H.-M., & Ebizuka, Y. (1999). p-Coumaroyltriacetic acid synthase, a new homologue of chalcone synthase, from Hydrangea macrophylla var. thunbergii. *European Journal of Biochemistry*, *263*(3), 834–839

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402.

Anschutz, P., & Blanc, G. (1996). Heat and salt fluxes in the Atlantis II Deep (Red Sea). *Earth and Planetary Science Letters*, *142*(1), 147–159.

Antunes, A., Ngugi, D. K., & Stingl, U. (2011). Microbiology of the Red Sea (and other) deep-sea anoxic brine lakes. *Environmental Microbiology Reports*, *3*(4), 416–433.

Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, *22*(2), 195–201.

Austin, M. B. (2005). *Structural and mechanistic determinants of biosynthetic functional diversity in the type III polyketide synthase superfamily* (Ph.D.). University of California, San Diego, United States . Retrieved from http://search.proquest.com.library.aucegypt.edu

Austin, M. B., & Noel, J. P. (2003). The chalcone synthase superfamily of type III polyketide synthases. *Natural Product Reports*, *20*(1), 79–110.

Austin, M. B., Saito, T., Bowman, M. E., Haydock, S., Kato, A., Moore, B. S., … Noel, J. P. (2006). Biosynthesis of Dictyostelium discoideum differentiation-inducing factor by a hybrid type I fatty acid-type III polyketide synthase. *Nature Chemical Biology*, *2*(9), 494–502.

Bangera, M. G., & Thomashow, L. S. (1999). Identification and characterization of a gene cluster for synthesis of the polyketide antibiotic 2,4-diacetylphloroglucinol from Pseudomonas fluorescens Q2-87. *Journal of Bacteriology*, *181*(10), 3155–3163.

Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E., & House, C. H. (2008). Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proceedings of the National Academy of Sciences*, *105*(30), 10583-10588.

Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J. C., … Hochstrasser, D. (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, *14*(10), 1023–1031.

Blanc, G., & Anschutz, P. (1995). New stratification in the hydrothermal brine system of the Atlantis II Deep, Red Sea. *Geology*, *23*(6), 543–546.

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., & Schwede, T. (2008). Protein structure homology modeling using SWISS-MODEL workspace. *Nature Protocols*, *4*(1), 1–13.

Brady, S. F., Simmons, L., Kim, J. H., & Schmidt, E. W. (2009). Metagenomic approaches to natural products from free-living and symbiotic organisms. *Natural Product Reports*, *26*(11), 1488–1503.

Brazelton, W. J., & Baross, J. A. (2009). Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *The ISME Journal*, *3*(12), 1420–1424.

Clarke, M., Lohan, A. J., Liu, B., Lagkouvardos, I., Roy, S., Zafar, N., … Loftus, B. J. (2013). Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biology*, *14*(2), R11.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., … Barrell, B. G. (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, *393*(6685), 537–544.

Collingro, A., Tischler, P., Weinmaier, T., Penz, T., Heinz, E., Brunham, R. C., … Horn, M. (2011). Unity in Variety-- The Pan-Genome of the Chlamydiae. *Molecular Biology and Evolution*, *28*(12), 3253–3270.

Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J., & Leadlay, P. F. (1990). An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. *Nature*, *348*(6297), 176–178.

Courtois, S., Cappellano, C. M., Ball, M., Francou, F.-X., Normand, P., Helynck, G., … Pernodet, J.-L. (2003). Recombinant Environmental Libraries Provide Access to Microbial Diversity for Drug Discovery from Natural Products. *Applied and Environmental Microbiology*, *69*(1), 49–55.

Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R., & Wittwer, P. (2005). Metagenomic gene discovery: past, present and future. *Trends in Biotechnology*, *23*(6), 321–329.

Dawe, J. H., Porter, C. T., Thornton, J. M., & Tabor, A. B. (2003). A template search reveals mechanistic similarities and differences in β-ketoacyl synthases (KAS) and related enzymes. *Proteins: Structure, Function, and Bioinformatics*, *52*(3), 427–435.

DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N.-U., … Karl, D. M. (2006). Community Genomics Among Stratified Microbial Assemblages in the Ocean's Interior. *Science*, *311*(5760), 496–503.

Demain, A. L. (1999). Pharmaceutically active secondary metabolites of microorganisms. *Applied Microbiology and Biotechnology*, *52*(4), 455–463.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., … Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, *36*(Web Server issue), W465–469.

Dick, G. J., Anantharaman, K., Baker, B. J., Li, M., Reed, D. C., & Sheik, C. S. (2013). The microbiology of deep-sea hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water column habitats. *Frontiers in Microbiology*, *4,* 124.

Djordjevic, M. A., Gabriel, D. W., & Rolfe, B. G. (1987). Rhizobium-The Refined Parasite of Legumes. *Annual Review of Phytopathology*, *25*(1), 145–168.

Donadio, S., Staver, M. J., Mcalpine, J. B., Swanson, S. J., & Katz, L. (1991). Modular Organization of Genes Required for Complex Polyketide Biosynthesis. *Science*, *252*(5006), 675–679.

Eden, P. A., Schmidt, T. M., Blakemore, R. P., & Pace, N. R. (1991). Phylogenetic analysis of Aquaspirillum magnetotacticum using polymerase chain reaction-amplified 16S rRNA-specific DNA. *International Journal of Systematic Bacteriology*, *41*(2), 324–325.

Edgar, R. C. (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*, 113.

Edgar, R. C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M.-Y., … Sali, A. (2007). Comparative protein structure modeling using MODELLER. *Current Protocols in Protein Science*, *Chapter 2*, Unit 2.9.

Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, *8*(3), 186–194.

Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, *8*(3), 175–185.

Ververidis ,F., Trantas, E., Douglas, C., Vollmer, G., Kretzschmar, G., & Panopoulos, N. (2007). Biotechnology of
flavonoids and other phenylpropanoid-derived natural products. Part I: Chemical diversity, impacts on plant
biology and human health. *Biotechnology Journal*, *2*(10), 1214–1234.

Frassinetti, S., Setti, L., Corti, A., Farrinelli, P., Montevecchi, P., & Vallini, G. (1998). Biodegradation of
dibenzothiophene by a nodulating isolate of Rhizobium meliloti. *Canadian Journal of Microbiology*, *44*(3),
289–297.

