





**Universidad de Cantabria**  
Departamento de Biología Molecular

**Instituto de Biomedicina y Biotecnología  
de Cantabria (IBBTEC)**



**Structural and biochemical  
characterization of proteins involved in  
the synthesis of triglycerides and  
polyunsaturated fatty acids**

Tesis Doctoral presentada por  
Omar Santín Martínez  
para optar al Grado de Doctor por la Universidad de Cantabria  
Santander, 2017



**Universidad de Cantabria**  
Departamento de Biología Molecular

**Instituto de Biomedicina y Biotecnología  
de Cantabria (IBBTEC)**



**Caracterización estructural y  
bioquímica de proteínas implicadas en  
la síntesis de triglicéridos y ácidos  
grasos poliinsaturados**

Tesis Doctoral presentada por  
Omar Santín Martínez  
para optar al Grado de Doctor por la Universidad de Cantabria  
Santander, 2017



**Don Gabriel Moncalian Montes**, Profesor Titular de Genética, perteneciente al  
Departamento de Biología Molecular de la Universidad de Cantabria,

**CERTIFICA:** Que Don. **Omar Santín Martínez** ha realizado bajo su dirección el trabajo que lleva por título “Caracterización estructural y bioquímica de proteínas implicadas en la síntesis de triglicéridos y ácidos grasos poliinsaturados”.

Considero que dicho trabajo se encuentra terminado y reúne los requisitos necesarios para su presentación como Memoria de Doctorado, al objeto de poder optar al grado de Doctor.

Santander, a 17 de Abril de 2017

Fdo. **Gabriel Moncalian Montes**



El presente trabajo ha sido realizado entre el Departamento de Biología Molecular de la Universidad de Cantabria y el Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), bajo la dirección del Dr. Gabriel Moncalián Montes, gracias a un contrato de la Universidad de Cantabria asociado a los proyectos "Aceites a la carta: bioingeniería de la síntesis de monoésteres y triglicéridos en microorganismos" (BIO2010-14809, del Ministerio de Ciencia e Innovación) y PLASMID OFFENSIVE: PROTEIN ENGINEERING OPERATIONS (BFU2014-55534-C2, Ministerio de Economía y Competitividad)

Durante este periodo se ha realizado una estancia de tres meses en el laboratorio del doctor Abel Baerga Ortiz (Universidad de Puerto Rico, Recinto de Ciencias Médicas)



*"Magic's just  
science that we don't  
understand yet"*

(Arthur C. Clarke)



## **Agradecimientos**

A mi jefe Gabriel por haber confiado en mí desde el principio y por darme la oportunidad de participar en este proyecto tan interesante. Sus consejos y son el pilar fundamental de este trabajo.

A toda la gente del laboratorio de genética. Solamente podría tener buenas palabras y elogios para todos. Gente entregada e inteligente de los que he aprendido un montón.

A todos los jefazos que han dedicado su tiempo a escuchar mis seminarios e intentar corregir mis errores de novato cabezudo: Fernando de la Cruz, Elena, Iñaki... Mil gracias por vuestra ayuda y apoyo.

A todos los post-docs del grupo de Fernando por tener siempre buenos consejos... Mapi, Ruli, Athanasia... Muchas gracias.

A los técnicos Mati, Sandra, Victor, Raquel, Carlos. Gracias por todo. Sin vosotros el edificio se cae!

A mis compañeras Lorena, Laura, Sandra y Bea y en general todos los que han pasado por el labo 02.08. Con vosotros he compartido las mejores discusiones constructivas y destructivas (mis favoritas). Sin vosotros este trabajo no podría haber existido. Gracias por aguantarme y ayudarme estos años.

A Serena, Carla y "Dj Jaime" por haber aguantado mis aburridas y obstinadas explicaciones y dudas sin resolver. Me habéis ayudado un montón.

Al doctor Abel Baerga y sus doctorandos Melisa, Carlos, Ramón, Yermay, Vilma y en general a todo Puerto Rico por hacerme sentir como en mi casa en mi estancia. Tenéis una isla preciosa y un corazón todavía mejor. Gracias a todos y espero veros pronto.

A los antiguos que dejaron huella, y nos permitieron subirnos a hombros de gigantes. Juan Villa por tu legado tesil. Jorge por hacerme ser el gallego segundo. Maria y Coral por su ancestral devoción a los "quesos" de tetilla. Alejandro por su pasioff genial. David porque se te quería aunque fueras fan de loquillo. Maris por ser la última de las antiguas, ayudarme y maldecir juntos el fastidioso Word, y Lillo por ayudarnos con tu lillopedia siempre que lo necesitamos ¡que esto se acaba!

Y los de las hornadas siguientes, mi ex-compañero de piso más coyote, Jorge, gracias por todo tío! A Carol por todas las risas nocturnas y diurnas y ser la santanderina más maja. A Antonio por rastrear tan bien, eres un genio, seguro que tenías un script. A Mark por esos guitarreos y por invitarme a tu único concierto en España, y yo quedarme en casa. A Emilio por tener siempre más risas que todos juntos y una perra tan estupenda, nunca cambies. A Jusep por tener siempre un comentario gracioso en el momento más insospechado. Al Litri por haber sido mi

mejor compi de labo aunque durase poco. A Marta y Cris por ser las andaluzas más majetas! A Lolu porque... ¿pensabas que me olvidaba de ti? A Silvia por liarla tantísimo en el Niágara, estas fatal! A Álvaro por gritar ¡Señor! conmigo en aquel concierto de reggae. A Yago por engordarnos con sus cupcakes. A Iñaki de los piero por compartir conciertos, charlas políticas y risas. A Ruth por sorprenderme siempre con sus historias raras. A Laurita por tener los abdominales de Swarzeneger y ser tan maja. A Yoli por ser la más punki del IBBTEC. Gracias a todos.

A mis amigos de Galicia que rápido y corriendo compartieron cervezas y risas en mis vueltas fugaces a la tierra. El poco tiempo que nos dedicamos estos años ha sido muy importante para mí. Antón con sus geocachings. Sike con sus rapeos. Rubens y Javi por venir con el a visitarme a Santander. Ska con sus madrugones para rapalear. Tarta y Kaze con sus ratos de estudio y zulo....Los churrasketes con Quirón, Lucia, Marcos, Marta, María, Champi, Patri, Arian, Chipi, Isma. El baile del robot a altas horas con Iñaki y Quinti. Y a Sonia porque cada vez que paso por Santiago me entra la morriña. Por supuesto a ti Marea por todos los momentos buenos! A mis primeros colegas cántabros Javi Guti y Rober. Todos unos putos jefes. Os quiero mazo.

A Lúa por su genialidad. Por arrastrarme a la locura transitoria. Por sacar lo mejor de mí. Por enseñarme a pintar marcianos. Por construir juntos el término “maquinita”. Gracias por tu compañía incondicional. Algún día te podré recompensar!? Te quiero!

A mi familia entera por su apoyo y por portarse tan bien conmigo siempre desde pequeño. Mi papá y mamá por unos padres increíbles y haberme dejado ser libre. A mi hermana por ser cada día más crack. A mis abuelos Tito, Marisa, Víctor y Aurora por ser los mejores abuelos que puedo tener. A mis tios Carlos, Viky, Juan, Ana, Malena, Tin, Lidia, y los tios de Lugo Paco y Elena y la prima Alba. A mi primo Victor por saberse más nombres científicos que yo.

Y no me puedo olvidar de Alexandra Elbakyan, por permitir que la información obtenida con dinero público sea de dominio público.

A todos de los que me he olvidado. Seguro sois un montón. GRACIAS.

Omar





# Index

---

<b>Abbreviations</b> .....	23
<b>1. Introduction</b> .....	27
<b>1.1. Commercial Interest of Added-Value lipids</b> .....	27
1.1.1. Triglycerides .....	28
1.1.2. Omega-3 fatty acids .....	31
<b>1.2. Triglyceride synthesis in bacteria</b> .....	34
1.2.1. Lipid Reservoirs in Bacteria .....	34
1.2.2. Mechanism of triglyceride synthesis: The WS/DGAT family .....	36
1.2.3. WS/DGAT proteins and formation of lipid bodies .....	40
1.2.4. Biotechnological use of WS/DGAT enzymes: tDGAT.....	41
<b>1.3. Fatty acid synthesis</b> .....	42
1.3.1. The fatty acid synthase systems I and II.....	43
1.3.2. Overall mechanism of fatty acid synthesis.....	44
<b>1.4. Polyketide synthesis</b> .....	47
1.4.1. PKS genetic organization .....	48
1.4.2. PKS type I: A vectorial synthesis.....	49
1.4.3. PKS type II: An iterative synthesis .....	50
<b>1.5. Structural biology of FAS and PKS domains</b> .....	52
1.5.1. Megasyntases and big complex structure.....	53
1.5.2. The Acyl carrier Protein.....	54
1.5.3. The condensing domains: KS and AT.....	56
1.5.3.1. The keto synthase domain .....	57
1.5.3.2. The Acyl transferase domain.....	58
1.5.3.3. The Chain length factor domain.....	60
1.5.4. The modifying domains.....	63
1.5.4.1. The Keto Reductase domain .....	63
1.5.4.2. The Dehydratase/Isomerase domain .....	65
1.5.4.3. The Enoyl Reductase domain .....	66
<b>1.6. Omega-3 Fatty Acid synthesis</b> .....	68
1.6.1. The aerobic pathway: desaturases and elongases.....	68
1.6.2. The anaerobic pathway: pfa synthases.....	70
<b>2. Aims and Scope</b> .....	77
<b>3. Materials and Methods</b> .....	81

<b>3.1. Bacterial Strains</b>	81
<b>3.2. Microbiological techniques</b>	81
3.2.1. General culture conditions	81
3.2.2. Culture conditions for neutral lipid production for TLC analysis	81
3.2.3. Culture conditions in microtiter plates	81
3.2.4. Bacterial glycerol stock	82
3.2.5. Preparation of competent cells	82
3.2.6. Electroporation transformation	82
<b>3.3. Oligonucleotides</b>	82
<b>3.4. Plasmids</b>	85
<b>3.5. DNA engineering</b>	85
3.5.1. DNA extraction and purification	85
3.5.2. Polymerase chain reaction (PCR)	85
3.5.3. Agarose electrophoresis	86
3.5.4. Restriction enzymes cloning	86
3.5.5. Isothermal assembly cloning	87
3.5.6. DNA sequencing	88
3.5.7. DNA quantification	88
<b>3.6. Protein purification techniques</b>	88
3.6.1. Recombinant protein production in <i>E. coli</i>	88
3.6.2. Protein purification	89
3.6.2.1. Nickel column affinity chromatography	89
3.6.2.2. Ion-exchange chromatography	89
3.6.2.3. Size Exclusion Chromatography	89
3.6.3. Buffer exchange	90
3.6.4. Protein electrophoresis	90
3.6.5. Protein quantification	90
3.6.6. Mass spectrometry analysis	90
<b>3.7. Directed evolution and protein engineering</b>	91
3.7.1. Error prone PCR	91
3.7.2. Site directed mutagenesis	91
3.7.3. Staggered extension process (StEP) in vitro recombination	92
3.7.4. Megawhop PCR	93
<b>3.8. Biochemical assays</b>	94
3.8.1. Lipid extraction and thin layer chromatography	94
3.8.2. Fluorescence-based lipid detection	94

3.8.3.	Radioactivity assays: Substrate-protein complexes .....	95
<b>3.9.</b>	<b>Biophysical methods</b> .....	95
3.9.1.	Ultracentrifugation analysis of tDGAT-TAG lipoprotein complexes.....	95
<b>3.10.</b>	<b>X-ray Crystalization</b> .....	95
3.10.1.	Crystallographic data collection .....	96
3.10.2.	Crystallographic data processing .....	96
3.10.3.	Molecular replacement.....	96
3.10.4.	Model building and refinement .....	97
<b>3.11.</b>	<b>Bioinformatic tools</b> .....	97
3.11.1.	Densitometry analysis of TLC .....	97
3.11.2.	Structural modeling.....	97
3.11.3.	Bioinformatic softwares .....	98
<b>4.</b>	<b>Results and Discussion</b> .....	101
<b>4.1.</b>	<b>Optimization of TAG production in E. coli</b> .....	101
4.1.1.	Directed evolution.....	101
4.1.1.1.	Evolving tDGAT using a fluorescence-based method .....	101
4.1.1.2.	tDGAT mutants with increased in vivo activity map at protein surface .....	103
4.1.1.3.	tDGAT mutants with monofunctional synthesis were found.....	105
4.1.1.4.	In vitro recombination of positive did not improved TAG accumulation .....	108
4.1.1.5.	Quantification by densitometric analysis of TAG accumulation .....	109
4.1.1.6.	Solubility of highly productive tDGAT mutants has not change .....	111
4.1.2.	Ultracentrifugation analysis of tDGAT-TAG lipoprotein complexes.....	111
4.1.3.	General Discussion .....	113
<b>4.2.</b>	<b>Structural and Biochemical analysis of PUFA synthases</b> .....	116
4.2.1.	<i>M. marina</i> Pfa functional domains.....	116
4.2.1.1.	PfaA is formed by ten individual domains that show five different folds.....	117
4.2.1.2.	PfaB has an acyltransferase and a pseudo keto synthase domain .....	126
4.2.1.3.	PfaC has two keto synthase and four dehydratase domains.....	128
4.2.1.4.	PfaD is a enoyl reductase .....	133
4.2.2.	<i>pfa</i> constructions design .....	135
4.2.3.	Biochemical analysis.....	139
4.2.3.1.	ACPs are self-acylated with malonylCoA but not acetylCoA.....	139
4.2.3.2.	PfaA KS-AT promotes malonyl-CoA binding to ACP .....	141
4.2.3.3.	PfaA KS-AT S707A promotes acetyl-CoA binding to ACP .....	143
4.2.3.4.	PfaB is able to charge ACPs with malonyl-CoA .....	144
4.2.3.5.	PfaC KS-CLF domain does not bind malonyl-CoA nor acetyl-CoA .....	145

4.2.4.	Structural analysis of PfaC KS-CLF domain .....	147
4.2.4.1.	Crystallization process.....	147
4.2.4.2.	Data collection and processing .....	148
4.2.4.3.	Overall structure of KS-CLF domain of PfaC.....	149
4.2.4.4.	Helical bundle fold could have structural implications.....	152
4.2.4.5.	KS-CLF C-terminal domain might have decarboxylation activity .....	153
4.2.4.6.	KS-CLF has a hydrophobic tunnel that connects both active centers.....	155
4.2.5.	General Discussion .....	158
<b>5.</b>	<b>Resumen en Castellano</b> .....	<b>165</b>
<b>5.1.</b>	<b>Introducción</b> .....	<b>165</b>
<b>5.2.</b>	<b>Resultados y Discusión</b> .....	<b>167</b>
5.2.1.	Evolución dirigida de tDGAT.....	167
5.2.2.	Estudio bioquímico y estructural de sintasas de ácidos grasos omega-3 .....	168
<b>6.</b>	<b>References</b> .....	<b>173</b>
<b>7.</b>	<b>Publications and Conferences</b> .....	<b>195</b>
<b>7.1</b>	<b>Scientific papers</b> .....	<b>195</b>
<b>7.2.</b>	<b>Scientific posters</b> .....	<b>195</b>
<b>8.</b>	<b>Supplementary Material</b> .....	<b>199</b>



---

# Abbreviations

---

<b>ACP</b>	Acyl carrier protein	<b>LC</b>	Long chain
<b>ALA</b>	$\alpha$ -Linolenic acid	<b>LD</b>	Lipid droplet
<b>ARA</b>	Arachidonic acid	<b>MAT</b>	Malonyl acyltransferase
<b>AT</b>	Acyl transferase	<b>NADPH</b>	Nicotinamideadenine dinucleotide phosphate
<b>Amp</b>	Ampiciline	<b>NMR</b>	Nuclear magnetic resonance
<b>CLF</b>	Chain length factor	<b>ORF</b>	Open reading frame
<b>CoA</b>	Coenzyme-A	<b>PHA</b>	Polyhydroxyalkanoate
<b>DAG</b>	Diacylglycerol	<b>PKS</b>	Polyketide synthase
<b>DEBS</b>	6-Deoxyerythronolide B synthase	<b>PMSF</b>	Phenylmethylsulfonyl fluoride
<b>DGAT</b>	Diacylglycerol acyltransferase	<b>PPT</b>	4-phosphopantetheine
<b>DH</b>	Dehydrogenase	<b>PPTase</b>	Phosphopantetheinyl transferase
<b>DHA</b>	Docosahexaenoic acid	<b>PUFA</b>	Polyunsaturated fatty acids
<b>DNA</b>	Deoxyribonucleic acid	<b>SAXS</b>	Small-angle X-ray scattering
<b>EPA</b>	Eicosapentaenoic acid	<b>SDR</b>	Short-chain dehydrogenase/reductase
<b>ER</b>	Enoyl reductase	<b>SDS-PAGE</b>	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
<b>FA</b>	Fatty acid	<b>TAG</b>	Triacylglycerol
<b>FAEE</b>	Fatty acid ethyl ester	<b>tDGAT</b>	WS/DGAT from <i>Thermomonospora curvata</i>
<b>FAME</b>	Fatty acid methyl ester	<b>TE</b>	Thioesterase
<b>FAS</b>	Fatty acid synthase	<b>WE</b>	Wax ester
<b>iPKS</b>	Iterative polyketide synthase	<b>WS/DGAT</b>	Wax ester synthase/acyl-CoA: diacylglycerol acyltransferase
<b>IPTG</b>	Isopropil- $\beta$ -D-1 tiogalactopiranósido	<b>Wt</b>	Wyld type
<b>kDa</b>	Kilodalton		
<b>KR</b>	Keto reductase		
<b>KS</b>	Keto synthase		
<b>Kn</b>	Kanamycin		



## **Introduction**

---



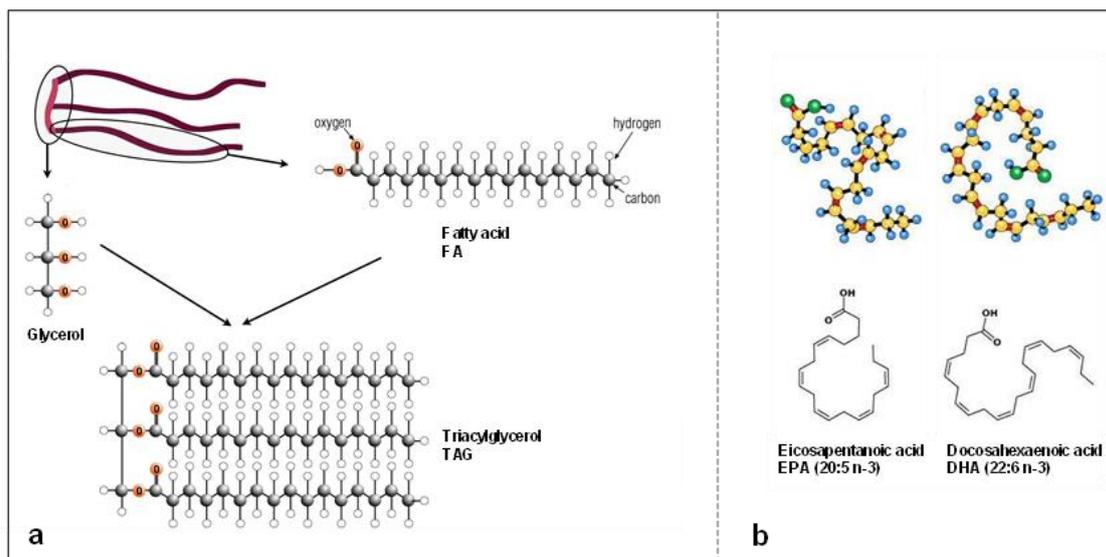
# 1. Introduction

---

## 1.1. Commercial Interest of Added-Value lipids

Lipids are molecules that are found in all living cells and contain hydrocarbons chains that make the building blocks of some important structures as the cellular membranes. Being non-polar, lipids are not soluble in water, and for that reason lipids that form cell membranes create fundamental barriers separating the aqueous cellular phase from the external complex environments. Lipids are highly reduced so they are very energetic and for that reason living organisms use them as the preferred long term form of stored energy, as their low oxidation rates and high calorific values make them release more energy upon oxidation than carbohydrates and proteins (Berg, Tymoczko, and Stryer 2002a). Other very important function of lipids, especially in higher eukaryotic organisms, is to act as signaling molecules as in the case of lipophilic hormones (Pucci, Chiovato, and Pinchera 2000). Thus, it is easy to understand the immense commercial importance of these molecules in the global market. Within the next paragraphs the biological and commercial value of some of these interesting molecules will be investigated.

One of the most common lipids that have been widely studied for being present in all life forms are fatty acids (FA). They are carboxylic acids with a long aliphatic chain, which normally have an unbranched chain of an even number of carbon atoms, from four to twenty-eight, due to some peculiarities of the synthesis mechanism which will be explained in detail later. According to the number of carbons, fatty acid are usually classified into three groups: short-chain FA when they have less than 6 carbons, medium-chain FA if they have from 6-10 carbons and long-chain FA when they have more than 12 carbons. Fatty acids can be saturated (without double bonds), unsaturated (with one double bond) or polyunsaturated (with more than 1 double bond) (Chow 1999). The more unsaturated is a FA, the less the melting temperature it has, which impacts on its biological properties. The position of the double bond nearest the methyl end ( $\text{CH}_3$ ) of the carbon chain is used to classified the polyunsaturated fatty acids (PUFAs) into omega-3 FA, when the PUFA has its first double bond 3 carbons away from the methyl end, and omega-6 FA when it has it 6 carbons away. Double bond can also occur in a cis or trans configuration (Keweloh and Heipieper 1996). A cis configuration means that the two hydrogen atoms adjacent to the double bond stick out on the same side of the chain and a trans configuration, by contrast, means that the adjacent two hydrogen atoms lie on opposite sides of the chain. Most biological fatty acids are in the cis configuration and normally trans fatty acids arise when they are subjected to hydrogenation processes in industrial processes. Fatty acids can be found in cells in the form of triglycerides (TAG) as the main form of energy storage. By keeping the FA into TAG molecules the cell undergoes osmotic pressure changes and avoids toxicities both normally associated with the existence of free fatty acids.

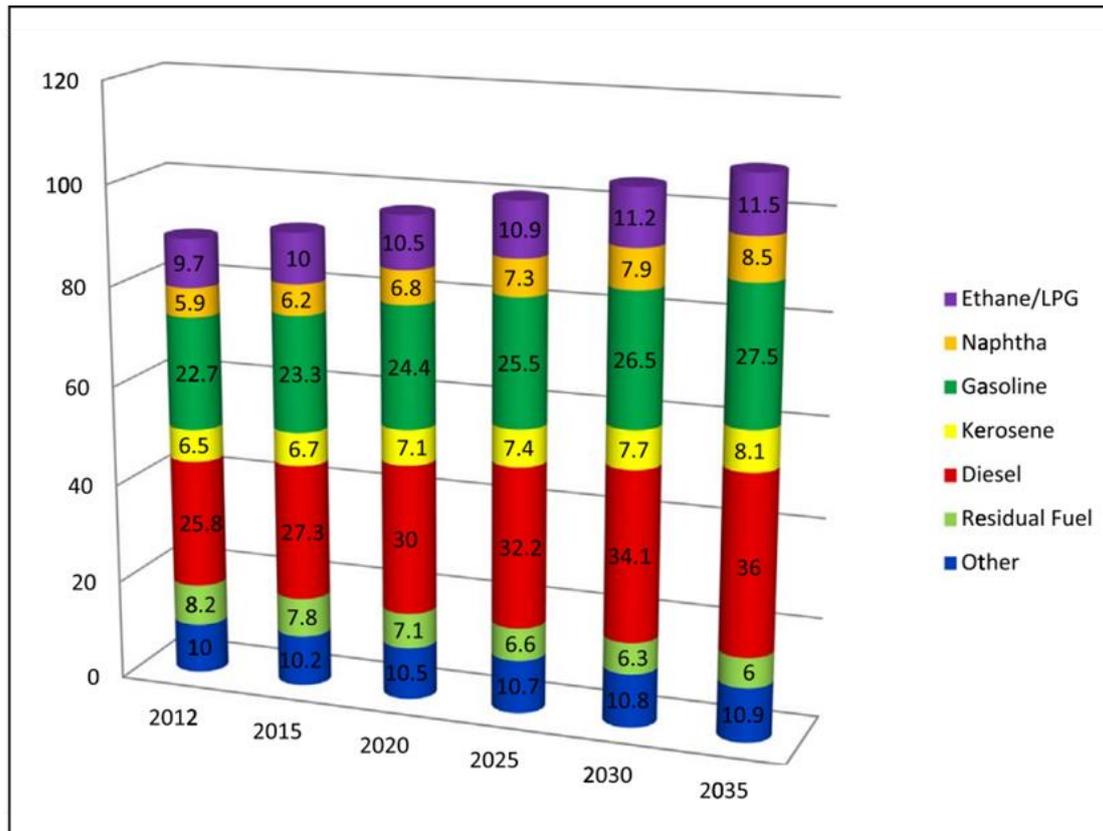


**Figure I-1.** Schematic representation of the molecular structures of triglycerides (a) and the most representative polyunsaturated fatty acids DHA and EPA (b)

Molecular representations of both TAG and PUFAs are shown in figure I-1. These two types of lipids have high economic impact and strong scientific relevance that will be discussed in this section. TAGs will be studied as very interesting reservoirs of fatty acids that could be used to produce biofuels (Ma and Hanna 1999) and PUFAs as powerful molecules from a medical and nutritional point of view (Tapiero et al. 2002).

### 1.1.1. Triglycerides

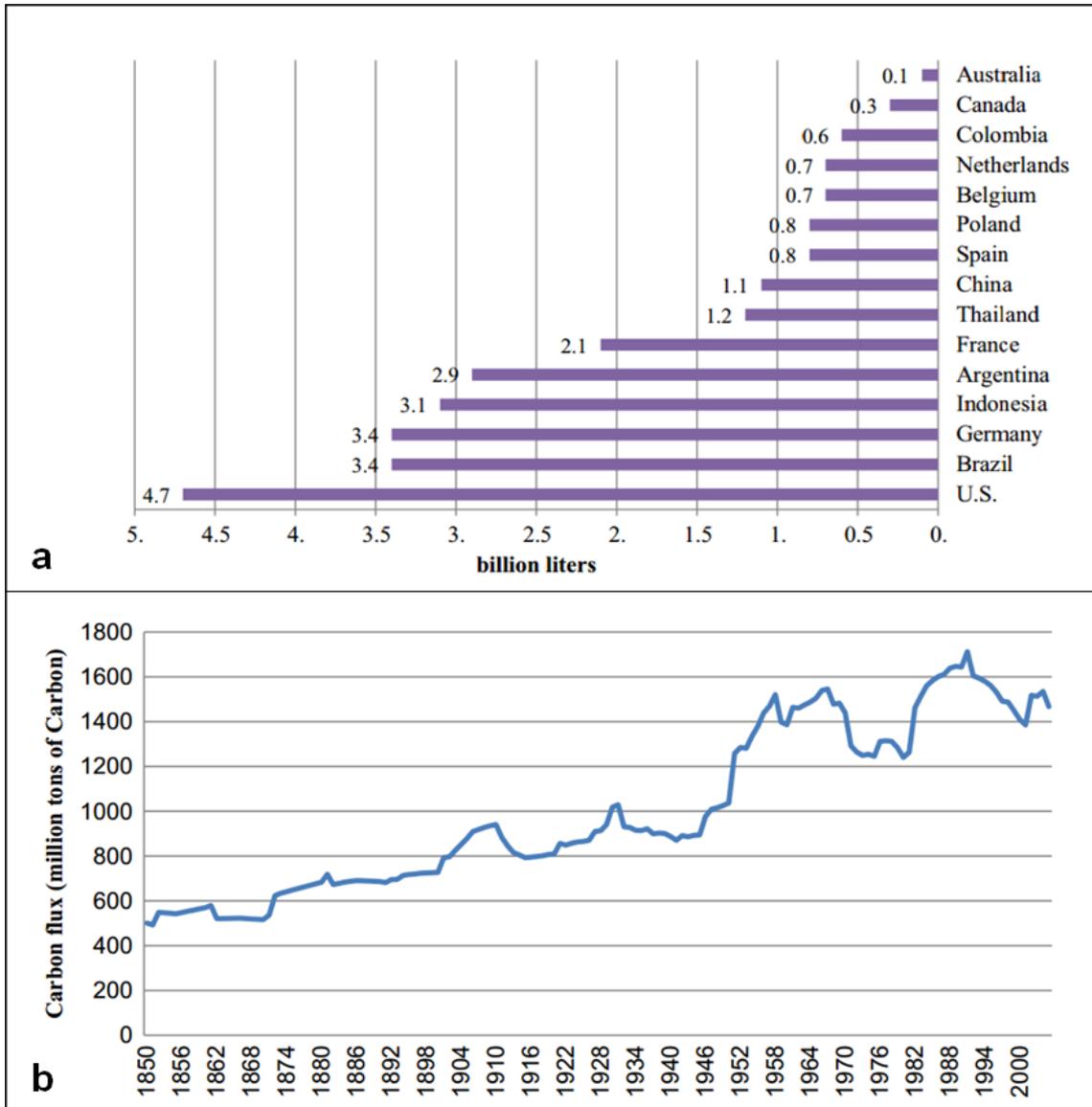
Normally, in living organisms, fatty acids are present as phospholipids in the cell membranes and accumulated as part of more complex molecules called triglycerides (TAG) which are esters derived from glycerol and three fatty acids. They are the main constituents of body fat in humans (Wertheimer and Shapiro 1948) and plants oils (Gurr et al., 1974), as well as forming lipid bodies in some particular bacterial species (Alvarez and Steinbüchel 2002). Besides constituting the primary source of FA in cells, TAG have always been used to manufacture many types of interesting products. They are used for cooking (olive, sunflower and palm oils), as artificial coatings for food preservation, as waterproof skin and garments, for cosmetic purposes, as food supplements, and more recently for the production of biofuels (Balat 2009; Elder 1984; Fox and Stachowiak 2007). TAG have been a well-appreciated good throughout history, mainly harvested from plant biomass of crops and animal fats from the livestock industry. In the last years, production of the major vegetable oils has risen more than 70%, mainly due to increases in palm, soybean, and rapeseed oil (Metzger and Meier 2011) which makes these biological resources important from an economic and social point of view since they are the most consumed.



**Figure I-2.** Estimates of the global demand for petroleum-based fuels from 2012 to 2035 (in million barrels/day). The growth in diesel demand is the most pronounced of all of them. Figure adapted from (Hajjari, Tabatabaei, Aghbashlo, & Ghanavati, 2017)

In recent decades there has been a concern about the increasing industrialization and motorization of the world that has led to a steep rise for the demand of petroleum-based fuels (Agarwal 2007). An example of this increase in demand is given in figure I-2 where a clear increase in diesel demand can be observed both in recent years and in forecast for the near future. Some studies point out that fossil fuels take up at least 80% of the primary energy consumed in the world, of which more than 50% alone is consumed by the transport sector (Escobar et al. 2009). There are a lot of evidences that clearly point out the negative environmental consequences of using fossil fuels that have spurred the search for renewable biofuels. The ease of manufacturing biodiesel using vegetable oils has made it a promising near-term alternative to fossil fuels, and for that reason, crop areas that were dedicated for human supply were progressively replaced to grow plants for biodiesel production. Some of the largest fossil-fuel consuming regions, such as the United States and Europe, have established higher biofuel targets that amount to at least 10 percent of transportation fuel by 2020. If such targets were to go global by 2050, meeting them would consume crops with an energy content equivalent to roughly 30 percent of the energy in today's global crop production. Consequently, the amount of crops needed to sustain this production would increase making the ideal of a sustainable agriculture even more difficult to achieve (ircsofia 2015). It seems obvious the unsustainability of these practices in the long run due to the limited space for cultivation and the rapid growth of human populations. It is needed a suitable hydrocarbon source for biodiesel production that can be producible in large quantities without reducing human food supplies. The next figure I-3 shows the biodiesel

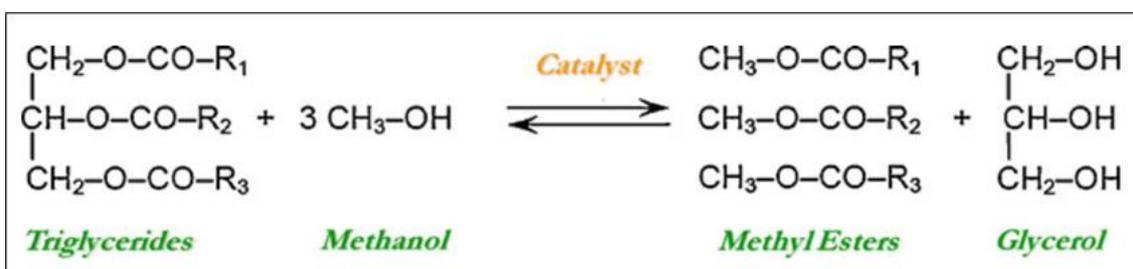
production levels of the leading countries and the associated increase of land use for biodiesel production.



**Figure I-3.** Global impact of biodiesel production through the use of land resources. Panel a show the world's biggest biodiesel producers in the year 2014, by country (in billion L). In the panel b is represented the annual net flux of carbon to the atmosphere as a result of land use change: 1850–2005. Figure adapted from (Hajjari, Tabatabaei, Aghbashlo, & Ghanavati, 2017)

Biodiesel is a mixture of fatty acid methyl esters (FAMES) produced from animal fats or vegetable oils. The transesterification from FA is the common method used for its production. This procedure is based on using short-chain alcohols, typically methanol, that react with TAGs to produce fatty acid methyl esters (FAMES) and glycerol (Fukuda, Kondo, and Noda 2001). Figure I-4 shows a scheme of a regular transesterification with methanol. Nowadays the growing of oil-enriched crops on a large scale is the established method for biodiesel production (Balat 2009; Ma and Hanna 1999), but as it was discussed in the last paragraph it causes some economic and ethical problems (Chu and

Majumdar, 2012). Biodiesels can be derived from conventional petrol, oilseeds or animal fat, but its production cannot meet realistic needs with these sources in actual economy and can only satisfy a small fraction of the existing demand. Thus, there is an emergence to find new sources and processes for the manufacture of modern biofuels.



**Figure I-4.** Triglycerides transesterification reaction with methanol. Figure taken from (Borges & Díaz, 2012)

Some natural unicellular organisms such as some microalgae, bacteria, fungi and yeasts are capable of produce and accumulate a variety of oils without any genetic modification. For that reason they have been extensively used as models in laboratories to try to generate an alternative source of biodiesel (Meng et al. 2009). It is believed that in the near future there will be techniques that use some of these micro-organisms to give rise to a “green replacement” for conventional petroleum diesel, due to its low cost, the manageability of their cell cultures and their long-term sustainability. The use of an alternative system for biodiesel production, like the ones based on culturing microorganism, will avoid the massive use of fertile and productive lands that currently increase the price of the human food. To date, lipid-production systems that use unicellular organisms that could be grown at a low cost in a very limited space are the most viable alternative to produce oils in a sustainable way that ensure the supply of future generations (Ochsenreither et al. 2016).

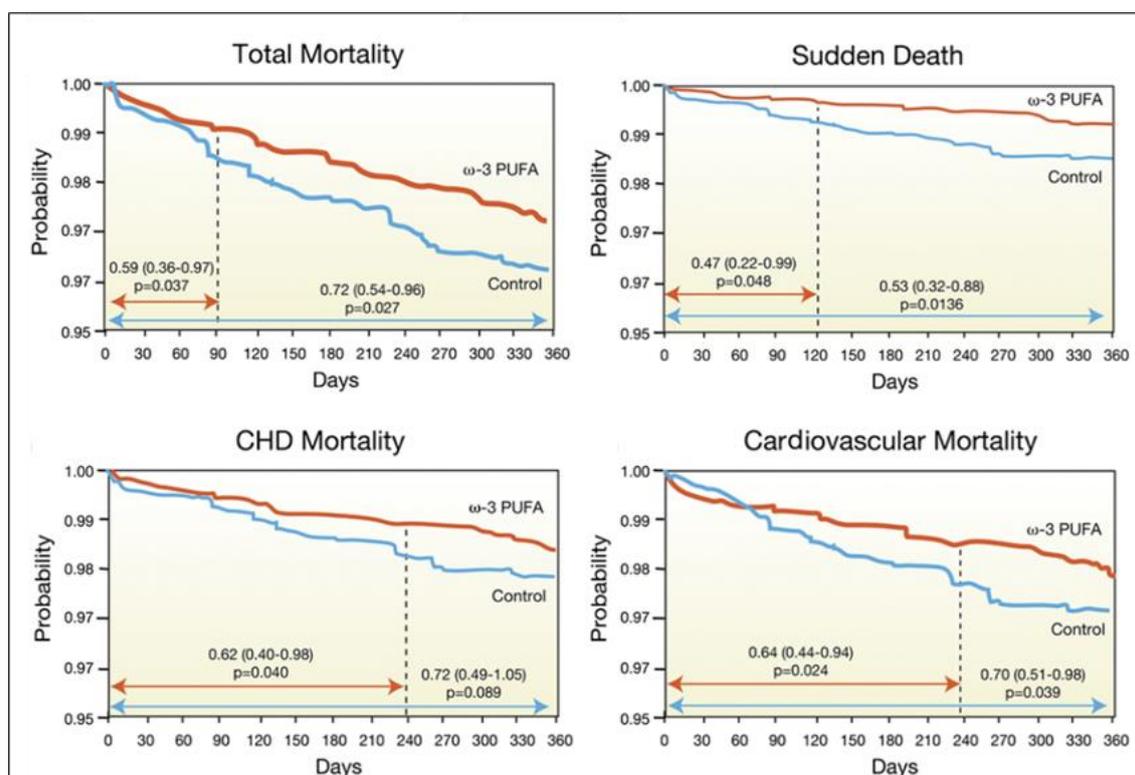
*Escherichia coli* is a model organism that has the potential to be easily cultured and modified using molecular biology and synthetic biology techniques in the laboratory. This bacteria can use different carbon sources to grow like hexose and pentose sugars such as glucose, xylose, arabinose and galactose that can be obtained from lignocellulosic biomass and other industrial waste substances. *E. coli* was engineered in the last decade to produce ethanol, triglycerides and fatty acids which can be directly used to produce biofuels, or even for directly produce biodiesel bypassing the downstream processing steps (Nawabi et al. 2011; Sherkhonov et al. 2016). In our laboratory, we have recently published an article in which we overexpress a protein of a thermophilic bacteria in *E. coli* achieving a rapid TAG accumulation (Lázaro et al., 2017, *in press*). One of the goals in this work was to improve the function of this metabolic system using direct evolution techniques to manipulate the behavior of the protein in its new host and therefore be able to improve the TAG production.