Fukumoto, T., Matsuo, J., Hayashi, M., Oguri, S., Nakamura, S., Mizutani, Y., … Yamaguchi, H. (2010). Impact of
Free-Living Amoebae on Presence of Parachlamydia acanthamoebae in the Hospital Environment and Its
Survival In Vitro without Requirement for Amoebae. *Journal of Clinical Microbiology*, *48*(9), 3360–3365.

Funa, N., Ohnishi, Y., Ebizuka, Y., & Horinouchi, S. (2002). Alteration of reaction and substrate specificity of a
bacterial type III polyketide synthase by site-directed mutagenesis. *The Biochemical Journal*, *367*(Pt 3), 781–
789.

Funa, N., Ohnishi, Y., Fujii, I., Shibuya, M., Ebizuka, Y., & Horinouchi, S. (1999). A new pathway for polyketide
synthesis in microorganisms. *Nature*, *400*(6747), 897–899.

Funa, N., Ozawa, H., Hirata, A., & Horinouchi, S. (2006). Phenolic lipid synthesis by type III polyketide synthases is
essential for cyst formation in Azotobacter vinelandii. *Proceedings of the National Academy of Sciences of
the United States of America*, *103*(16), 6356–6361.

Funabashi, M., Funa, N., & Horinouchi, S. (2008). Phenolic lipids synthesized by type III polyketide synthase confer
penicillin resistance on Streptomyces griseus. *The Journal of Biological Chemistry*, *283*(20), 13983–13991.

Gajendiran, N., & Mahadevan, A. (1990). Growth of Rhizobium sp. in the presence of catechol. *Plant and Soil*, *125*(2),
207–211.

Gao, X., Wang, P., & Tang, Y. (2010). Engineered polyketide biosynthesis and biocatalysis in Escherichia coli. *Applied
Microbiology and Biotechnology*, *88*(6), 1233–1242.

Ghai, R., Rodriguez-Valera, F., McMahon, K. D., Toyama, D., Rinke, R., Cristina Souza de Oliveira, T., … Henrique-
Silva, F. (2011). Metagenomics of the Water Column in the Pristine Upper Course of the Amazon River.
PLoS ONE, *6*(8), e23785.

Ghosh, R., Chhabra, A., Phatale, P. A., Samrat, S. K., Sharma, J., Gosain, A., … Gokhale, R. S. (2008). Dissecting the
Functional Role of Polyketide Synthases in Dictyostelium discoideum: biosynthesis of the differentiation
regulating factor 4-methyl-5-pentylbenzene-1,3-diol. *Journal of Biological Chemistry*, *283*(17), 11348–
11354.

Ginolhac, A., Jarrin, C., Gillet, B., Robe, P., Pujic, P., Tuphile, K., … Nalin, R. (2004). Phylogenetic Analysis of

    Polyketide Synthase I Domains from Soil Metagenomic Libraries Allows Selection of Promising Clones.

    Applied and Environmental Microbiology, *70*(9), 5522–5527.

Giovannoni, S. J., DeLong, E. F., Schmidt, T. M., & Pace, N. R. (1990). Tangential flow filtration and preliminary

    phylogenetic analysis of marine picoplankton. *Applied and Environmental Microbiology*, *56*(8), 2572–2575.

Gokhale, R. S., Saxena, P., Chopra, T., & Mohanty, D. (2007). Versatile polyketide enzymatic machinery for the

    biosynthesis of complex mycobacterial lipids. *Natural Product Reports*, *24*(2), 267–277.

Gordon, D. (2002). Viewing and Editing Assembled Sequences Using Consed. In Andreas Baxevanis Al (Ed). *Current*

    Protocols in Bioinformatics. Chapter 11: Unit 11.2 .John Wiley & Sons, Inc.

Gordon, D., Abajian, C., & Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Research*, *8*(3),

    195–202.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., & Lopez, R. (2010). A new bioinformatics

    analysis tools framework at EMBL–EBI. *Nucleic Acids Research*, *38*(suppl 2), W695–W699.

Gross, F., Luniak, N., Perlova, O., Gaitatzis, N., Jenke-Kodama, H., Gerth, K., … Müller, R. (2006). Bacterial type III

    polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites

    by heterologous expression in pseudomonads. *Archives of Microbiology*, *185*(1), 28–38.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and

    Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0.

    Systematic Biology, *59*(3), 307–321.

Haapalainen, A. M., Meriläinen, G., & Wierenga, R. K. (2006). The thiolase superfamily: condensing enzymes with

    diverse reaction specificities. *Trends in Biochemical Sciences*, *31*(1), 64–71.

Handelsman, J., Liles, M., Mann, D., Riesenfeld, C., & Goodman, R. M. (2002). Cloning the metagenome: Culture-

    independent access to thediversity and functions of the uncultivated microbial world. In Brendan Wren (Ed.),

    Methods in Microbiology .Volume 33: 241–255. Academic Press.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the

    chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, *5*(10), R245–

    R249.

Hårdeman, F., & Sjöling, S. (2007). Metagenomic approach for the isolation of a novel low-temperature-active lipase

    from uncultured bacteria of marine sediment. *FEMS Microbiology Ecology*, *59*(2), 524–534.

Hartmann, M., Scholten, J. C., Stoffers, P., & Wehner, F. (1998). Hydrographic structure of brine-filled deeps in the Red Sea—new results from the Shaban, Kebrit, Atlantis II, and Discovery Deep. *Marine Geology*, *144*(4), 311–330.

Hertweck, C., Luzhetskyy, A., Rebets, Y., & Bechthold, A. (2007). Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Natural Product Reports*, *24*(1), 162–190.

Hopwood, D. A., & Sherman, D. H. (1990). Molecular Genetics of Polyketides and its Comparison to Fatty Acid Biosynthesis. *Annual Review of Genetics*, *24*(1), 37–62.

Hranueli, D., Peri, N., Borovicka, B., Bogdan, S., Cullum, J., Waterman, P. G., & Hunter, I. S. (2001). Molecular Biology of Polyketide Biosynthesis. *Food Technology and Biotechnology*, *39*(3), 203–213.

Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research*, *9*(9), 868–877.

Huse, S. M., Dethlefsen, L., Huber, J. A., Welch, D. M., Relman, D. A., & Sogin, M. L. (2008). Exploring Microbial Diversity and Taxonomy Using SSU rRNA Hypervariable Tag Sequencing. *PLoS Genet*, *4*(11), e1000255.