### 1.1.2. Omega-3 fatty acids

Omega-3 fatty acids are polyunsaturated fatty acids (PUFAs) that are named in this way because their last double bond is located between the third and fourth carbons starting from the FA methyl end. They have a regular aliphatic chain and differ in chain length, number and position of double bonds, which are all of them in a cis configuration.

The most relevant PUFAs in nature are known as arachidonic acid (ARA), an omega-6 FA, and the omega-3 FAs eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). These relevant PUFAs are composed by chains of 20 (ARA and EPA) and 22 carbons (DHA) with 4 (ARA), 5 (EPA) and 6 (DHA) unsaturations per molecule, and for that reason they are usually noted as 20:4, 20:5 and 22:6 respectively, where the first number indicates the carbon chain length and the second one indicates the number of unsaturations. For being longer than 12 carbons, these FA are classified into the long-chain polyunsaturated fatty acids (LC-PUFAs) group.

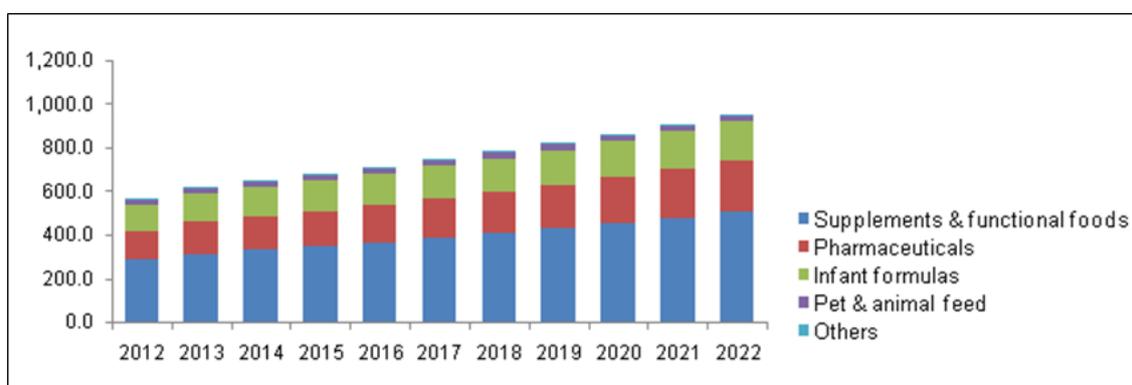
PUFAs are present in a small proportion in higher organisms like mammals, but they are very important from the physiological point of view because their implications in numerous important biological processes. EPA and DHA have long been known as essential membrane components in the brain and retina as well as precursors of signaling molecules such as prostaglandins, thromboxanes and leukotrienes (Tapiero et al. 2002). These omega-3 FA stimulate growth and exert a protective effect on the development of the cardiovascular system, reduce inflammatory symptoms in diseases like rheumatoid arthritis, ulcerative colitis or atopic dermatitis and psoriasis (Wall et al. 2010). EPA and DHA are also associated with the suppression of various diseases such as Alzheimer's disease, allergic diseases and cancer (Cole et al. 2005; Rose and Connolly 1999). It has been shown that DHA deficiency is associated with abnormalities in brain function probably due to the high accumulation of this type of FA in certain areas of the brain such as the frontal cortex (Ruxton et al. 2004). Figure I-5 shows some of the early benefits of PUFA ingestion.



**Figure I-5.** Early benefits of omega-3 polyunsaturated fatty acid (PUFA) ingestion on total mortality, sudden death, coronary heart disease (CHD) mortality, and cardiovascular mortality. Probability measurements represent relative risk (95% confidence interval). Figure adapted from (Lavie, Milani, Mehra, & Ventura, 2009)

PUFAs can arise in nature by more than one biosynthetic route, but most organisms produce them by desaturation and elongation of shorter saturated FA. Eukaryotic organisms possess special desaturases that introduce *cis* double bonds into preexisting saturated FA to synthesize PUFAs (Pereira, Leonard, and Mukerji 2003). For example, within the human metabolism,  $\alpha$ -linolenic acid (ALA) (18:3) is obtained from plant material in the diet and processed in some tissues to form EPA and DHA that, as was explained before, are essential for the correct development and function of the body. Terrestrial plants can produce small to moderate amounts of ALA, but the mammalian system cannot synthesize it. Hence, ALA deficiency could lead to EPA and DHA deficiency. Generally, mammalian systems cannot elongate and desaturate enough ALA into EPA and/or DHA upon consumption because the desaturase enzyme that is involved in this conversion competitively transform other substrates besides ALA (Plourde and Cunnane 2007). Thus, humans have to directly intake ALA, EPA and/or DHA, to obtain physiological effects of them. Many countries recommend daily EPA and DHA intakes of 0.3–0.5 g.

The major sources of EPA and DHA in nature are marine products. It has been described that microalgae and some proteobacteria that live in the depth sea are able to produce PUFAs (Adarme-Vega et al. 2012; Adarme-Vega, Thomas-Hall, and Schenk 2014). As a consequence, marine lipids generally contain a wider range of these FA than their terrestrial counterparts including both plants and animals. Vegetable oils usually consist of saturated and unsaturated fatty acids with 16 and 18 carbon chain lengths, while marine oils mainly consist of C14-C22 FA. From these marine microenvironmental niches, PUFAs enter the trophic chains and reach organisms like fishes that human can easily obtain to include them in the diet. Marine bacteria of *Colwellia*, *Moritella*, *Shewanella* and *Photobacterium* genus are a source of omega-3 FA, because they use biosynthetic pathways that do not require desaturation and elongation of FA precursors. They carry genes termed as *pfa* clusters coding for PUFA synthases that have very similar domain organization to the bacterial polyketide synthases (PKS) and the fatty acid synthases (FAS). The study of these recently discovered proteins is an area of great interest because of the relevance they have for the maintenance of omega 3 levels in aquatic ecosystems. These system will be revised in an specific subsection of this introduction (1.6.2. The anaerobic pathway: *pfa* synthases).



**Figure I-6.** EPA and DHA-based product market size estimation for the next 5 years. This market is expected to gain at over 8% to surpass USD 4 billion sales by 2022. Figure taken from (Omega 3) EPA/DHA Ingredients Market Size, Share Report 2022.

Currently the main source of EPA and DHA in the human diet is sea-fish oil that is obtained from direct ingestion of oil-rich fishes like tuna or salmon or using food supplements which are in turn also made with fish. Because of the limited sources of omega-3 in nature, there is a problem of overexploitation of fish stocks in the oceans (Rubio-Rodríguez et al. 2010). As shown in figure I-6, there is a growing market that demands this type of FAs for the production of functional food, pharmaceutical uses, aliment for pets etc. Unfortunately there is not enough biomass that can be sustainably harvested from the oceans to extract omega-3 FA to manufacture these products and meet the existing demand in the market. Thus, a sustainable alternative source that replace fish is required soon. Bacteria and eukaryotic microorganisms have been regarded as a promising source of LC-PUFAs. The benefits to using microorganisms include a stable supply of LC-PUFAs and the fact that microbial oils are not associated with marine pollution, fish odor, or fish allergies. Currently, some LC-PUFAs are produced at the industrial level using eukaryotic microorganisms. However, LC-PUFA-producing bacteria have not been extensively used as a profitable source because of their low productivity. Therefore, there is a need to understand the mechanism of synthesis of this type of valuable molecules to have the capacity to improve yields in the laboratory and thereby design new sustainable biological systems to produce these essential oils.

## 1.2. Triglyceride synthesis in bacteria

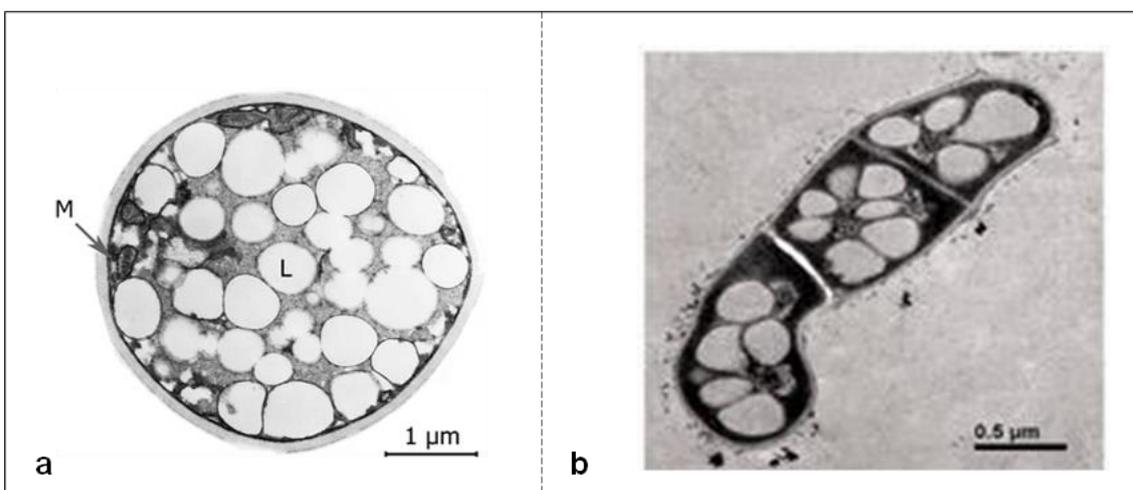
Most of the living cells have the ability to deposit intracellular carbon storage molecules, as it is an evolutive advantage when sources of energy are depleted. Triglycerides (TAGs) are esters derived from glycerol and three fatty acids and they are highly concentrated stores of metabolic energy because they are reduced and anhydrous, and show a higher calorific value than carbohydrates and proteins and therefore yield significantly more energy upon oxidation (Berg, Tymoczko, and Stryer 2002b; Berg, Tymoczko, and Stryer 2002b). There are diverse types of TAG with different properties depending on their fatty acid composition. Three fatty acid chains are bonded to each of the three -OH groups of a glycerol molecule by the reaction of the carboxyl end of the fatty acid (-COOH) with the alcohol. Water is eliminated and the carbons are linked by an -O- bond through dehydration synthesis. TAGs are described as the major storage molecules of metabolic energy in the nature, found in plant seeds, in animal adipocytes and in the cytosol of unicellular eukaryotic organisms like yeasts. They are also synthesized by some bacteria, where sometimes other energy reservoirs like wax esters and polyhydroxyalkanoates (PHAs) can be found (Alvarez and Steinbüchel 2002; Santala et al. 2014; Wältermann and Steinbüchel 2005).

### 1.2.1. Lipid Reservoirs in Bacteria

The occurrence of intracellular TAG reservoirs is widespread in eukaryotes, while its accumulation in prokaryotes is restricted only to a few aerobic heterotrophic bacteria and some cyanobacteria. Nearly all prokaryotes known are able to accumulate at least one type of storage compound and some species that belong to the gram-positive actinomycetes group were described to biosynthesize TAG (*Mycobacterium* sp., *Nocardia* sp., *Rhodococcus* sp., *Micromonospora* sp., *Dietzia* sp., *Gordonia* sp. and *Streptomyces*). For instance, fatty acids in acylglycerols in cells of *Rhodococcus opacus* PD630 accounted for up to 87% of the cellular dry weight. TAG inclusions have also been reported for some gram-negative proteobacteria like *Alcanivorax* sp.,

*Marinobacter* sp. or *Acinetobacter* sp (Manilla-Pérez et al. 2011; Santala et al. 2011). All these species can be defined as oleaginous organisms when they accumulate more than 20% of their biomass are lipids (Alvarez and Steinbüchel 2002). Accumulation of wax esters (WEs) (esters of a fatty acid and a fatty alcohol) as storage lipid is not usual in eukaryotes as occurs only in jojoba whereas PHAs are not synthesized in eukaryotes (Bugnicourt et al. 2014). Like TAGs, WEs accumulates in few prokaryotic species, acting in some of them (such as *Acinetobacter baylyi*) as the main storage compound. In contrast, PHAs are common in the entire bacteria spectrum (Ishige et al. 2003). The principal function of all these bacterial storage lipids seems to be as a reserve compound. In most bacteria, accumulation of TAG and other neutral lipids is usually stimulated if a carbon source is present in the medium in excess and if the nitrogen source is limiting growth. Cellular growth is impaired under these conditions, and the cells utilize the carbon source mainly for the biosynthesis of neutral lipids during the stationary growth phase (Wältermann et al. 2000). Other functions that have been discussed include regulation of cellular membrane fluidity by keeping unusual fatty acids away from membrane phospholipids, or acting as a sink for reducing equivalents. Figure I-7 shows photographs of the oleogenic microorganisms *R. opacus* and *Candida curvata* where the lipid droplets are perfectly differentiated as they occupy most of the cellular volume.

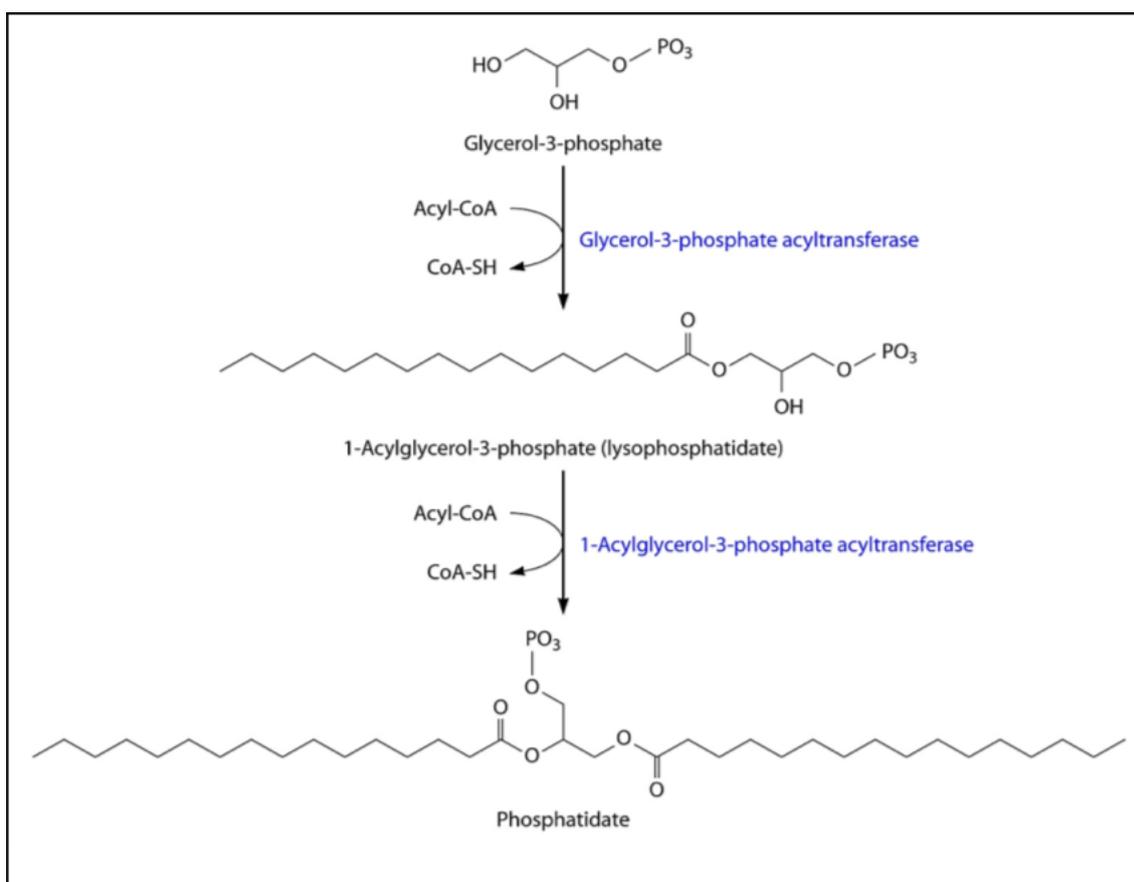
A wide range of carbon sources can be utilized by bacteria for biosynthesis of TAG. These carbon sources include sugars, organic acids, alcohols, n-alkanes, branched alkanes, phenylalkanes, oils and coal (Nigam and Singh 2011). This wide range of substrates for lipid production has always been very interesting from the point of view of the industry since these microorganisms could be used as biological factories that take advantage of waste residues. Understanding how these organisms are able to synthesize and accumulate these substances and what biological processes are behind is very interesting due to its scientific and industrial relevance.