Izumikawa, M., Shipley, P. R., Hopke, J. N., O'Hare, T., Xiang, L., Noel, J. P., & Moore, B. S. (2003). Expression and characterization of the type III polyketide synthase 1,3,6,8-tetrahydroxynaphthalene synthase from Streptomyces coelicolor A3(2). *Journal of Industrial Microbiology & Biotechnology*, *30*(8), 510–515.

Jeya, M., Kim, T.-S., Kumar Tiwari, M., Li, J., Zhao, H., & Lee, J.-K. (2012). A type III polyketide synthase from Rhizobium etli condenses malonyl CoAs to a heptaketide pyrone with unusually high catalytic efficiency. *Molecular bioSystems*, *8*(12), 3103–3106.

Jez, J. M., Dixon, R. A., Bowman, M. E., & Ferrer, J.-L. (1999). Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nature Structural Biology*, *6*(8), 775–784.

Jez, J. M., Ferrer, J. L., Bowman, M. E., Austin, M. B., Schröder, J., Dixon, R. A., & Noel, J. P. (2001). Structure and mechanism of chalcone synthase-like polyketide synthases. *Journal of Industrial Microbiology & Biotechnology*, *27*(6), 393–398.

Jez, J. M., Ferrer, J. L., Bowman, M. E., Dixon, R. A., & Noel, J. P. (2000). Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry*, *39*(5), 890–902.

Jez, J. M., & Noel, J. P. (2000). Mechanism of Chalcone Synthase pKa of the catalytic cysteine and the role of the conserved histidine in a plant polyketide synthase. *Journal of Biological Chemistry*, *275*(50), 39640–39646.

Jiang, C., Kim, S. Y., & Suh, D.-Y. (2008). Divergent evolution of the thiolase superfamily and chalcone synthase family. *Molecular Phylogenetics and Evolution*, *49*(3), 691–701.

Katsuyama, Y., Miyazono, K., Tanokura, M., Ohnishi, Y., & Horinouchi, S. (2011). Structural and Biochemical
Elucidation of Mechanism for Decarboxylative Condensation of ?-Keto Acid by Curcumin Synthase. *The
Journal of Biological Chemistry*, *286*(8), 6659–6668.

Katsuyama, Y., & Ohnishi, Y. (2012). Type III polyketide synthases in microorganisms. *Methods in Enzymology*, *515*,
359–377.

Kennedy, J., Flemer, B., Jackson, S. A., Lejon, D. P. H., Morrissey, J. P., O'Gara, F., & Dobson, A. D. W. (2010).
Marine Metagenomics: New Tools for the Study and Exploitation of Marine Microbial Metabolism. *Marine
Drugs*, *8*(3), 608–628.

Kennedy, J., Marchesi, J. R., & Dobson, A. D. (2008). Marine metagenomics: strategies for the discovery of novel
enzymes with biotechnological applications from marine environments. *Microbial Cell Factories*, *7*(1), 27.

Koressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design program Primer3.
Bioinformatics, *23*(10), 1289–1291.

Kreuzaler, F., & Hahlbrock, K. (1972). Enzymatic synthesis of aromatic compounds in higher plants: Formation of
naringenin (5,7,4′-trihydroxyflavanone) from p-coumaroyl coenzyme A and malonyl coenzyme A. *FEBS
Letters*, *28*(1), 69–72.

Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., & Hugenholtz, P. (2008). A bioinformatician's guide to
metagenomics. *Microbiology and Molecular Biology Reviews*, *72*(4), 557–578.

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*,
227(5259), 680–685.

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L., & Pace, N. R. (1985). Rapid determination of 16S
ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences*,
82(20), 6955–6959.

Letunic, I., & Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and
annotation. *Bioinformatics (Oxford, England)*, *23*(1), 127–128.

Letunic, I., & Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made
easy. *Nucleic Acids Research*, *39*(Web server issue), W475–W478.

Lewin, A., Wentzel, A., & Valla, S. (2013). Metagenomics of microbial life in extreme temperature environments.
Current Opinion in Biotechnology, *24*(3), 516–525.

Lim, H. K., Chung, E. J., Kim, J.-C., Choi, G. J., Jang, K. S., Chung, Y. R., … Lee, S. W. (2005). Characterization of a Forest Soil Metagenome Clone That Confers Indirubin and Indigo Production on Escherichia coli. *Applied and Environmental Microbiology*, *71*(12), 7768–7777.

Malakhov, M. P., Mattern, M. R., Malakhova, O. A., Drinker, M., Weeks, S. D., & Butt, T. R. (2004). SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *Journal of Structural and Functional Genomics*, *5*(1-2), 75–86.

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., … Bryant, S. H. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*, *39*(Database issue), 225–229.

Mardis, E. R. (2008). Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics*, *9*(1), 387–402.

Mavrodi, O. V., McSpadden Gardener, B. B., Mavrodi, D. V., Bonsall, R. F., Weller, D. M., & Thomashow, L. S. (2001). Genetic Diversity of phlD from 2,4-Diacetylphloroglucinol-Producing Fluorescent Pseudomonas spp. *Phytopathology*, *91*(1), 35–43.

Melo, F., Sánchez, R., & Sali, A. (2002). Statistical potentials for fold assessment. *Protein Science: A Publication of the Protein Society*, *11*(2), 430–448.

Mirete, S., Figueras, C. G. de, & González-Pastor, J. E. (2007). Novel Nickel Resistance Genes from the Rhizosphere Metagenome of Plants Adapted to Acid Mine Drainage. *Applied and Environmental Microbiology*, *73*(19), 6001–6011.

Moffitt, M. C., & Neilan, B. A. (2003). Evolutionary Affiliations Within the Superfamily of Ketosynthases Reflect Complex Pathway Associations. *Journal of Molecular Evolution*, *56*(4), 446–457.

Moore, B. S., & Hopke, J. N. (2001). Discovery of a new bacterial polyketide biosynthetic pathway. *Chembiochem: A European Journal of Chemical Biology*, *2*(1), 35–38.

Ng, C., DeMaere, M. Z., Williams, T. J., Lauro, F. M., Raftery, M., Gibson, J. A., … Cavicchioli, R. (2010). Metaproteogenomic analysis of a dominant green sulfur bacterium from Ace Lake, Antarctica. *The ISME Journal*, *4*(8), 1002–1019.

Niehaus, F., Bertoldo, C., Kähler, M., & Antranikian, G. (1999). Extremophiles as a source of novel enzymes for industrial application. *Applied Microbiology and Biotechnology*, *51*(6), 711–729.

Noguchi, H., Taniguchi, T., & Itoh, T. (2008). MetaGeneAnnotator: Detecting Species-Specific Patterns of Ribosomal Binding Site for Precise Gene Prediction in Anonymous Prokaryotic and Phage Genomes. *DNA Research*, *15*(6), 387–396.

Parke, D., Rynne, F., & Glenn, A. (1991). Regulation of phenolic catabolism in Rhizobium leguminosarum biovar trifolii. *Journal of Bacteriology*, *173*(17), 5546–5550.

Parsley, L. C., Linneman, J., Goode, A. M., Becklund, K., George, I., Goodman, R. M., … Liles, M. R. (2011). Polyketide synthase pathways identified from a metagenomic library are derived from soil Acidobacteria. *FEMS Microbiology Ecology*, *78*(1), 176–187.

Piel, J., Butzke, D., Fusetani, N., Hui, D., Platzer, M., Wen, G., & Matsunaga, S. (2005). Exploring the Chemistry of Uncultivated Bacterial Symbionts: Antitumor Polyketides of the Pederin Family. *Journal of Natural Products*, *68*(3), 472–479.

Piel, J., Hui, D., Fusetani, N., & Matsunaga, S. (2004). Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Environmental Microbiology*, *6*(9), 921–927.

Pitel, S. B. R. (2009). *Type III polyketide synthases: Discovery, characterization, and engineering*. University of Illinois at Urbana-Champaign, United States.

Poinar, H. N., Schwarz, C., Qi, J., Shapiro, B., MacPhee, R. D. E., Buigues, B., … Schuster, S. C. (2006). Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. *Science*, *311*(5759), 392–394.

Poonthrigpun, S., Pattaragulwanit, K., Paengthai, S., Kriangkripipat, T., Juntongjin, K., Thaniyavarn, S., … Pinphanichakarn, P. (2006). Novel intermediates of acenaphthylene degradation by Rhizobium sp. strain CU-A1: evidence for naphthalene-1,8-dicarboxylic acid metabolism. *Applied and Environmental Microbiology*, *72*(9), 6034–6039.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., … Finn, R. D. (2011). The Pfam protein families database. *Nucleic Acids Research*, *40*(D1), D290–D301.

Qian, P.-Y., Wang, Y., Lee, O. O., Lau, S. C. K., Yang, J., Lafi, F. F., … Wong, T. Y. (2011). Vertical stratification of microbial communities in the Red Sea revealed by 16S rDNA pyrosequencing. *The ISME Journal*, *5*(3), 507–518.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., … Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59–65.

Ramboz, C., Oudin, E., & Thisse, Y. (1988). Geyser-type discharge in Atlantis II Deep, Red Sea; evidence of boiling from fluid inclusions in epigenetic anhydrite. *Canadian Mineralogist*, *26*, 765–786.

Redburn, A. C., & Patel, B. K. (1993). Phylogenetic analysis of Desulfotomaculum thermobenzoicum using polymerase chain reaction-amplified 16S rRNA-specific DNA. *FEMS Microbiology Letters*, *113*(1), 81–86.

Rees, H. C., Grant, S., Jones, B., Grant, W. D., & Heaphy, S. (2003). Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles*, *7*(5), 415–421.

Roesch, L. F. W., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K. M., Kent, A. D., … Triplett, E. W. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME Journal*, *1*(4), 283–290.

Roessner, C. A., & Scott, A. I. (1996). Genetically engineered synthesis of natural products: From alkaloids to corrins. *Annual Review of Microbiology*, *50*(1), 467.

Rothschild, L. J., & Mancinelli, R. L. (2001). Life in extreme environments. *Nature*, *409*(6823), 1092–1101.

Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., … Venter, J. C. (2007). The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol*, *5*(3), e77.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics (Oxford, England)*, *16*(10), 944–945.

Samappito, S., Page, J. E., Schmidt, J., De-Eknamkul, W., & Kutchan, T. M. (2003). Aromatic and pyrone polyketides synthesized by a stilbene synthase from Rheum tataricum. *Phytochemistry*, *62*(3), 313–323.

Saxena, P., Yadav, G., Mohanty, D., & Gokhale, R. S. (2003). A new family of type III polyketide synthases in Mycobacterium tuberculosis. *The Journal of Biological Chemistry*, *278*(45), 44780–44790.

Schirmer, A., Gadkari, R., Reeves, C. D., Ibrahim, F., DeLong, E. F., & Hutchinson, C. R. (2005). Metagenomic Analysis Reveals Diverse Polyketide Synthase Gene Clusters in Microorganisms Associated with the Marine Sponge Discodermia dissoluta. *Applied and Environmental Microbiology*, *71*(8), 4840–4849.

Schloss, P. D., & Handelsman, J. (2005). Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome Biology*, *6*(8), 229.

Schmeisser, C., Steele, H., & Streit, W. R. (2007). Metagenomics, biotechnology with non-culturable microbes. *Applied Microbiology and Biotechnology*, *75*(5), 955–962.

Schmidt, T. M., DeLong, E. F., & Pace, N. R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of Bacteriology*, *173*(14), 4371–4378.

Schuster, S. C. (2008). Next-generation sequencing transforms today's biology. *Nature Methods*, *5*(1), 16–18.

Shen, B. (2000). Biosynthesis of Aromatic Polyketides. In Leeper, F.J. & Vederas, J.C. (Eds.), *Biosynthesis* (pp. 1–51). Springer Berlin Heidelberg.

Shin-ya, K., Furihata, K., Hayakawa, Y., & Seto, H. (1990). Biosynthetic studies of naphterpin, a terpenoid metabolite of streptomyces. *Tetrahedron Letters*, *31*(42), 6025–6026.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., … Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*, 539.

Simon, C., & Daniel, R. (2011). Metagenomic Analyses: Past and Future Trends. *Applied and Environmental Microbiology*, *77*(4), 1153–1161.

Simoneit, B. R. (1993). Aqueous high-temperature and high-pressure organic geochemistry of hydrothermal vent systems. *Geochimica et Cosmochimica Acta*, *57*, 3231–3243.