**Figure I-7.** Electron microscopy photographs of oleogenic microorganisms. a: *Candida curvata* D grown with limiting nitrogen showing a total lipid accumulation of approximately 40%. Mitochondrion (M) and lipid droplets (L) are indicated in the figure. b: Ultra-thin sections of *Rhodococcus opacus* PD630 cultured with limited nitrogen for 48 hours. Lipid droplets can also be perfectly differentiated inside the cytoplasm as clear spheres. Figure adapted from (Ageitos, Vallejo, Veiga-Crespo, & Villa, 2011; Chen et al., 2014)

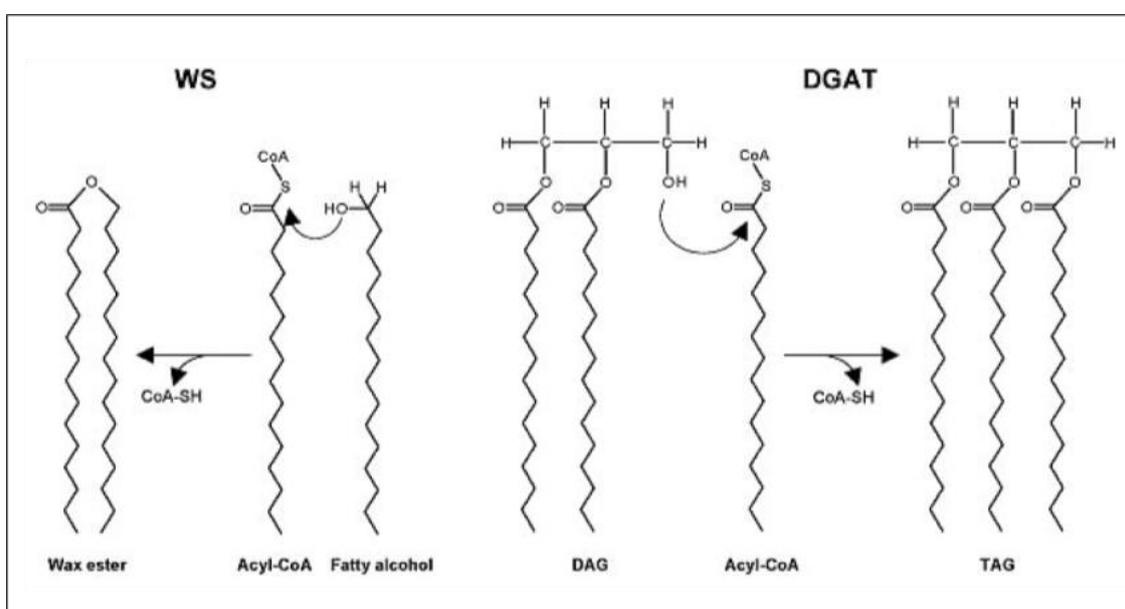
## 1.2.2. Mechanism of triglyceride synthesis: The WS/DGAT family

There is a general pathway that organisms use to produce diacylglycerol (DAG) precursors for *de novo* TAG production: the glycerol phosphate or Kennedy pathway. This biochemical route is based on the use of fatty acyl-CoA intermediates that are covalently bound to glycerol in successive reactions to form DAG and finally TAG as is represented in figure I-8. This synthetic mechanism involves membrane-bound acyltransferases and a phosphatidic acid phosphatase. Each of the acylations needed to produce TAGs is catalyzed by a different acyltransferase protein. The particularities of these acyltransferases determine the distribution of the various acyl groups in the hydroxyl groups of the glycerol molecule (Gibellini and Smith 2010). Therefore, the third step of acylation in the molecule of glycerol is the only specific enzymatic reaction for the biosynthesis of TAG. Three different classes of enzymes are known to mediate TAG formation using DAG as precursor: Acyl-CoA:DAG acyltransferase (DGAT) that catalyzes the final acylation of DAG and are found in animals and plants; acyl-CoA-independent TAG synthesis mediated by a phospholipid:DAG acyltransferase found in yeast and plants that uses phospholipids as acyl donors for DAG esterification; and a third alternative mechanism present in animals and plants that is the TAG synthesis by a DAG-DAG-transacylase, which uses DAG as acyl donor and acceptor yielding TAG and monoacylglycerol although no gene coding such a transacylase could be identified as yet (Kalscheuer and Steinbüchel 2003).



**Figure I-8.** Kennedy pathway: First two acylation steps in membrane lipid and triacylglycerol synthesis from glycerol-3-phosphate to the central intermediate phosphatidate. Enzymes are labelled in blue and molecule names in black. Figure taken from (Röttig & Steinbüchel, 2013).

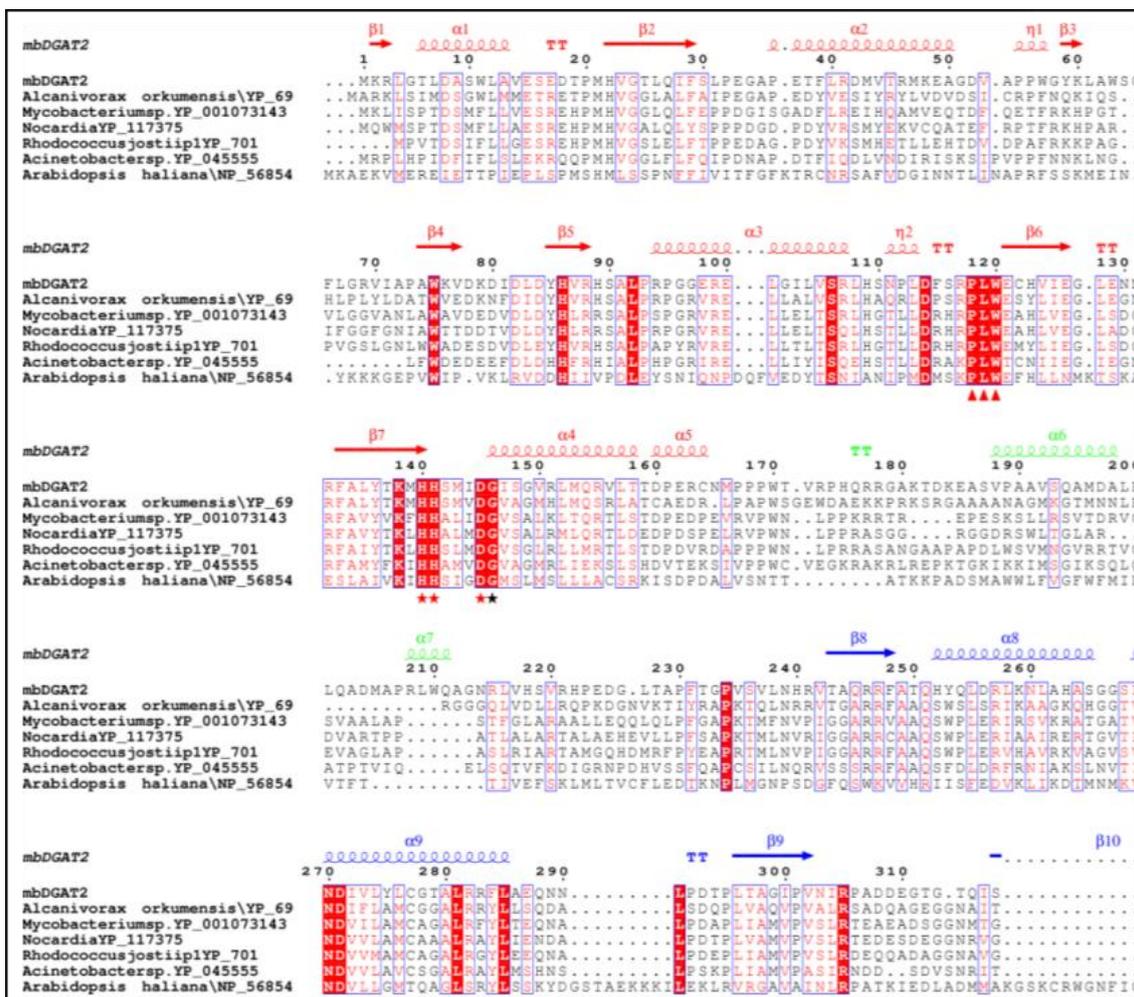
In addition to these systems for TAG production, it is known an alternative system to perform the final step of TAG synthesis that is present in bacterial species that naturally produce lipid droplets. The major synthesis of TAG and WE in these organisms is dependent of an enzyme named as wax ester synthase/acyl-CoA:diacylglycerol acyltransferases (WS/DGAT) whose first member was described in 2003 from *Acinetobacter baylyi* (Kalscheuer and Steinbüchel 2003). In gram-positive actinomycetes, there is a high genetic redundancy of these enzymes, including 15 putative genes in *Mycobacterium tuberculosis* (Daniel et al. 2004) and 14 genes in *Rhodococcus jostii* RHA1 (Hernández et al. 2008). This family of enzymes are not related in sequence or structure to those involved in formation of similar lipids in eukaryotes. WS/DGAT mediate the formation of both TAG and WE storage molecules condensing acyl-CoA molecules as acyl donors and long-chain fatty alcohols or diacylglycerols (DAGs) as respective acyl acceptors (Stöveken et al. 2005). Figure I-9 shows a schematic diagram of the two reactions catalyzed by WS/DGAT proteins.



**Figure I-9.** Biochemical reactions catalyzed by the enzymes of the WS/DGAT family. WE and TAG are synthesized from acyl-CoA and fatty alcohol or DAG, respectively. Figure taken from (Kalscheuer & Steinbüchel, 2003).

Sequence analysis of the WS/DGAT family show that all of them share a conserved catalytic motif "HHXXXDG" also present in other acyl-CoA-dependent acyltransferases as is shown in the sequence alignment of figure I-10. Generally, enzymes with this motif have the ability to transfer thioester-activated acyl substrates to a hydroxyl or amine acceptor to form an ester or amide bond. A few enzymes can be found with this catalytic motif such as acyltransferases that synthesize glycerolipids, nonribosomal peptide synthetases, acyltransferases involved in lipid biosynthesis, polyketide-associated acyltransferases, or chloramphenicol acetyltransferase (CAT). Both histidines within the HHxxxDG motif were reported to be essential for the correct function of the protein and the second one acts as a general base that promotes deprotonation of the hydroxyl group of the alcohol to catalyze the transfer of the acyl group (Lewendon et al. 1994; Stöveken, Kalscheuer, and Steinbüchel 2009). The

aspartic acid of this motif is also critical for the enzyme activity playing a structural role in the organization of the active site (Villa et al. 2014).

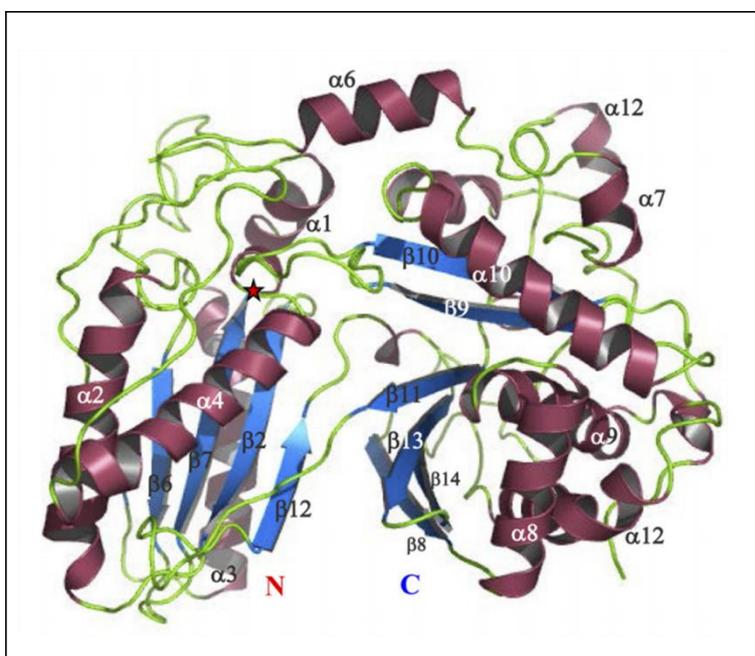


**Figure I-10.** Multiple sequence alignment of representative WS/DGAT proteins from different organisms. Identical residues are shown in white on a red background, while similar residues are shown in red. Prediction of structural elements for Ma2 from *Marinobacter hydrocarbonoclasticus* is shown above the alignment. Secondary structure representation is colored red for the N-terminal domain, blue for the C-terminal domain, and green for the connecting helices. The stars indicate the active site residues and the arrows highlight the structural motifs "PLW" and "ND". Figure taken from (Villa et al., 2014).

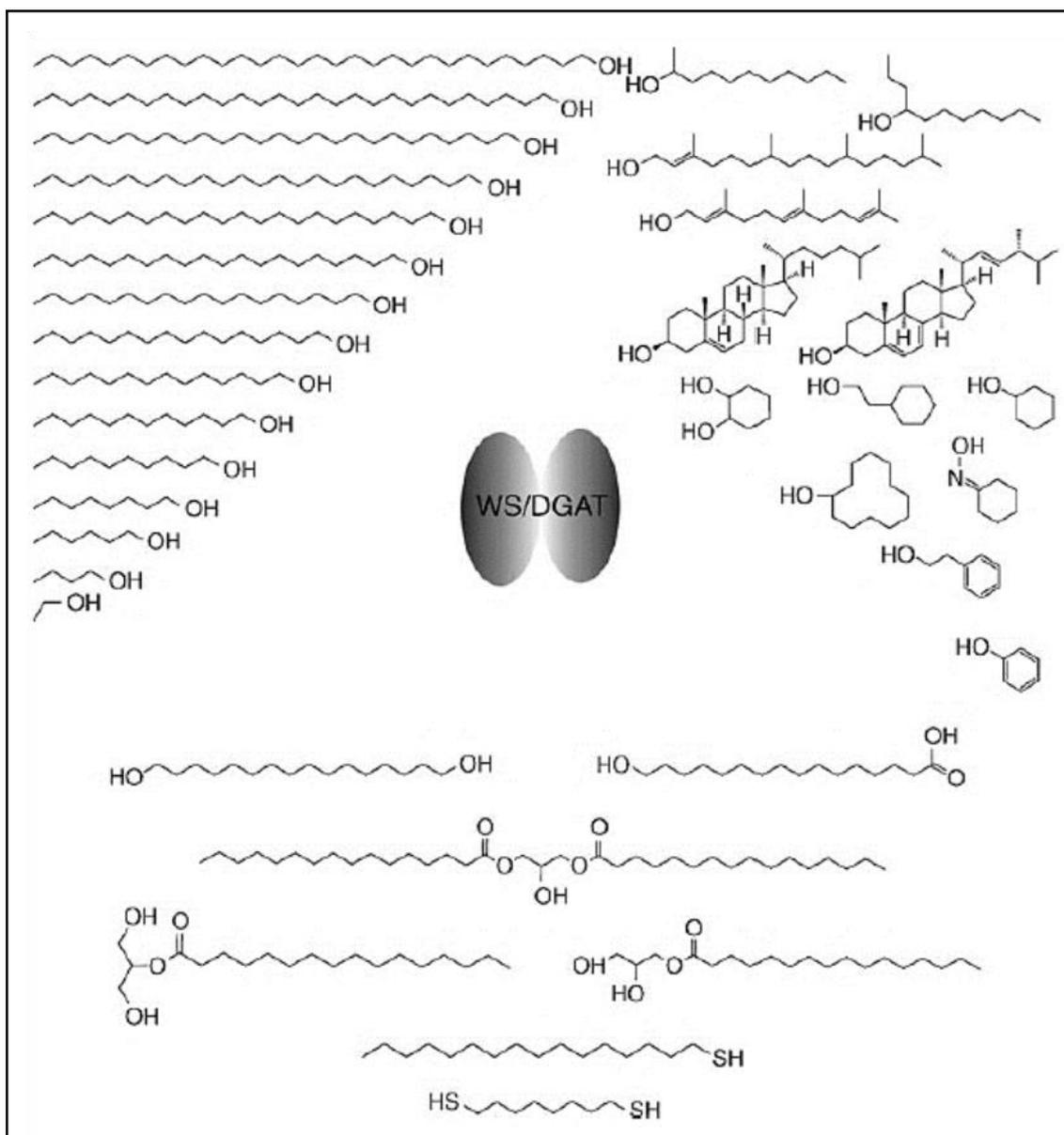
There is no structure available for any protein belonging to the WS/DGAT family but structural models are predicted to be quite accurate due to the high homology they have with other protein domains that has been solved as the tyrocidine synthetase, TycC (2V5Q) from a nonribosomal peptide synthetase (Samel et al. 2007). Model suggests that WS/DGAT proteins can adopt a 3D structure with two domains (N-terminal and C-terminal domains) connected by a helical linker, similar to other HHxxxDG acyltransferases as is shown in the three-dimensional model in figure I-11. To confirm the model prediction, limited proteolysis and mutagenesis experiments were carried out in previous experiments in our research group, demonstrating the presence of two structural and independent domains within the protein. Some other conserved residues

of WS/DGAT enzymes were also defined. These residues are not directly involved in the catalysis of the acyltransferase reaction. The conserved residues defined as “PLW” or “motif I” and the “ND” or “motif II” may be essential for the correct folding of the protein core as any of their mutants drastically reduced the enzymatic activity of the WS/DGAT from *M. hydrocarbonoclasticus*. R305 was proposed to orientate the D residue of the catalytic motif in a correct position (Villa et al. 2014).

WS/DGAT enzymes have been reported to have a remarkably broad spectrum. Experiments done with the WS/DGAT protein from *Acinetobacter sp.* demonstrated that it is able to accept a wide range of substrates when *in vitro* experiments were done (Stöveken et al. 2005). High enzymatic activities were observed for medium-chain-length alcohols (C14 to C18) and palmitoyl-CoA substrates, but alcohols and acyl-CoAs with different lengths were also accepted as well as cyclic and aromatic alcohols. In general, it is only the availability of hydrophobic substrates that can limit the products synthesized by this family of enzymes (Stöveken et al. 2005; Stöveken and Steinbüchel 2008). It was also reported that WS/DGAT enzymes can utilize mono acyl glycerols as substrates, to directly synthesize TAG. This means that in some organisms the production of TAG is not exclusively occurring via the Kennedy pathway by sequential acylation of glycerol-3-phosphate, but also by a possible involvement of WS/DGAT in the formation of DAGs *in vivo* as well (Stöveken and Steinbüchel 2008). This extremely wide range of substrates which is represented in figure I-12, makes this family of enzymes a key target for the design of lipid production systems.



**Figure I-11.** Structural prediction analysis of the Ma2 protein from *Marinobacter hydrocarbonoclasticus* revealed a monomer with a two-domain structure: beta sheets are shown in blue, alpha helices in magenta, and connecting loops in green. The position of the active site motif HHxxxDG is shown by a red star. Taken from (Villa et al. 2014).

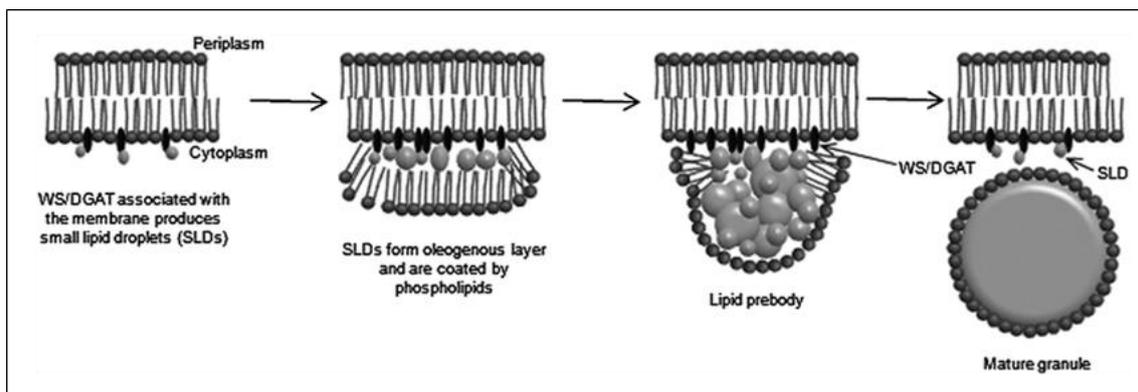


**Figure I-12.** Overview of some substrates for bacterial acyltransferases belonging to the WS/DGAT family (Stöveken & Steinbüchel, 2008).