Snyder, S. A., Tang, Z.-Y., & Gupta, R. (2009). Enantioselective Total Synthesis of (−)-Napyradiomycin A1 via Asymmetric Chlorination of an Isolated Olefin. *Journal of the American Chemical Society*, *131*(16), 5744–5745.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., … Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere." *Proceedings of the National Academy of Sciences*, *103*(32), 12115–12120.

Song, L., Barona-Gomez, F., Corre, C., Xiang, L., Udwary, D. W., Austin, M. B., … Challis, G. L. (2006). Type III Polyketide Synthase β-Ketoacyl-ACP Starter Unit and Ethylmalonyl-CoA Extender Unit Selectivity Discovered by Streptomyces coelicolor Genome Mining. *Journal of the American Chemical Society*, *128*(46), 14754–14755.

Staunton, J., & Weissman, K. J. (2001). Polyketide biosynthesis: a millennium review. *Natural Product Reports*, *18*(4), 380–416.

Steele, H. L., Jaeger, K.-E., Daniel, R., & Streit, W. R. (2009). Advances in Recovery of Novel Biocatalysts from Metagenomes. *Journal of Molecular Microbiology and Biotechnology*, *16*(1-2), 25–37.

Streit, W. R., & Schmitz, R. A. (2004). Metagenomics – the key to the uncultured microbes. *Current Opinion in Microbiology*, *7*(5), 492–498.

Suh, D. Y., Fukuma, K., Kagami, J., Yamazaki, Y., Shibuya, M., Ebizuka, Y., & Sankawa, U. (2000). Identification of amino acid residues important in the cyclization reactions of chalcone and stilbene synthases. *Biochemical Journal*, *350*(Pt 1), 229–235.

Ueda, K., Kim, K. M., Beppu, T., & Horinouchi, S. (1995). Overexpression of a gene cluster encoding a chalcone synthase-like protein confers redbrown pigment production in Streptomyces griseus. *The Journal of Antibiotics*, *48*(7), 638–646.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), e115.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., … Smith, H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science (New York, N.Y.)*, *304*(5667), 66–74.

Wang, Y., Cao, H., Zhang, G., Bougouffa, S., Lee, O. O., Al-Suwailem, A., & Qian, P.-Y. (2013). Autotrophic Microbe Metagenomes and Metabolic Pathways Differentiate Adjacent Red Sea Brine Pools. *Scientific Reports*, *3*,1748.

Wang, Y., Yang, J., Lee, O. O., Dash, S., Lau, S. C. K., Al-Suwailem, A., … Qian, P.-Y. (2011). Hydrothermally generated aromatic compounds are consumed by bacteria colonizing in Atlantis II Deep of the Red Sea. *The ISME Journal*, *5*(10), 1652–1659.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191.

Wooley, J. C., & Ye, Y. (2009). Metagenomics: Facts and Artifacts, and Computational Challenges*. *Journal of Computer Science and Technology*, *25*(1), 71–81.

Xu, J. (2006). Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Molecular Ecology*, *15*(7), 1713–1731.

Xu, M., Xiao, X., & Wang, F. (2008). Isolation and characterization of alkane hydroxylases from a metagenomic library of Pacific deep-sea sediment. *Extremophiles: Life under Extreme Conditions*, *12*(2), 255–262.

Yessica, G.-P., Alejandro, A., Ronald, F.-C., José, A. J., Esperanza, M.-R., Samuel, C.-S. J., … Ormeño-Orrillo, E. (2013). Tolerance, growth and degradation of phenanthrene and benzo[a]pyrene by Rhizobium tropici CIAT 899 in liquid culture medium. *Applied Soil Ecology*, *63*, 105–111.

Yohei Katsuyama. (2010). 1.05 - Microbial Type III Polyketide Synthases. In Lew Mander & Hung-Wen (Ben) Liu (Eds.), *Comprehensive Natural Products II* (pp. 147–170). Oxford: Elsevier.

Yu, D., Xu, F., Zeng, J., & Zhan, J. (2012). Type III polyketide synthases in natural product biosynthesis. *IUBMB Life*, *64*(4), 285–295.

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., … Brinkman, F. S. L. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics (Oxford, England)*, *26*(13), 1608–1615.

Zeng, J., Decker, R., & Zhan, J. (2012). Biochemical characterization of a type III polyketide biosynthetic gene cluster from Streptomyces toxytricini. *Applied Biochemistry and Biotechnology*, *166*(4), 1020–1033.

Zhao, J., Yang, N., & Zeng, R. (2008). Phylogenetic analysis of type I polyketide synthase and nonribosomal peptide synthetase genes in Antarctic sediment. *Extremophiles*, *12*(1), 97–105.

SUPPLEMENTARY DATA

Section A:

Figure (S 1) Colony PCR for p-GEM-T® transformants

lane 1: GeneRuler™ 1 kb Plus DNA ladder (Thermo Scientific). lane 2-6: 1038bp positive amplicon [F-Orf+R_before_stop] for five picked colonies. lane7: Negative control of the amplicon 1038bp. lane9-13: 1128 bp positive amplicon, [F_Orf+R_downstream1] and lane 8: Negative Control of the amplicon 1128bp.
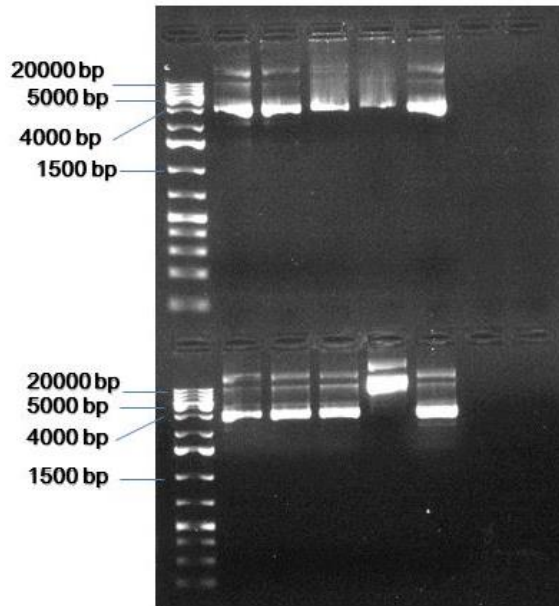
Figure (S 2) Plasmid purification of correct p-GEM-T® transformants

lane 1 & 7: GeneRuler™ 1 kb Plus DNA ladder (Thermo Scientific). lane 2-6: extracted p-GEM-T® with 1038pb insert (total 4038bp) [F-Orf+R_before_stop]. lane 8-12: extracted p-GEM-T® with 1128pb insert (total 4128bp) [F_Orf+R_downstream1].

- **Expression of "ATII-ChSyn" in two expression systems pET SUMO / pET28b+:**
- **Sample preparation: pET SUMO expression system:**

Sample p-GEM-T®/clone2 was amplified [F_Orf+R_downstream1] to amplify the insert (1128bp) creating A-overhangs & gel extracted, Figure (S3). This insert was ligated into pET SUMO plasmid then transformed in *E.coli* BL21 (DE3). Colony PCR was done using Forward SUMO and R_Downstream1, Figure (S4)
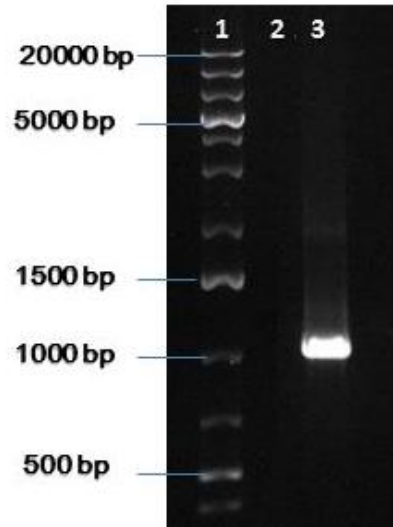
Figure (S 3) Amplification of the ATII-ChSyn ORF for pET SUMO vector ligation

lane1: GeneRuler™ 1 kb Plus DNA ladder (Thermo Scientific). lane2: Negative Control.lane3: 1128 bp positive amplicon [F_Orf+R_downstream1].
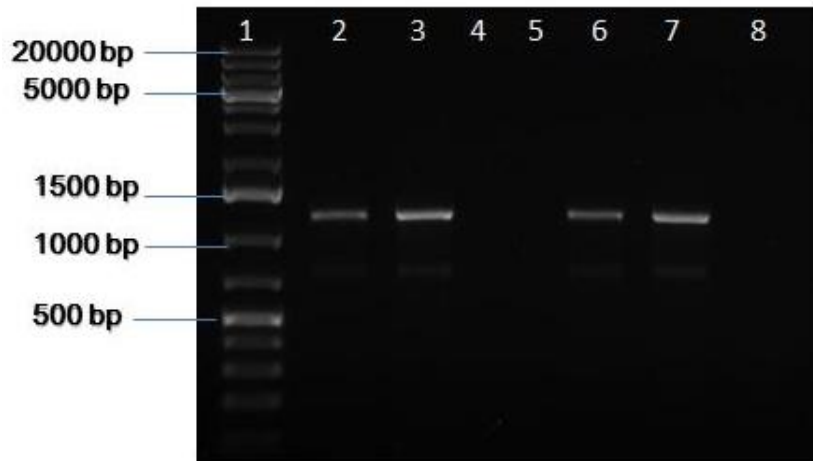

Figure (S 4) Colony PCR for pET SUMO/ATII-ChSyn transformants in *E.coli* BL21 (DE3)

lane1: GeneRuler™ 1 kb Plus DNA ladder (Thermo Scientific). lane 2-7: 1455 positive amplicon using Forward SUMO and R_Downstream1 (lane 4 & 5 showed no bands). lane8: Negative Control.

- **Sample preparation of pET -28b+ expression system:**

The synthesized gene version of ATII-ChSyn was purchased as pET -28b+/ATII-ChSyn vector. It was transformed to *E.coli* Top 10 to maintain as glycerol stocks. Plasmids were purified to be transformed to *E.coli* BL21 (DE3). Colony PCR was done with the pET - 28b+ primers.
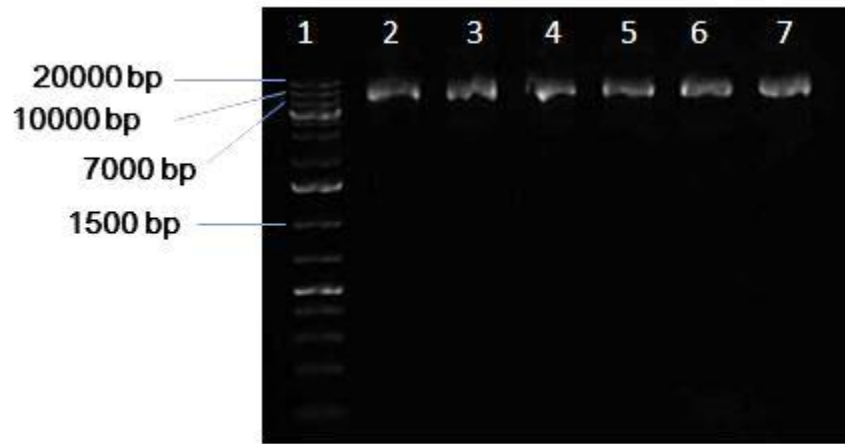

Figure (S 5) Plasmid extraction of pET -28b+/ATII-ChSyn

lane1: GeneRuler™ 1 kb Plus DNA ladder (Thermo Scientific) lane 2- 7: purified pET - 28b+/ATIIChSyn of total size 6422 bp.
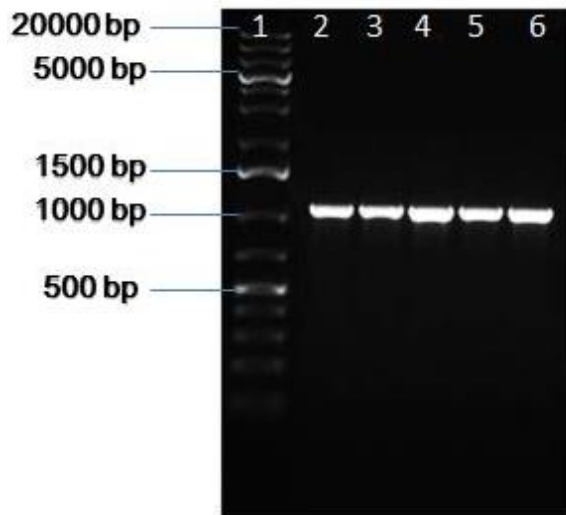
Figure (S 6) Colony PCR for pET -28b+ /ATII-ChSyn
lane1: GeneRuler™ 1 kb Plus DNA ladder (Thermo Scientific). lane2 – 6: 1199bp positive amplicon using T7 forward and T7 reverse primers.