### 1.2.3. WS/DGAT proteins and formation of lipid bodies

Prokaryotes accumulate lipids like TAG or WE as carbon storage compounds in special structures called lipid droplets (LDs) or granules. LDs are cytoplasmic structures constituted by a neutral lipid core where the lipids are stored in a non-polar environment separated from the cytosol with a monolayer membrane formed by phospholipids. There are also numerous proteins necessary for the assembly and correct function of these structures (Ding et al. 2012; Walther and Farese Jr. 2009; Yang et al. 2012). Different lipidic substances like the ones described before (TAG, WE or PHA) accumulated into these structures resulting in a wide variety of different morphologies ranging from spherical and ellipsoid to disc-like structures or even rectangular shapes (Wältermann et al. 2005).

WS/DGAT proteins are involved in the lipid body formation so they are key components to understand how they grow. These proteins are usually found attached to the cell membrane or to the lipid droplet monolayer membrane where they perform their synthesis. The lipid-body formation starts at peripheral lipid domains close to the cytoplasm membrane where small lipid droplets “seeds” remain in complex with WS/DGAT proteins due to hydrophobic surface interactions. These unstable small-lipid-droplet-emulsion accumulates in an oleogenous layer at the cytoplasm membrane at certain parts of the cell, coated by a by a half-unit membrane of PLs, to form lipid-prebodies. When these lipid-prebodies are big enough they merge with each other until they reach a critical size and are matured into proper lipid bodies to be released into the cytoplasm (Wältermann and Steinbüchel 2005). Figure I-13 show the most plausible model for the lipid body formation process to date.



**Figure I-13.** Prokaryotic TAG- and wax ester-lipid body formation. In prokaryotes neutral lipid-body formation starts with attachment of WS/DGAT to the cytoplasm membrane and subsequent synthesis of SLDs forming an oleogenous layer coated by a monolayer of PLs. Lipid-prebodies are formed by conglomeration and coalescence of SLDs leading to the formation of membrane bound lipid-prebodies and afterwards cytoplasmic lipid-bodies. Figure adapted from (Thomson, Summers, & Sivaniah, 2010)

WS/DGAT were described to be proteins with a high index of surface hydrophobicity. This may indicate that these proteins have a tendency to bind to hydrophobic elements since they are always in association with cell membranes. Due to its intrinsic hydrophobic nature, unexpected interactions may arise when proteins of this family are heterologously expressed in hosts that do not naturally produce lipid bodies. In addition to this, there are still unknown mechanisms by which lipid droplets take their form, for which certain proteins that are missing in the new host may be required. In this thesis experiments aimed to test such membrane-protein interactions will be conducted to understand how these heterologous production systems behave.

#### 1.2.4. Biotechnological use of WS/DGAT enzymes: tDGAT

It has not been established yet an economic process for microbial production of TAGs at an industrial level. Bacterial species capable of synthesizing sufficient yields of valuable lipids (*A. baylyi* or *R. opacus* for instance) require strict culture conditions, show low carbon source flexibilities, lack efficient genetic modification tools and in some cases poses safety concerns (Alvarez, Kalscheuer, and Steinbüchel 2000; Kalscheuer and

Steinbüchel 2003). For these reasons, many efforts have been made in the genetic manipulation of bacteria for gene expression to use them in the laboratories in a controlled environment. *E. coli* has been successfully engineered to synthesize jojoba oil-like WEs (Röttig and Steinbüchel 2013), fatty acid butyl esters (FABEs) and FAEEs (Kalscheuer, Stölting, and Steinbüchel 2006; Röttig and Steinbüchel 2013). TAG production has also been reported in *E. coli* (Janßen and Steinbüchel 2014; Lin et al. 2013; Rucker et al. 2013). However, recombinant oil production in a non-modified *E. coli* only produced insignificant quantities of product that does not meet the initial expectations. In a recent work in our laboratory a codon-optimized gene coding for a WS/DGAT from a gram-positive thermophilic bacteria, *Thermomonospora curvata* (Chertkov et al. 2011) was patented. When this gene was cloned into an expression plasmid and transformed into *E. coli* a very rapid *in vivo* activity was proved (Lázaro et al., 2017, in press). This enzyme that we called “tDGAT” was shown to be active in the industrially appropriate microorganism *E. coli*. This way we successfully introduced the TAG biosynthesis pathway in a gamma-proteobacteria, which gives rise to the significant production and accumulation of TAGs and sets the basis for further research on the achievement of a suitable method for oil production in microorganisms. At the same time, the lipid composition observed led us to think that the WS/DGAT might have possible new specificities, making it highly attractive for biotechnological applications such as biodiesel production.

Despite being able to produce TAG in *E. coli* in a relatively short time by overexpressing a WS/DGAT enzyme, we believe that it is possible to improve the production system by doing directed evolution experiments. By introducing certain changes in the protein sequence it would be possible to improve either the protein-protein or protein-membrane for interactions, as well as the mechanism of catalysis. In the first part of this thesis work it will be detailed how we have managed to improve the production of triglycerides by using of directed evolution and how this can help to reach scalable yields at industrial level for other proteins.

### 1.3. Fatty acid synthesis

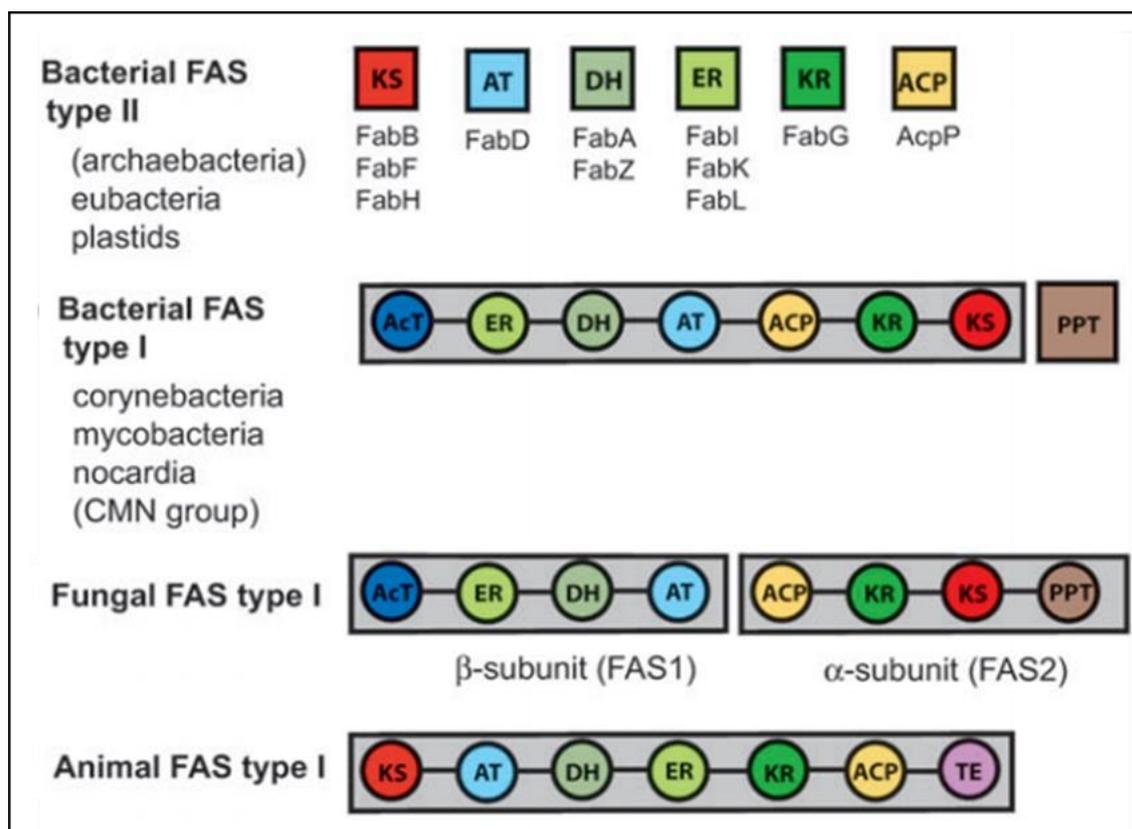
The primary role of bacterial fatty acids is to act as the hydrophobic component of the membrane lipids forming phospholipids. They can also be found as components of storage lipids in lipid droplets, being part of more complex molecules as TAG and waxes in some *Actinomycetes* (Alvarez and Steinbüchel 2002; Ishige et al. 2003) or even forming polyhydroxyalkanoic acids in organisms like *Bacillus subtilis* (Bugnicourt et al. 2014). Bacterial FA are very similar to the ones found in eukaryotic cells but they tend to be slightly shorter and generally saturated or monounsaturated. The mechanism of synthesis of regular saturated FA is conserved through evolution, so there are no strong differences between the ones found in Bacteria and Eukarya (some unconventional FA moieties can be found in Archaea since they can synthesize isoprenoid-derived lipids, but we will not go into detail in this work). Additionally, some bacteria are able to produce various types of rare FA, like branched chain fatty acids or hydroxyacyl acids (Kaneda 1991).

In the next paragraphs the mechanisms by which living organisms are able to produce regular FA will be explained as well as the role of the different protein domains that carry out the process. Due to the similarities found between regular FA, polyketides and omega-3 FA synthesis we will use the basics of fatty acid and polyketide synthesis

as a model to understand the results shown in this manuscript. The information available in the literature about the individual contribution of the different protein domains involved in the process will be used to introduce basic biochemistry and structural notions that will help to understand the subsequent scientific work.

### 1.3.1. The fatty acid synthase systems I and II

Although the lipids found in Bacteria and Eukarya are similar, they show important differences in terms of the arrangement of the protein domains that perform their synthesis. Fatty acid synthases (FAS) are divided into two distinct molecular forms called type I and II (figure I-14). FAS type I are typical from eukaryotic cells but can be found as a remarkable procaryotic exception in the mycolic acid producing subgroup *Actinomycetale* (Bloch and Vance 1977). FAS type I proteins consist of large polypeptides with catalytic domains distributed in single domains that form huge megasynthases (up to 2.6 MDa). These enzymes can be organized as single proteins (as in mammals and mycobacteria) or be split in two interacting proteins (as in fungi and *Corynebacterium*). These big polypeptides can be assembled as homodimers like the human FAS of 540 kDa (Maier, Leibundgut, and Ban 2008a) or in more complicated structures as the heterododecamer of 2.6 MDa (Jenni et al. 2007) found in fungi. In contrast with this, each enzymatic activity is found as a discrete protein in type II FAS systems, which are naturally present in bacteria (Marrakchi, Zhang, and Rock 2002). Both systems are organized in different ways but they actually share the same protein domains. Therefore, the biochemical reactions they perform to produce FA are very similar. However, it should be pointed out that FAS I is a more specialized system and usually produces only palmitate (Smith, Witkowski, and Joshi 2003). In contrast, FAS II is a more versatile system that is capable of producing a diversity of products for cellular metabolism. Not only different chain lengths are produced by FAS II systems, but different degrees of unsaturation, branched-chain fatty acids, and hydroxy fatty acids. In addition, FAS II intermediates are used in the synthesis of key cellular constituents, such as lipoic acid and quorum-sensing molecules (White et al. 2005).



**Figure I-14.** Schematic representation of fatty acid synthases genetic arrangement in different organisms. Distinct proteins are indicated as squares and domains integrated within proteins as circles, respectively. Abbreviations: KS, ketosynthase; AT, acyltransferase; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; ACP, acyl carrier protein; AcT, acetyltransferase; PPT, phosphopantetheinyl transferase. Figure adapted from (Jenke-Kodama, Sandmann, Müller, & Dittmann, 2005).

### 1.3.2. Overall mechanism of fatty acid synthesis

Generally, FA biosynthesis begins with acetyl-CoA that is firstly incorporated to seed a new FA molecule. After that, through a series of carboxylation rounds, malonyl-CoA building blocks are incorporated to the growing chain. As a consequence of that, malonyl-CoA will be required for all the elongation steps and have to be synthesized in a previous reaction catalyzed by acetyl-CoA carboxylase that uses bicarbonate and acetate as precursors (Alberts and Vagelos 1968). The new FA molecule is going to be reduced at the end of each condensation step in an iterative fashion until the FA matures for being use by the cell. Almost all animal FA synthesis are done by assembling segments of two carbons length together (malonyl-CoA extender units), and as a consequence of this the resulting fatty acid has always an even number of carbon atoms (Wakil, Stoops, and Joshi 1983). The following paragraphs will go into the details of this biochemical mechanisms necessary for the production of FA.

Both, type I and type II FAS, rely on a small protein known as the acyl carrier protein (ACP) to shuttle the fatty acid cargo from domain to domain to perform the condensation and reduction steps (Chan and Vogel 2010). To start, the synthesis the

ACP must be activated via post-translational modification by an enzyme called phosphopantetheinyl transferase (PPTase) that install a coenzyme A derived 4'-phosphopantetheine arm (PPT), very important to form the tether upon which the growing fatty acid is bound (Flugel et al. 2000). The ACP then plays a central role in the synthesis keeping the FA attached and exposed to the rest of the protein domains.

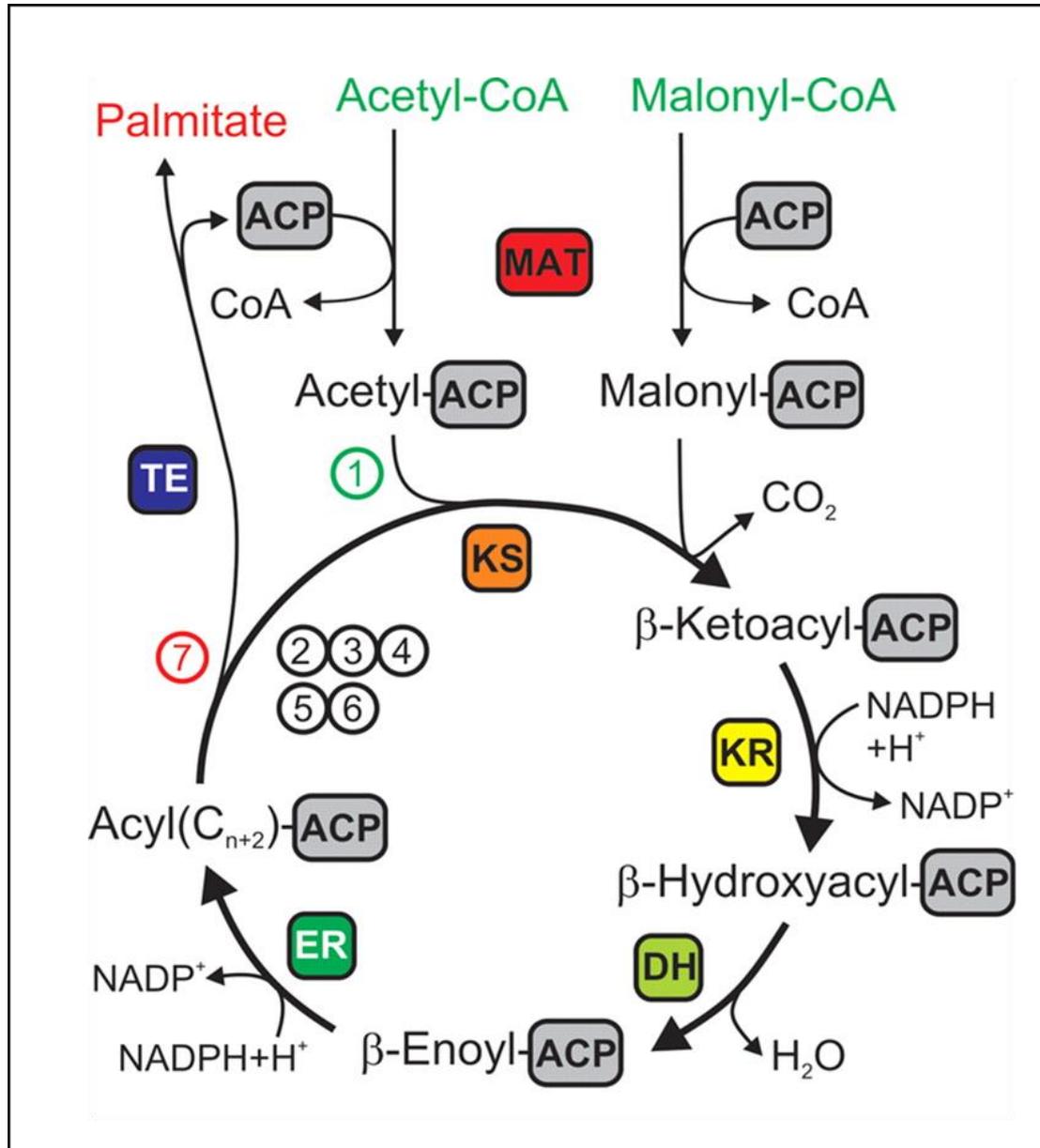
The malonyl-CoA units are transferred to the ACPs by special domains called Malonyl-CoA-ACP acyl transferases, generally known as acyl transferases (AT or MAT if specific for malonyl) as is represented in figure I-15. These domains (*fabD* in *E.coli*) catalyze the loading of fatty acid building blocks by conversion of malonyl-CoA to malonyl-ACP (Rock and Cronan 1996). The AT domain is firstly malonylated at an active site serine residue. Then the malonyl groups are transferred to the ACP PPT, activating them for the successive synthesis reactions. Despite the fact that malonyl is always the basic extender unit, the synthesis is always initiated with a molecule of acetyl-CoA or acetyl-ACP (John E. Cronan and Rock 2008) that is condensed with malonyl-ACP to form acetoacetyl-ACP in the first synthesis cycle (performed by *FabH* in *E.coli*). In this way the acetate unit will become the methyl end of the future FA. Thus, AT domains are responsible for starting and leading the biosynthesis by selecting the fundamental blocks with which the FA will be constructed. Numerous experiments demonstrate that AT domains are potentially promiscuous being able to accept a variety of CoA analogs *in vitro* like  $\beta$ -ketobutyryl,  $\beta$ -hydroxybutyryl and crotonyl moieties but showing a clear preference for malonyl-CoA (Ghayourmanesh and Kumar 1981; Joshi, Witkowski, and Smith 1997).

Once the ACPs is charged with malonyl units, the FA chain will be iteratively extended by other domains known as ketoacyl synthases (KS) that are normally located adjacent to the AT domains and very close to the ACPs within the protein architecture. The KS (either *FabH*, *FabB* or *FabF* in *E.coli*) catalyze the chain extension through an iterative process achieved by successive attacks of the  $\alpha$ -carbon of the malonyl-thioester onto the acyl-thioester intermediate with decarboxylation of the malonyl group, resulting in  $\beta$ -keto-thioesters (Johan Gotthardt Olsen et al. 2001). Thus, the KS domain channels malonyl groups previously selected and bound to the ACP by the AT domain into the FA chain, extending in this way the growing molecule. In this process a  $\beta$ -keto group is generated after each condensation and needs to be subsequently reduced to produce the characteristic bonding pattern of the FA.