Section B:

- **pET -28b+ with synthesized gene of ATII-ChSyn:**

- **Optimized Sequence Length:1058, GC%:57.04 with two restrictions sites NcoI, HindIII:**

CCATG
GCACAGGCTACGAAACTGATTAGTATTGGCACGGCGGTCCCGTCTCACATGATTGAACAACGCGACGCAGCAGCG
GTGGCACATCGTGCATTTAGCGATCGCTTTAAAGATTTCGACCGTCTGGCGAAGGTTTTCGAAAGCTCTGGCATT
CGTCGCCGTTATGCCATCCGCCCGCTGGAATGGTACCTGGAACCGCTGGGTTGGCCGGAACGTAACGCAGCATAT
ATTGAGGGTGCGTGCTCTCTGTTTGTCAGTGCAGCTCAGCAAGCCCTGGAAAAAGCAGACCTGCGCGGCGAAGAT
GTGGACGTGGTTGTCACCGTTAGTTCCACGGGTATTGCCACCCCGAGTCTGGAAGCACGTGTGGCTCATCAGATG
GGTTTTCGCGAAGATATCGAACGTGTCCCGGTGTTCGGTCTGGGTTGTGCTGGCGGTGTGTCAGGCTTTTCGATT
GCAAGCCGTTTCGCAGCAGCACGTCCGGGTAAAATCGTTCTGCTGGTTGCCGTCGAACTGTGCACCCTGGCATTT
CGCCTGGATCAGCTGACGAAAGCGAACATTGTTGCAACCGCACTGTTCGGTGACGGTGCTGCAGCCTGTATCCTG
AAGGCGGGCGATGCCGGTATTGCAAATATCGAAATGTCTGGCCAGCACACGTGGCCGGATACCCTGGACATTATG
GGTTGGAGTGTCGACAAACAAGGCTTTGGTGTGATTTTCGATCGCGCTATCCCGCCGTTTGCGGAAGAACGTATC
GCACCGGCTATTTCCGGCATCCTGAACCGCGCGGATCTGGCGATGGCCGATATTGACCGTTTCGCCTGCCATCCG
GGCGGTACGAAGGTGATCGCAGCTCTGGAAACCGCACTGTCCCTGGGTCAGGGTTCACTGGATCACGAACGTGAC
GTTCTGTCGGATTACGGCAATATGAGCGCTCCGACGGCGCTGTTTGTCCTGGATCGCGTGGTTCAAGCGGGTCTG
CCGTCTCGTACCCTGCTGACGGCACTGGGTCCGGGTTTCTCTGTCTCCTGTGTTTCCCTGCAACGCGCTGCT
AAGCTT

- DNA Alignment (red characters are optimized regions):

```
Optimized 6    GCACAGGCTACGAAACTGATTAGTATTGGCACGGCGGTCCCGTCTCACATGATTGAACAA
Original  6    GCTCAAGCTACGAAGCTGATCTCGATCGGAACAGCTGTTCCATCGCATATGATCGAACAG

Optimized 66   CGCGACGCAGCAGCGGTGGCACATCGTGCATTTAGCGATCGCTTTAAAGATTTCGACCGT
Original  66   CGAGATGCAGCCGCCGTAGCGCATCGCGCATTTTCGGATCGGTTCAAGGATTTTGATCGT

Optimized 126  CTGGCGAAGGTTTTCGAAAGCTCTGGCATTCGTCGCCGTTATGCCATCCGCCCGCTGGAA
Original  126  TTGGCAAAGGTGTTCGAGAGCTCGGGCATTCGTAGGCGCTATGCGATCAGGCCTCTTGAG

Optimized 186  TGGTACCTGGAACCGCTGGGGTTGGCCGGAACGTAACGCAGCATATATTGAGGGTGCGTGC
Original  186  TGGTATCTTGAACCGCTGGGGCTGGCCGGAACGAAATGCCGCCTATATCGAAGGGGCATGC

Optimized 246  TCTCTGTTTGTCAGTGCAGCTCAGCAAGCCCTGGAAAAAGCAGACCTGCGCGGCGAAGAT
Original  246  AGTTTGTTCGTCAGTGCGGCTCAGCAGGCGCTGGAGAAAGCGGACCTTCGAGGTGAAGAT

Optimized 306  GTGGACGTGGTTGTCACCGTTAGTTCCACGGGTATTGCCACCCCGAGTCTGGAAGCACGT
Original  306  GTCGATGTCGTCGTCACGGTTTCCTCAACAGGTATAGCCACGCCTAGTCTCGAAGCGCGC

Optimized 366  GTGGCTCATCAGATGGGTTTTCGCGAAGATATCGAACGTGTCCCGGTGTTCGGTCTGGGT
Original 366   GTCGCTCATCAAATGGGGTTTCGTGAAGACATCGAAAGGGTGCCTGTCTTCGGCCTTGGC

Optimized 426  TGTGCTGGCGGTGTGTCAGGCTTTTCGATTGCAAGCCGTTTCGCAGCAGCACGTCCGGGT
Original  426  TGTGCCGGCGGTGTCTCAGGATTCTCCATTGCATCTCGATTTGCAGCAGCGCGGCCCGGA

Optimized 486  AAAATCGTTCTGCTGGTTGCCGTCGAACTGTGCACCCTGGCATTTCGCCTGGATCAGCTG
Original  486  AAAATTGTCCTGCTGGTTGCAGTTGAACTGTGCACGCTCGCGTTCCGACTGGATCAATTG

Optimized 546  ACGAAAGCGAACATTGTTGCAACCGCACTGTTCGGTGACGGTGCTGCAGCCTGTATCCTG
Original  546  ACAAAGGCCAACATCGTCGCAACTGCGCTGTTCGGGGATGGGGCGGCCGCCTGCATTCTC

Optimized 606  AAGGCGGGCGATGCCGGTATTGCAAATATCGAAATGTCTGGCCAGCACACGTGGCCGGAT
Original  606  AAGGCGGGTGATGCAGGTATCGCAAACATCGAAATGTCCGGCCAGCATACTTGGCCTGAT

Optimized 666  ACCCTGGACATTATGGGTTGGAGTGTCGACAAACAAGGCTTTGGTGTGATTTTCGATCGC
Original  666  ACGCTCGACATCATGGGTTGGAGCGTTGATAAGCAGGGGTTCGGCGTTATTTTTGATCGA

Optimized 726  GCTATCCCGCCGTTTGCGGAAGAACGTATCGCACCGGCTATTTCCGGCATCCTGAACCGC
Original  726  GCAATACCGCCATTTGCTGAGGAACGCATTGCACCGGCTATATCGGGTATTTTAAACCGT

Optimized 786  GCGGATCTGGCGATGGCCGATATTGACCGTTTCGCCTGCCATCCGGGCGGTACGAAGGTG
Original  786  GCTGATCTTGCCATGGCAGATATCGACAGGTTTGCCTGTCATCCAGGTGGTACCAAGGTG

Optimized 846  ATCGCAGCTCTGGAAACCGCACTGTCCCTGGGTCAGGGTTCACTGGATCACGAACGTGAC
Original  846  ATTGCTGCTCTGGAGACAGCACTGTCGCTTGGTCAAGGCAGTTTGGACCATGAGCGTGAC

Optimized 906  GTTCTGTCGGATTACGGCAATATGAGCGCTCCGACGGCGCTGTTTGTCCTGGATCGCGTG
Original  906  GTGTTGTCCGATTATGGAAACATGTCGGCTCCAACCGCACTTTTTGTATTGGACCGTGTC

Optimized 966  GTTCAAGCGGGTCTGCCGTCTCGTACCCTGCTGACGGCACTGGGTCCGGGTTTCTCTGTC
Original  966  GTCCAAGCCGGCTTACCGTCACGTACCCTCCTTACTGCACTGGGACCGGGTTTTAGCGTG

Optimized 1026 TCCTGTGTTTCCCTGCAACGCGCTGCT
Original  1026 AGCTGTGTATCTCTCCAGAGAGCAGCA
```