At this point is when a perfectly coordinated action of the chain-modifier domains keto reductase (KR), dehydratase (DH) and enoyl reductase (ER) is essential. The keto-group generated in each round of condensation needs to be firstly reduced by a KR domain (*FabG* in *E.coli*), that is typically an NADPH-dependent 3-ketoacyl-ACP reductase (Heath and Rock 1995). After that, a water molecule is removed by a DH domain (*FabZ* and *FabA* in *E.coli*), a 3-hydroxyacyl-ACP dehydratase that generate a trans or cis double bond (Heath and Rock 1996). The final reduction is catalyzed by an ER (*FabI* in *E.coli*), to give an acyl-ACP, which can serve as a substrate for another round of elongation or, if it has reached the expected chain length, the FA will be ready for being released. The reduction step does not have to occur at all cycles and this is very important to be able to generate FA with unsaturations at different positions. Normally the ER's ability to remove the double bond will be conditioned by its isomeric form (cis or trans), catalyzing normally the reduction of the trans-2-acyl-ACP isomers (Massengo-Tiassé and Cronan 2009).

When the FA that is being synthesized reaches its definitive size the iterative process ends and the FA is released with the help a thioesterase (TE) domain that carries

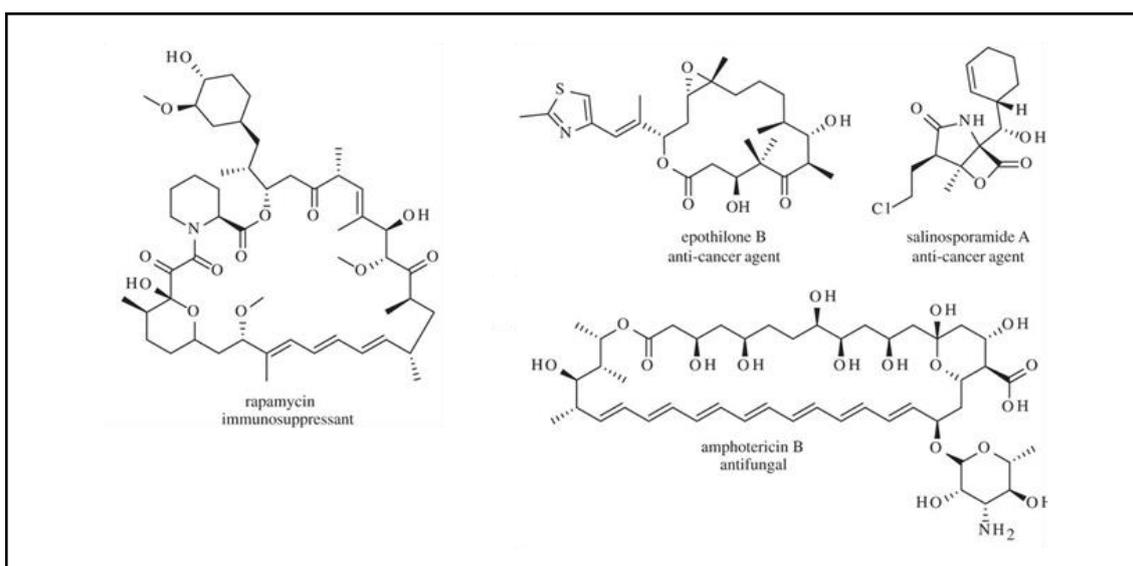
out the chain-terminating step of fatty acid synthesis, leading to the release of the synthesized molecule by the hydrolysis of the acyl-S-PPT thioester bound to the ACP domain (Chakravarty et al. 2004). The complete process of palmitate synthesis is exemplified in figure I-15.



**Figure I-15.** Catalytic cycle model for palmitate synthesis. The reaction cycle of FAS is initiated by the transfer of the acyl moiety of acetyl-CoA to the acyl carrier protein (ACP, gray) catalyzed by the malonyl-CoA-/acetyl-CoA-ACP-transacylase (MAT, red). The  $\beta$ -ketoacyl synthase (KS, orange) catalyzes the decarboxylative condensation of the acyl intermediate with malonyl-ACP to a  $\beta$ -ketoacyl-ACP intermediate. The  $\beta$ -carbon is processed by reduction through  $\beta$ -ketoacyl reductase (KR, yellow). The resulting  $\beta$ -hydroxyacyl-ACP is dehydrated by a dehydratase (DH, light green) to a  $\beta$ -enoyl intermediate, which is reduced by the NADPH-dependent  $\beta$ -enoyl reductase (ER, dark green) to yield a four-carbon acyl substrate for further cyclic elongation with two-carbon units derived from malonyl-CoA until a substrate length of  $C_{16}$  to  $C_{18}$  is reached. Finally, the product is released from the ACP by the thioesterase (TE, blue). Figure taken from (Maier, Jenni, & Ban, 2006).

## 1.4. Polyketide synthesis

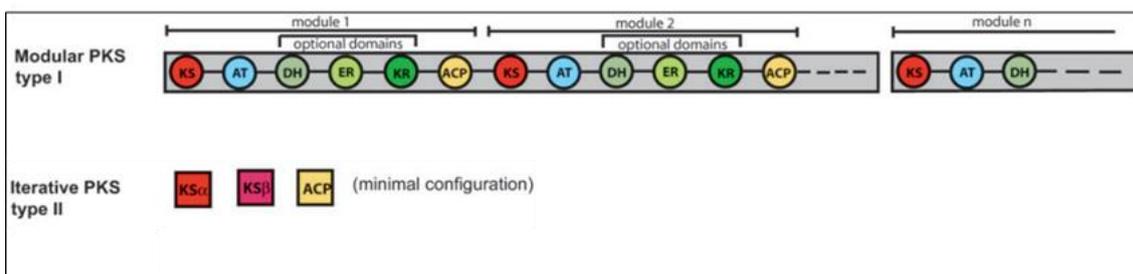
Polyketide synthases (PKSs) are a family of multi-domain enzymes that produce polyketides and are present in some bacteria, fungi, plants and a few animals. They have always been biologically striking because of their ability to produce very valuable secondary metabolites as antibiotic (Clardy, Fischbach, and Walsh 2006) or anticancer substances (Vignot et al. 2005) as the ones shown in figure I-16. Like fatty acids, polyketides are assembled by successive decarboxylative condensations of simple precursors like malonyl-CoA, but whereas the intermediates in fatty acid biosynthesis are fully reduced to generate alkyl chains, the intermediates in polyketide biosynthesis may be only partially processed, giving rise to different functional groups. The use of different starter and extender units such as methylmalonyl-CoA, propionyl-CoA or butyryl-CoA, the presence of multiple chiral centers, the possibility of modifying functional groups and the introduction of cyclizations give rise to an immense range of possible products that the living systems produce as secondary metabolites. PKS are in close relationship with FAS because they share most of their protein domains and their structural organization (Staunton and Weissman 2001). For these reasons, we considered it necessary to include them in this work, as we did with the FAS system, to use them as a complementary model to understand the omega-3 FA synthesis.



**Figure I-16.** Diversity of polyketide natural products. The panel shows different structures of polyketides that mainly arise from the diversity in extender units incorporation and the action of different accessory proteins like cyclases. Figure taken from (Dunn & Khosla, 2013).

### 1.4.1. PKS genetic organization

PKS are modular multienzyme systems that present identical ACP, KS, AT, KR, DH and ER functional domains as the ones present in FAS systems that were described in the previous section. The evolutionary pressure in combination with horizontal gene transfer events have led to the diversity of polyketides produced by bacteria combining these domains and organizing them in different ways (Ridley, Lee, and Khosla 2008). PKS and FA synthesis are very similar since both use ACPs to carry the growing molecule and shuttles it between the different modules to perform the necessary biochemical reactions. This common thioester chemistry and the adaptable architecture of PKS have promoted the emergence of hybrid PKS-FAS pathways in genera such as myxobacteria where they produce extremely complex lipids (Jenke-Kodama et al. 2005). Unlike FA synthesis, in PKS each chain elongation step is not always followed by a fixed sequence of ketoreduction, dehydration and enoyl reduction. The PKS synthesis can undergo any of these modifications in any of the intermediate steps of the route, giving rise to an enormous complexity of possible products. The following paragraphs describe how these systems are organized to produce an immense range of metabolites.

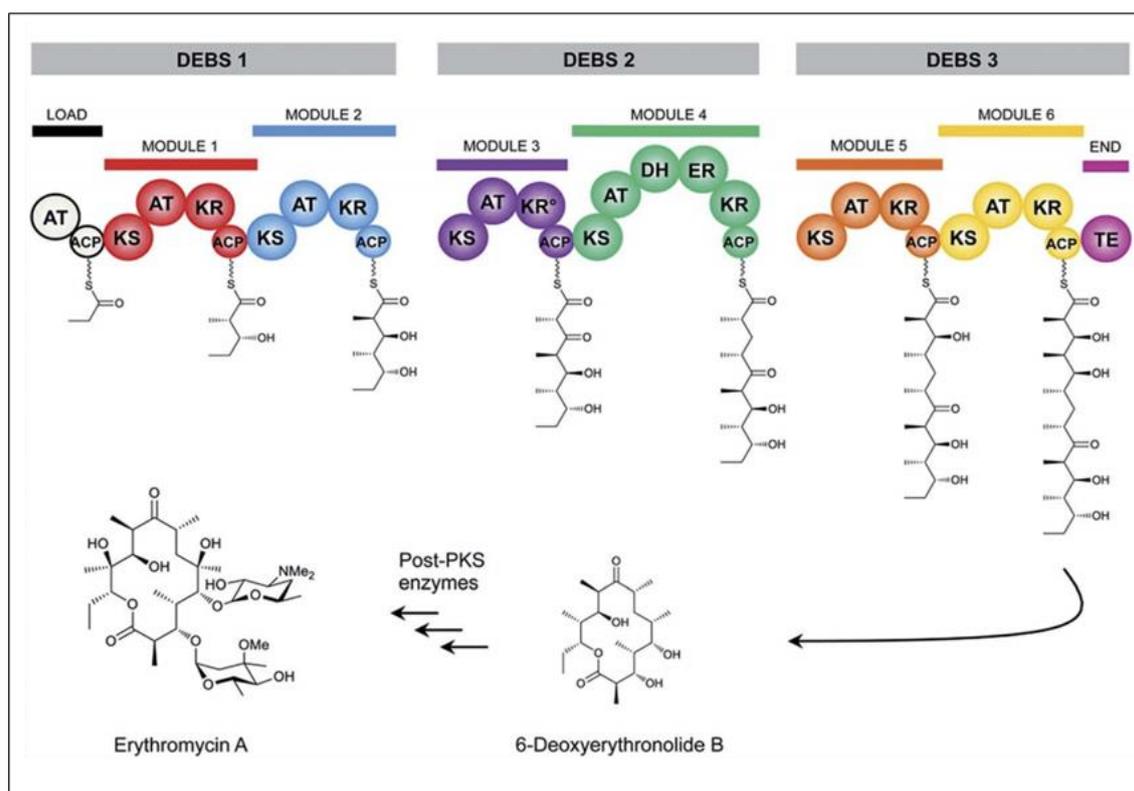


**Figure I-17.** Schematic representation of the two basic polyketide synthases genetic arrangement in different organisms. Distinct proteins are indicated as squares and domains integrated within proteins as circles, respectively. Abbreviations: KS, ketosynthase; AT, acyltransferase; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; ACP, acyl carrier protein; AcT, acetyltransferase; PPT, phosphopantetheinyl transferase. Figure adapted from (Jenke-Kodama, Sandmann, Müller, & Dittmann, 2005).

Bacterial PKS are classified into three different types according to the biosynthesis pathways they use (Jenke-Kodama et al. 2005). In figure I-17 it is shown the genetic organization of two of the three main PKs systems; the ones using ACPs. First, type I PKSs are multifunctional enzymes that are organized into modules, each of which harbors a set of active domains that catalyze one cycle of polyketide chain elongation and reduction. Type I PKS synthesis is a vectorial non-iterative process since intermediaries flow in one direction passing only once by each of the modules of the system. On the other hand, type II PKSs are multienzyme complexes that carry a single set of iteratively acting domains that perform all the condensation rounds by themselves. Type III PKSs, are enzymes very similar to the previous ones but do not use ACP domains in their synthesis. In this introduction we will focus on type I and II PKS since they share interesting characteristics with FAS and Omega-3 synthases that will provide a better global vision of our results.

### 1.4.2. PKS type I: A vectorial synthesis

The best-studied type I PKS is probably the 6-Deoxyerythronolide B Synthase (DEBS) which is found in *Saccharopolyspora erythraea* and is responsible for the synthesis of the macrolide ring precursor of the antibiotic erythromycin (Lowry et al. 2013). Its modular organization was extensively studied from a biochemical and structural point of view and will help us to understand the PKS type I synthesis. The DEBS system is composed of three large and multifunctional proteins named DEBS 1, 2 and 3. All of them are composed of two modules and each of them exist as a dimer in the cell (Caffrey et al. 1992; Khosla et al. 2007). Each of the modules is formed by a set of domains in which are always present an ACP, KS and AT domain as the minimal condensation unit, but that may also contain a set of reduction domains like KR, DH or ER. The intermediate molecules flow in one direction, from the first module to the sixth, being condensed one extender unit per module by each of the KS domains. In each step of the synthesis the extender blocks are modified according to the reduction domains present in each module. At the end of the synthesis, the TE domain release the 6-deoxyerythronolide B molecule that will be processed by other enzymes to produce erythromycin. The whole synthesis process performed by the DEBS PKS is illustrated in figure I-18.



**Figure I-18.** Biosynthesis pathway of PKS type I DEBS system. Erythromycin A synthase consists of three multi-enzymes (DEBS 1, DEBS 2 and DEBS 3) comprising loading, chain extension and termination modules. In the scheme each color represents a module of the PKS. There are several biochemical reactions that lead to the production of the final molecule which are spatially and temporally separated. The chain-extension intermediates are transferred through the ACPs of the different modules in a directional scheme. Modules are differentiated by color. Figure taken from (Weissman, 2015).

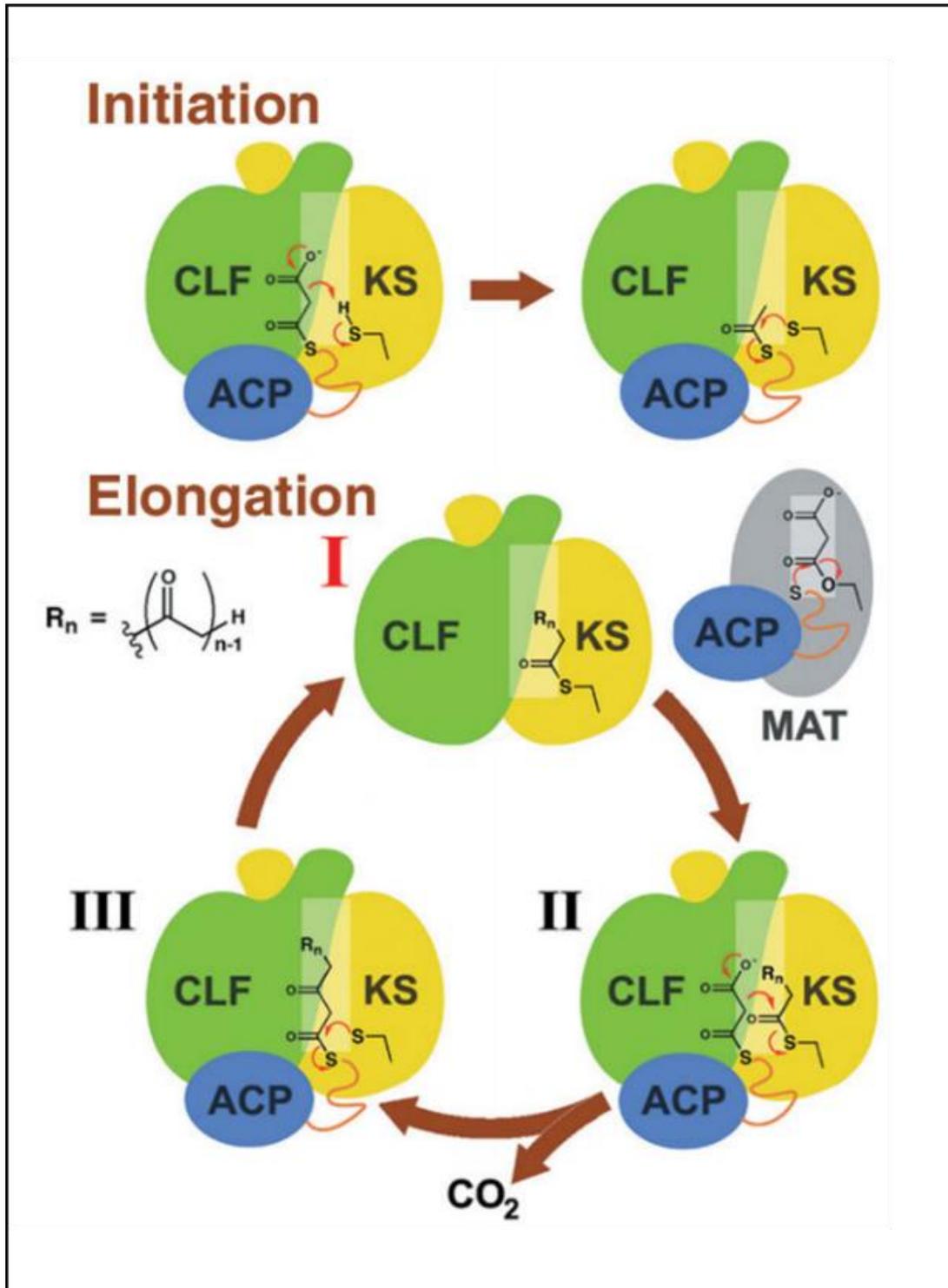
An important characteristic of these modular PKS is the presence of special docking domains at the beginning and at the end of each polypeptide that maintain in a close keep all modules of the system together, organized to maintain its directionality (Broadhurst et al. 2003; Weissman 2006). Another important proof of concept to define the essential domains that allow the synthesis of a polyketide has been the development of a minimal modular PKS using the DEBS system. This artificial construction is a diketide synthase consisting of a single chimaeric extension module, derived from the DEBS first module KS, that is sandwiched between a loading module and a chain-terminating thioesterase (Böhm et al. 1998).

### 1.4.3. PKS type II: An iterative synthesis

Type II polyketide synthases (PKSs) or iterative PKS (iPKS) are a family of multienzyme systems that catalyze the biosynthesis of polyketides such actinorhodin or tetracenomycin using a very small set of at least three protein domains for doing all the condensation rounds (Dreier and Khosla 2000). These systems are composed of discrete monofunctional proteins (like FAS type II systems) rather than the large, highly modular of type I PKS. These enzymes include a heterodimeric KS (called KS-chain length factor or KS-CLF), an ACP, and an AT domain. A central component of these systems is the KS-CLF heterodimer that in the presence of the ACP synthesizes a polyketide chain of defined length using malonyl-CoA or any other precursor (Bao, Wendt-Pienkowski, and Hutchinson 1998).

KS-CLF is only found in type II PKS within PKS systems and is absent in FAS. It was demonstrated its ability to decarboxylate malonyl extender units to prime the KS domain with the acetyl seed needed in some cases to initiate the synthesis. KS-CLF domains are also able to polymerize extender units and perform condensation rounds lengthening the nascent chain. In addition to its ability to initiate the synthesis and polymerize the chain, these special domains can limit the number of carbons that the final product will have due to the presence of a hydrophobic tunnel that somehow limits the growth of the polyketide (Dreier and Khosla 2000; Szu et al. 2011; Tang, Tsai, and Khosla 2003a).

The condensation mechanism in PKS type II is similar to the one found in iterative FAS. Firstly the AT domain load the extender units to the ACP and the synthesis is initiated when a malonyl-ACP encounters an unoccupied KS-CLF heterodimer. Decarboxylation of the malonyl group is followed by transfer of the resulting acetyl group to the active cysteine of the KS, thus priming the PKS. The growing chain is now polymerized via a series of iterative condensation rounds performed by the KS-CLF domain. The final product is supposedly released when the hydrophobic tunnel of the CLF is occupied, however the mechanism by which this occurs is not fully understood yet (Keatinge-Clay et al. 2004). A simplified overview of the synthesis is illustrated in figure I-19 that represents the actinorhodin synthesis.



**Figure I-19.** Actinorhodin production by a PKS type II system. Synthesis is divided in two steps: initiation and elongation. During the initiation the malonyl-ACP is decarboxylated by CLF to form an acetyl group that primes KS. In the elongation the malonyl acyl transferase (MAT) transfers malonyl groups (I) from malonyl-CoA to ACP, which shuttles them to KS-CLF. KS decarboxylates the malonyl group that is condensed with the growing polyketide (II). The growing chain extrudes into the polyketide tunnel and is transferred back to KS before ACP dissociates (III). Figure taken from (Keatinge-Clay, Maltby, Medzihradzky, Khosla, & Stroud, 2004)

## 1.5. Structural biology of FAS and PKS domains

The reactions for FA and PKS production, as was explained before, are described as a series of consecutive enzymatic reactions that are carried out by protein domains and lead to the production of compounds that are repetitions of single molecular blocks. These blocks can be fully reduced as in FA synthesis or may show some reductions and other modifications in its chains in PKS synthesis. In this part of the introduction we focused our research in the description of the structural knowledge accumulated in the last years for the different domains that carry out this process. For this purpose FAS systems will be described as a model for understanding the whole synthesis process, but also some examples of PKS systems will be used in order to complete this general overview.

FAS proteins may be encoded by discrete and small genes, as in the case of type II FAS, or by single genes that lead to the production of multidomain large polypeptides in FAS type I systems. In both cases the synthesis is similar: intermediate FA molecules are shuttled between the different protein active sites (KS, AT, KR, DH and ER) to be condensed and reduced as thioesters of an ACP domain. Both systems have been extensively studied in model organisms like *E.coli*. There are available high-resolution X-ray and/or NMR structures of every domain present in them. Experiments focused on FAS type II single domains have been historically the ones that led to the resolution of most of the protein structures available in the literature since they are single polypeptides, but there were also some successes in the study of entire FAS type I megasynthases.

Similar events have occurred in relation to the PKS research field and the structure of most of their protein domains has been solved as well as the structure of some entire modules carrying multiple domains. As has been introduced in the last two sections, their biochemistry and structural biology are closely related with FAS as they share most of their domain functionality and structural organization. As PKS are especially relevant for the production of certain commercial products with medical interest, much effort has been put into elucidating the structure of many of their domains. In addition to the structural study of individual domains of PKS, there have been structures of entire modules of PKS complexes, as the case of the pikromycin polyketide synthase (Dutta et al. 2014).

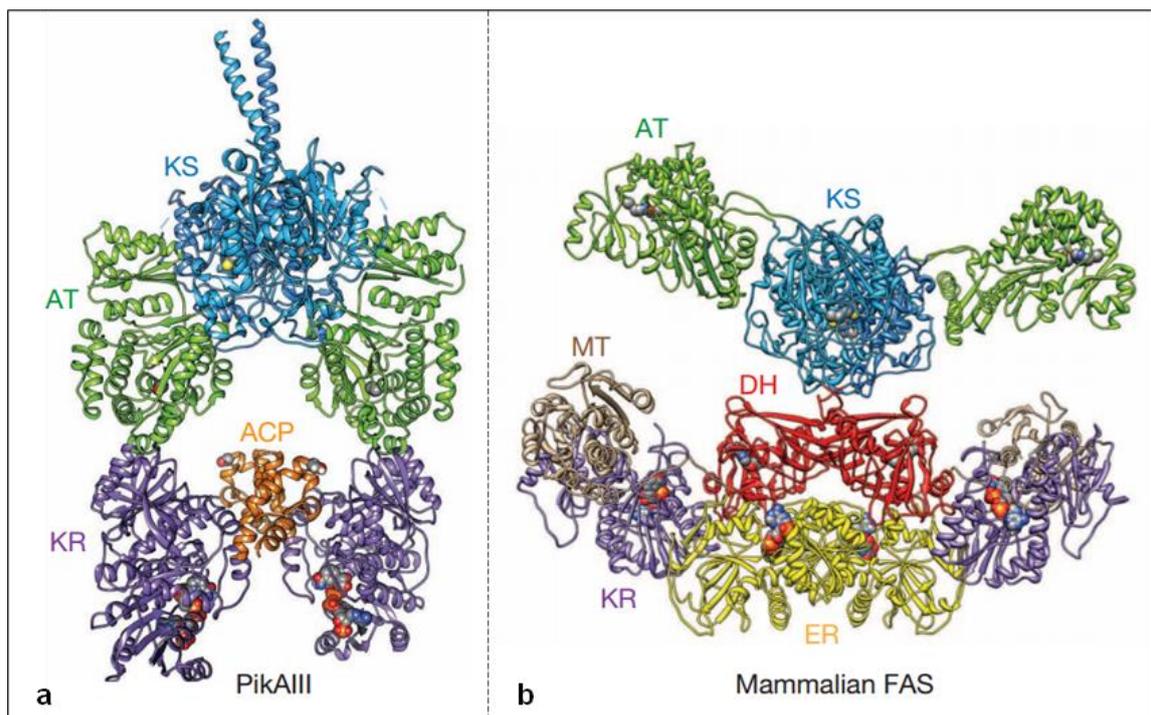
The structures solved (from FAS and PKS systems) and the associated biochemical experiments have generated an extensive knowledge about the role of their catalytic protein motifs. Now the scientific community has a lot of information about the substrate recognition, the catalytic mechanisms or the protein-protein interactions of the enzymes that are part of these systems. This helps to understand the way by which FA and polyketides are produced despite the intrinsic difficulties these systems show as their individual domains act most of the times as a whole machinery and not as individual entities. In this part of the introduction some basic structural concepts about each of the different domains will be explained based on the information available in the bibliography. Increasing the knowledge about the function of these proteins will facilitate the understanding of the results obtained in the subsequent study of omega-3 synthases since they rely on a similar biochemical route.

### 1.5.1. Megasyntases and big complex structure

Before introducing the structure of each domain needed for FA synthesis, it is convenient to have an overview of the most representative megastructures that have been solved in the last years using techniques of X-ray crystallography and electron microscopy. FAS type I systems as well as modular and iterative PKS systems have been partially understood thanks to the global vision provided by some structures solved of the whole complexes. Interactions between domains, spatial localization of them and refinement of the old biochemical models have been some of the achievements driven by the structural understanding of these complexes.

FAS type I systems of fungi (Jenni et al. 2007) and human FAS (Maier, Leibundgut, and Ban 2008a) were the first representative structures obtained with a good resolution using X-ray crystallography. Human FAS is a “small” homodimer of 540 kDa in comparison with the huge heterododecamer of 2.6 MDa of fungi, but both of them mostly share the same catalytic domains. Electron density of both structures covers the basic five enzymatic units (KS, AT, KR, DH, ER domains) but the ACP remain unresolved due to its intrinsic flexibility. Despite this, the chambers in which the ACP must be located are defined. Both structures reveal architectures of alternating linkers and domains with separation between enzymatic units. In the human FAS there is a clear separation between the condensing and modifying portions, which are mainly connected by linkers rather than direct interaction.

All catalytic FAS domains are closely related to PKS domains at the sequence level. In fact the human FAS can be considered as a single PKS module that is strongly specialized for iterative fatty acid synthesis. This has been proved with the solution of some PKS entire module structures in the last years such as the modular PKS PikAIII from *Streptomyces venezuelae*, using electron cryo-microscopy (Dutta et al. 2014), or the pseudo-iterative PKS from *Mycobacterium smegmatis*, solved very recently using hybrid crystals (Herbst et al. 2016). PikAIII structure resembles the structure of human FAS. Although PIKA has an open single ACP chamber, the overall structural organization is very similar. Both structures have the condensation domains at the top and the modifier domains separated in a lower region as is shown in figure I-20. The main difference found at the overall structural arrangement is a more open architecture of mammalian FAS with its two ACP chambers instead of one. The single closed chamber of PikAIII probably helps to make a more exclusive active-site entrances so that it prevents incorrect substrates deliver by other near-neighbor ACPs, facilitating the efficient substrate channeling from the upstream module.

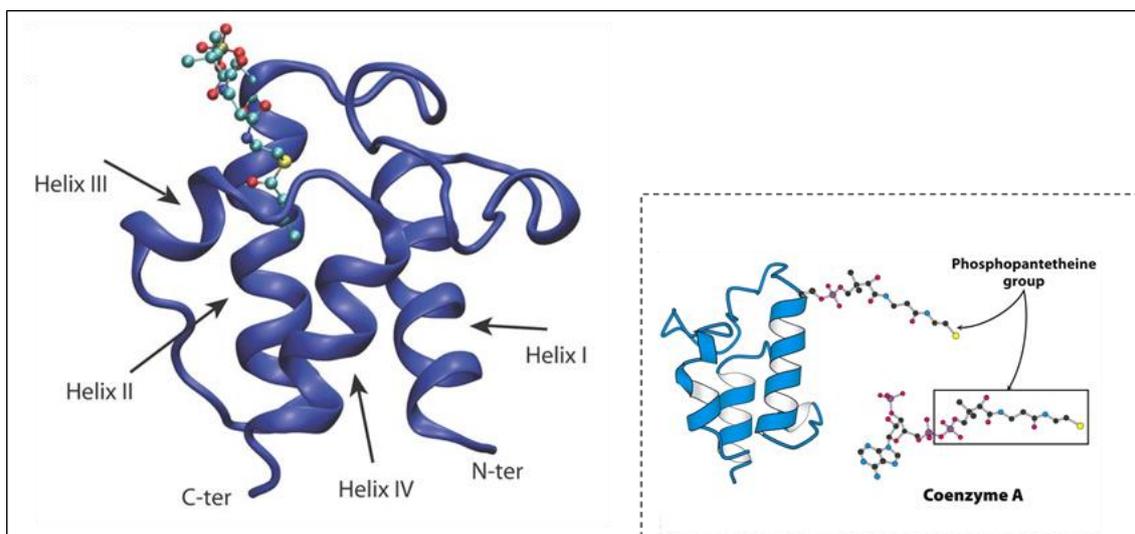


**Figure I-20.** Structure and organization of two representative megasynthases: the animal fatty acid synthase (FAS) and the PikAIII from the pikromycin polyketide synthase (PKS). Panel A shows the structure of one module of the pikromycin PKS whereas panel B shows the solved structure of the animal FAS. Domains of both representations have been colored with the same color code: blue (KS), green (AT), purple (KR), red (DH), grey (MT) and yellow (ER). ER, DH and MT domains are absent in PikAIII. Active sites are highlighted with spheres. The domains present in both structures retain the same folding and are organized as dimers, although there are topological differences in their structural organization. In PikAIII, each AT is rotated by  $120^\circ$  relative to its position in the FAS forming an extensive interface with the KS domain. With this spatial arrangement a single ACP reaction chamber is created in the centre of the PikAIII dimer, whereas in the mammalian FAS there are two of them due to the central position the DH domains occupy (Weissman, 2015).

### 1.5.2. The Acyl carrier Protein

The ACP plays a central role in both FAS systems and PKS systems due to its implication in almost all the steps of the pathway carrying the intermediate molecules and shuttling them between the different enzymatic regions of the protein complex. Indeed, ACP was reported to be the third most abundant protein present in *E. coli* (Lu et al. 2007) which gives an idea of the importance for this small domain for the correct function of the cell. The first ACP solved structure was from *E. coli* in the 80's using NMR (Gally et al. 1978). It is not trivial that many of the later arose structures have had to be solved using the same technique since these domains have a strong structural flexibility and therefore show difficulties to produce crystals. In fact, in all megastructures published to date (from FAS type I) the ACP domains cannot be visualized due to a loss of definition in the electronic density, probably because their high mobility within the crystal (Jenni et al. 2007; Maier, Leibundgut, and Ban 2008a).

Not all the ACPs in nature are part of FAS or PKS complexes but all of them show a clear structural homology (White et al. 2005). The general architecture shared between all known ACP consists of four  $\alpha$ -helices with a long but structured loop that connects the first and the second helices. These helices are oriented in an up-down-down topological arrangement to form a helical bundle plus a short fourth helix of lower stability, normally parallel and almost perpendicular to the three helix bundle in all available structures (Johnson et al. 2006; Kim and Prestegard 1990). There are quite similar inter-helical hydrophobic interactions in all ACPs known to help in stabilizing the helix bundle (Wong et al. 2002). The loop connecting first and second helices is the most flexible region of the protein that lacks a defined structure (Crump et al. 1997). Some studies suggest that the second  $\alpha$ -helix is the one that recognizes some of the other FAS or PKS domains (Zhang et al. 2001; Zhang et al. 2003). Figure I-21 shows the *E. coli* ACP structure.



**Figure I-21.** Acyl carrier protein (ACP) nuclear magnetic resonance (NMR) structure from *E. coli*. ACP is a small protein in charge of exchange intermediaries of the synthesis in FAS and PKS systems. ACP structural motifs are highly conserved in nature and consist of a four  $\alpha$ -helix bundle indicated with arrows in the left panel of the figure. A serine residue at the N terminus of the helix II is subject to post-translational modification that transform the apo- form into the holo- form. This modification consists in the transfer of a 4'-phosphopantetheine (PPT) moiety from coenzyme A (CoA) to the serine. The activity of ACP is dependent on this conversion from the inactive apo- to the active holo- form. Figure adapted from (Chan & Vogel, 2010).

As it was introduced before in this manuscript, the ACP does not bind directly the FA. It has to be activated with a 4'-phosphopantetheine prosthetic group (PPT) that acts as a linkage between the ACP and the FA. In this way the ACP can exist in two forms, the apo-ACP, that is not activated so there is no presence of the PPT arm, and the holo-ACP, totally active and able to be malonylated (Evans et al. 2008; Morris et al. 1993). The active form was described to show some dynamic and structural differences compared to the inactive form (Kim, Kovrigin, and Eletr 2006). The PPT is then attached to the protein via a phosphodiester linkage to the serine of the Asp-Ser-Leu conserved motif at the end of the second  $\alpha$ -helix and has a terminal sunfhydryl group that is going to be covalently bound to the FA intermediates during the synthesis (White et al. 2005). Right panel of figure I-21 shows a representation of the ACP PPT arm.

### 1.5.3. The condensing domains: KS and AT

For a better understanding of the structural biology of the FAS and PKS systems, their study will be dissected in the same way than nature has done: separating the condensing domains from the modifiers. To start each cycle of elongation the holo-ACP must be correctly charged with malonyl extender units (or any other extender blocks in the case of more complex PKS synthesis) and they should be accessible for the keto synthase (KS) domain that perform the condensation rounds. It is convenient to analyze the biology of both domains in the same section since in all solved mega-structures both are in close contact and most of the times they are even transcribed as a unique biochemical block. Both domains together with the ACP are known as the condensing unit as they are able to perform condensation round by themselves.

The transfer of the malonyl moieties to the terminal sulfhydryl group of the ACP's PPT is carried out by an AT domain. First, the AT binds the malonyl moiety from a malonyl-CoA molecule transferring it to the active serine. This serine together with two histidines form the catalytic triad that acts as a general base/acid catalyst in the acyl transfer reaction (Liew et al. 2012). After this first binding step, the Coenzyme A is released from the reaction chamber and the malonyl moiety is transferred from the serine to the terminal sulfhydryl of the ACP's PPT in a reversible reaction. Normally ATs that are part of FA synthases show strong specificity for malonyl, but analogous proteins within PKS pathways can generate different products by using different extender units due to different specificities. Although it seems that the AT modules are in charge of loading the extender units to the ACPs, it has been shown that the ACP domains can also exhibit intrinsic acyltransferase activity in some PKS. This activity is completely necessary in some type II PKS in which the AT module is missing (Arthur et al. 2005; Hitchman et al. 1998).

The KS condensation acts when all the ACPs are activated and with their specific extender units loaded. The first reaction in FAS synthesis use acyl-CoA as a primer that is condensed with malonyl-ACP to form acetoacetyl-ACP. Successive condensations occur between the malonyl and the acyl-ACP growing chain. In bacterial systems that lack a specific AT domain with specificity for acetyl-CoA, FabH, that is a KS protein, can directly accept acetyl-CoA as starter substrate. Further acyl chain extensions from C4 to C14 and from C14 to C16 are carried out by other KS proteins, FabB and FabF (Maier, Leibundgut, and Ban 2008a) respectively. Moreover, in iterative FAS as the ones from fungi and mammals only one KS is required for FA synthesis simplifying in this way the system but reducing the enzymatic versatility. On the other hand, modular PKS contain specialized KS that show specificity for certain substrates but can accept a wide range of substrate lengths (Smith and Tsai 2007).