- **Using solubilizing and detergent agents to purify inclusion bodies in pET - 28b+/ATIIChSyn:**
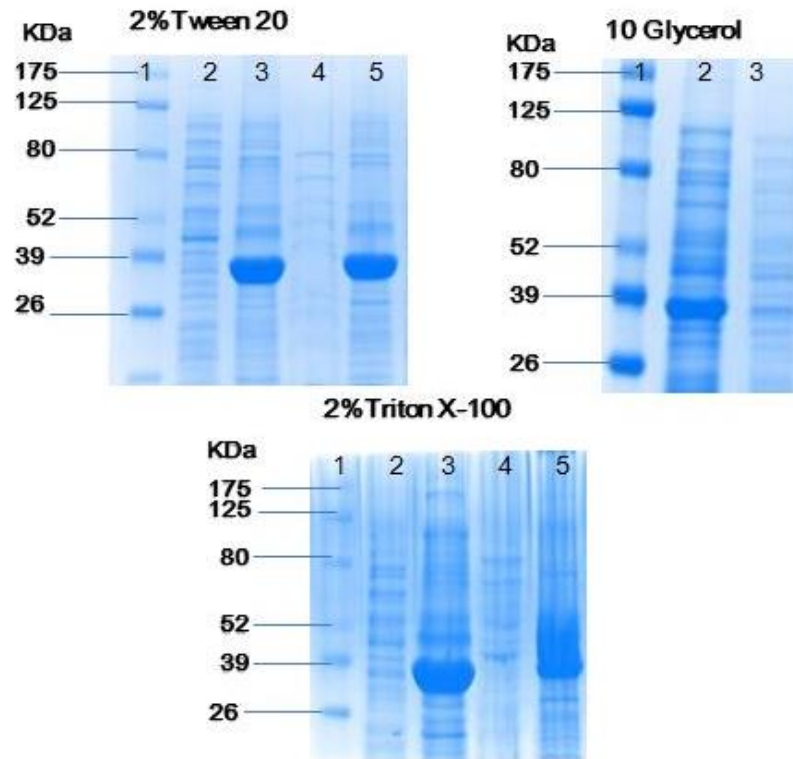

Figure (S 7) Localization of the expressed pET SUMO /ATIIChSyn after using detergents, 0.1mM IPTG for 1 hour at 37°C on 12 % SDS-PAGE

lane 1: ProSieve[®] Colour Protein Marker (Lonza), lane 2: Supernatant compartment without tween treatment, lane 3: Cell debris compartment without tween treatment, lane 4: Supernatant compartment with tween treatment, lane 5: Cell debris compartment with tween treatment. 10% Glycerol: lane 1: ProSieve[®] Colour Protein Marker (Lonza), lane 2: Cell debris compartment with 10% Glycerol treatment, lane 3: Supernatant compartment with 10% Glycerol treatment. 2% Triton X-100: Lane 1: ProSieve[®] Colour Protein Marker (Lonza), lane 2: Supernatant compartment without triton x-100 treatment, lane 3: Cell debris compartment without triton x-100 treatment, lane 4: Supernatant compartment with triton x-100 treatment, lane 5: Cell debris compartment with triton x-100 treatment.

- **Using mild detergent to purify inclusion bodies in pET -28b+ / ATII-ChSyn:**
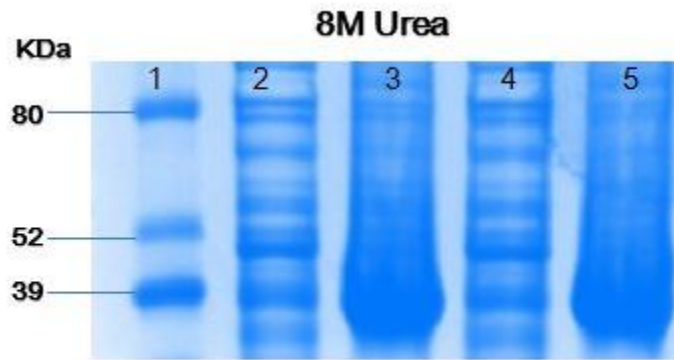

Figure (S 8) Localization of the expressed pET -28b+ /ATII-ChSyn after 8M Urea treatment for 1 hour and overnight with agitation at room temperature on 12 % SDS-PAGE

lane 1:ProSieve® Colour Protein Marker (Lonza), lane 2: Supernatant compartment after 8M Urea treatment for 1 hour, lane 3: Cell debris compartment after 8M Urea treatment for 1 hour. lane 4: Supernatant compartment after 8M Urea overnight treatment, lane 5: Cell debris compartment after 8M Urea overnight treatment.