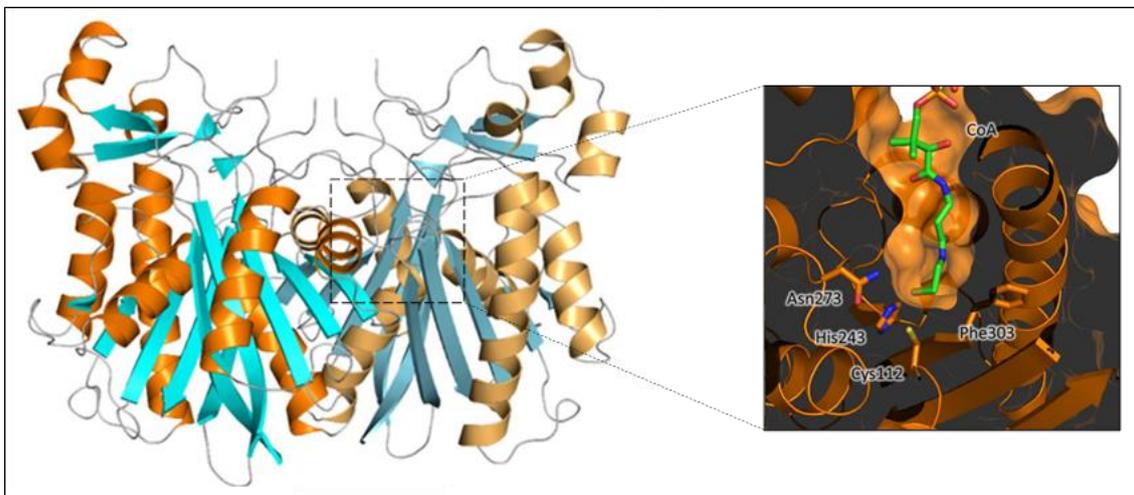
KS and AT structures belonging to FAS type II systems, have been published as individual proteins, while PKS and FAS type I KS-AT structures have been solved as unique two-domain polypeptides. Mammalian FAS and the PikAIII PKS module, whose structure were previously shown in figure I-20 are good examples of megastructures with the both domains assembled together as homodimers by the side face of the KS domains (Maier, Leibundgut, and Ban 2008a). To our knowledge, all known KS-AT structures contain a linker that have two  $\alpha$ -helices facing the KS and three antiparallel parallel  $\beta$ -sheets on the AT side acting as adapter between both domains and prevents direct interaction between them (Maier, Leibundgut, and Ban 2008a). This linker was also studied in KS-AT domains from PKS systems, like the one from the 6-deoxyerythronolide

B synthase (DEBS) that was described to influence the KS substrate specificity due to its proximity to the ACP docking site that is where the reaction intermediaries between both active centers are exchanged (Tang et al. 2007; Tang et al. 2006).

The biochemistry and the structural biology of individual KS and AT domains from FAS type II systems have been extensively studied demonstrating that their architectures are similar to the ones found in di-domain polypeptides. Examples of structures available in the literature of KS proteins from *E.coli* are FabH (Davies et al. 2000; Qiu et al. 2001; Qiu et al. 2005), FabB and FabF (Bagautdinov et al. 2008; Moche et al. 1999; J. G. Olsen et al. 2001; Price et al. 2001) and AT proteins like FabD (Serre et al. 1995).

### 1.5.3.1. The keto synthase domain

The condensing KS enzymes are part of the thiolase superfamily, that shows a core structure consists of two  $\beta\alpha\beta\alpha\beta\alpha\beta$  motifs related by pseudo dyad duplication (Hopwood and Sherman 1990; White et al. 2005; White et al. 2005). The catalytic triad involved in this reactions typically has a cysteine where the acyl chain is going to be transiently bound at the time of decarboxylation, and two histidines as the case of FabB in *E.coli* or an Asn/His in some bacterial FAS type II KS domains like FabH (Meurer and Hutchinson 1995). The active cysteine is located at the bottom of a tunnel that measure about 20 Å deep and 5 Å in diameter. This cysteine and the histidines of the active site are found in equivalent locations of the repeated structure of the protein core that comprise the top two thirds of the protein. It was recently described the involvement of other residues in the decarboxylation, translocation, C-C bond formation and in the coupling of decarboxylation and formation of the C-C bond (Robbins et al. 2016). Figure I-22 shows a cartoon representation of the protein structure of a typical KS domain as well as a zoom of the active center more important residues.

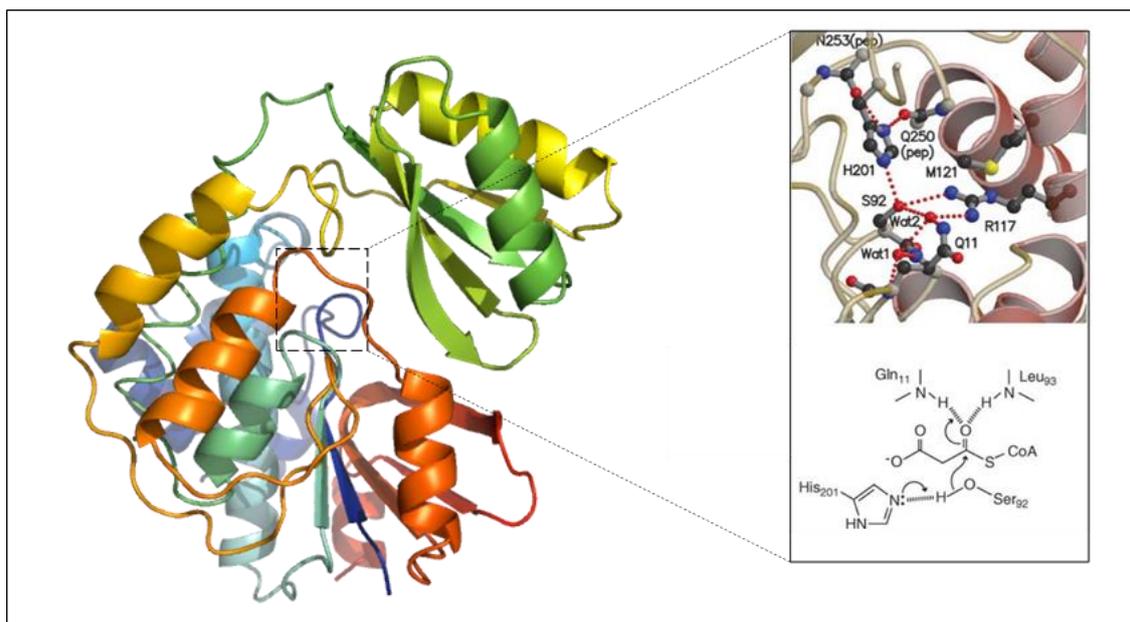


**Figure I-22.** The structure of *Yersinia pestis* FabH (YpFabH) as a model for the KS domains. Figure shows a cartoon representation of YpFabH as a dimer on the left panel that shows a canonical thiolase fold.  $\beta$ -sheets are displayed in cyan and  $\alpha$ -helices are displayed in orange. The right panel represents a zoom of one of the active centers of the structure, in which the catalytic triad, formed by a Cys-His-Asn, can be seen. Other KS domains present a second His instead of the Asn residue. A superposition of CoA (green) also shows the phosphopantetheine arm of CoA extending the length of the binding pocket to reach the active site cysteine. Figure adapted from (Nanson, Himiari, Swarbrick, & Forwood, 2015).

KS domains are in close contact with ACP due to the constant exchange of intermediaries between them and their direct interaction has been proved many times (Bruegger et al. 2013; McDaniel et al. 1997; Ranganathan et al. 1999). In fact, it was postulated that the protein-protein interactions and not the substrate recognition dominates the turnover of chimeric assembly line polyketide synthases (Klaus et al. 2016). In FAS it was demonstrated that the transfer of saturated acyl moieties from the PPT thiol to the active site cysteine thiol is an inherent property of the KS domain (Witkowski, Joshi, and Smith 1997). Thus the KS domains seem to be active elements during all the biosynthetic route and not only during the condensation step.

### 1.5.3.2. The Acyl transferase domain

The acyl transferases proteins have an  $\alpha/\beta$ -hydrolase core fold with a central, four-stranded parallel  $\beta$ -sheet surrounded by 12  $\alpha$ -helices and a ferredoxin-like subdomain composed of four antiparallel  $\beta$ -sheets and 2  $\alpha$ -helices, which together form the active site cleft (Serre et al. 1995). They share a conserved active site with a catalytic Ser-His dyad where the catalytic serine is always found in a turn between a  $\beta$ -strand and an  $\alpha$ -helix forming a hydrogen bond with the cognate histidine. The imidazole side chain of the histidine orientates the serine residue for the substrate attack and in this way the serine can destabilize the labile thioester bond of the incoming substrate, which leads to the formation of an acyl-enzyme complex and free CoASH (Oefner et al. 2006). Figure I-23 shows the cartoon representation of the structure of FabD from *E.coli* as well as a zoom with a detail of the active center residues.



**Figure I-23.** The acyltransferase FabD (*E.coli*) active site and catalytic mechanism. Figure shows the three dimensional structure of the FabD protein active site. FabD core shows an  $\alpha/\beta$ -hydrolase fold with an alpha/beta sheet of eight beta-sheets connected by alpha-helices. Right panel shows a zoom of the residues involved in the catalysis and the hydrogen bond interactions involving the active site Ser92 and the His201, that together form the catalytic dyad. His201 assists in the abstraction of a proton from Ser92. Ser92 also interacts with a tightly bound water (Wat2) that is held in place by interactions with Arg117, Gln11, and Wat1. Arg117 and Gln11 interact with the malonyl carboxylate of the incoming substrate, which displaces the water molecules. The overall reaction occurs via a ping-pong mechanism. The first step is the transfer of malonate from malonyl-CoA to the active Ser92. His201 activates Ser92 for nucleophilic attack on the incoming thioester. CoASH is then released from the enzyme and followed by ACP binding (White et al., 2005).

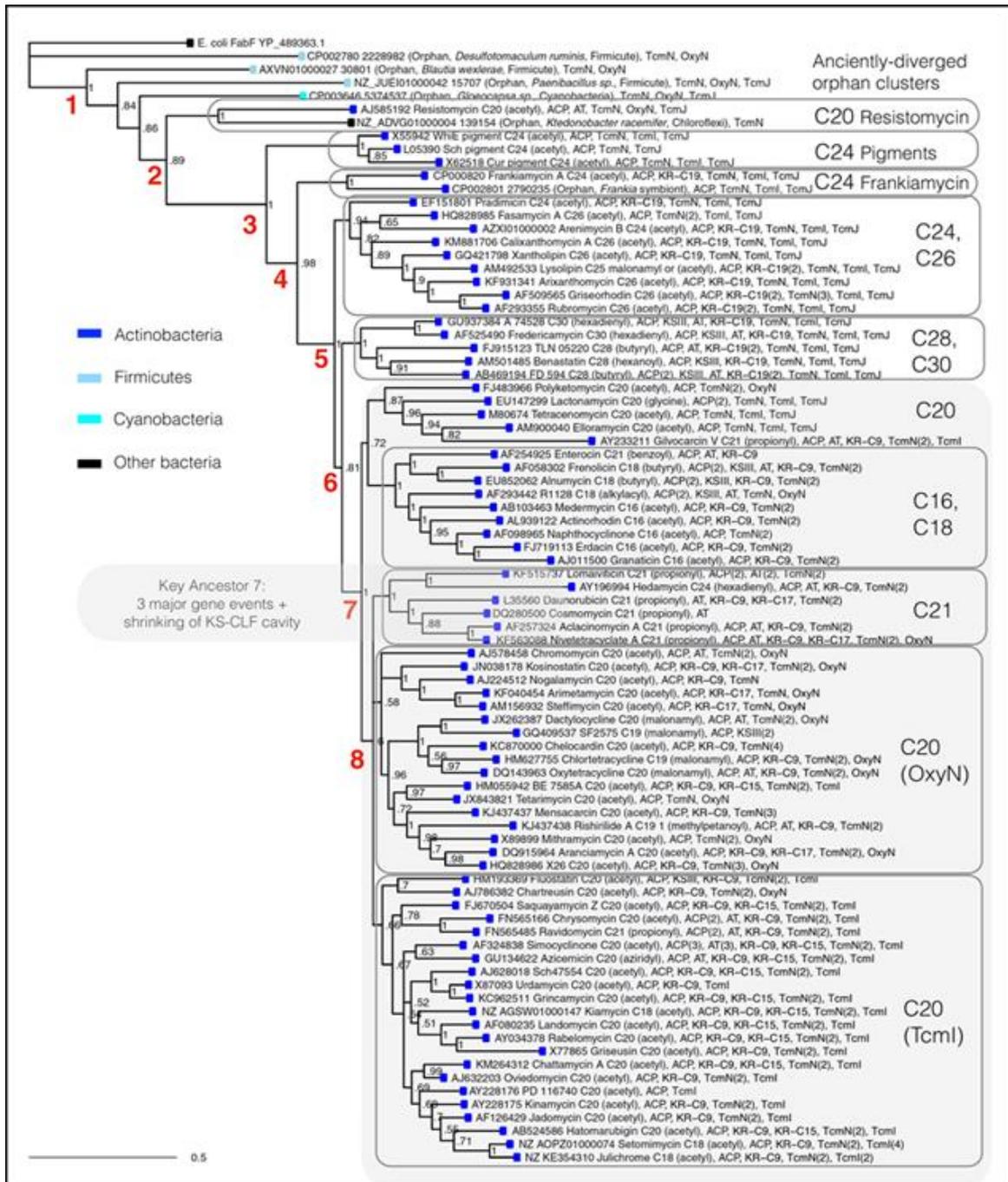
There are some particular residues that were described to affect the selectivity of the AT proteins, for example, AT that show specificity for malonyl (malonyl acyl transferases or MAT) usually have an arginine within the active site (Jenni et al. 2007) that forms a bidentate salt bridge with the malonyl carboxylate (Oefner et al. 2006). Some experiments show that simply changing this arginine residue for an alanine, the specificity change and the protein cannot accept malonyl anymore (Reeves et al. 2001). In PKSs, the methylmalonyl-CoA is the extender unit most commonly used by AT domains (Katz and Donadio 1993). When an AT domain shows methylmalonyl-CoA specificity it is usual to find a YASH motif in their sequence, while a HAFH motif is found when the AT domain malonyl-CoA specific (Del Vecchio et al. 2003; Reeves et al. 2001; Sundermann et al. 2013). It is generally accepted that more hydrophobic active centers promote enzymatic promiscuity in AT domains. Within the oxyanion hole of AT domains there are always key residues that interact with the carboxylate group of the extender unit and form hydrogen bonds and salt bridge interactions with it. For instance in FabD of *E.coli*, Arg117, Gln11, and Gln63 are the amino acids that carry out this substrate recognition and avoid the water to achieve the hydrophobic cavity of the protein, making possible the existence of transient but stable acyl-AT intermediates (White et al. 2005).

### 1.5.3.3. The Chain length factor domain

The chain length factor (CLF) is a KS-like domain found in some iterative PKS always nearby a KS domain with which forms a dimer called KS-CLF. It has a special relevance in PKSs that produce long chain polyketides (more than 16 carbon atoms) since in these circumstances KS-CLF is always present within the protein sequence. KS-CLF was proposed to catalyze the decarboxylation of malonyl units and elongate the polyketide chain in PKS type II, being responsible of condensing extender units sometimes even bigger than malonyl or methylmalonyl. In this way, it is a central protein for some polyketide synthesis due to its ability to synthesize the desired products when incubated with malonyl-CoA and only with the help of the ACPs (Bao, Wendt-Pienkowski, and Hutchinson 1998; Carreras and Khosla 1998; Matharu et al. 1998).

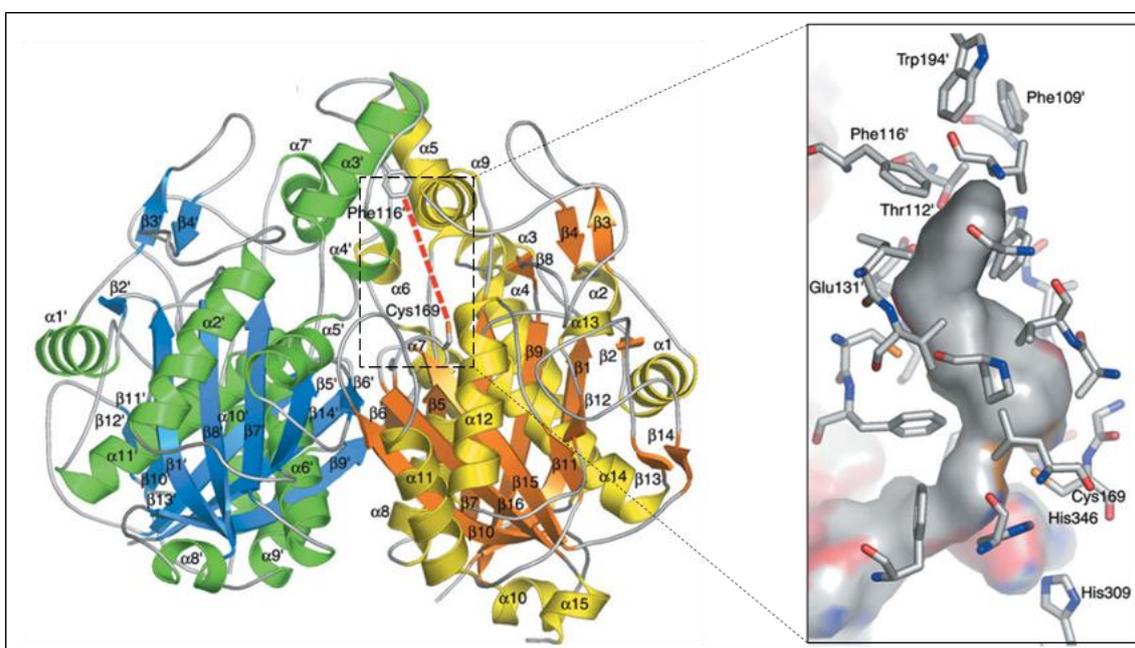
The KS-CLF di-domain protein is composed of two KS domains referred as KS $\alpha$  and KS $\beta$  or KS-CLF di-domain. In contrast to homodimeric KS found in fatty acid synthase from FAS or PKS systems the active site cysteines are absent from the CLF subunit of this heterodimer. The CLF has always a glutamate or an aspartate replacing the regular cysteine present in a normal KS domain, and this residue is the one that show decarboxylation activity against malonyl and is able to prime the cognate KS subunit to start the synthesis (Bisang et al. 1999). Some studies showed that KS-CLF can be primed by medium-chain-length CoAs more efficiently than malonyl-coA, giving rise to different polyketide molecules. This means that the KS-CLF prefers non-acetate priming over decarboxylative priming. This results suggest that KS-CLF could act sometimes condensing mature and longer extender units at the last steps of the synthesis (Tang, Koppisch, and Khosla 2004; Tang, Tsai, and Khosla 2003a).

Both KS domains from KS-CLF proteins are descendants of a common ancestor arose from a genetic duplication (Ridley, Lee, and Khosla 2008). Phylogenetic analysis of characterized CLF genes shows that they are grouped by the length of the chain they elongate. The amplitude of their hydrophobic cavity is also correlated with the polyketide length, being able to bear larger molecules when they have bigger cavities (Hillenmeyer et al. 2015). A possible explanation for this ability to control the chain length is the presence of a hydrophobic tunnel that connects both active centers (from KS and CLF subdomains). The nascent molecule supposedly grow attached to the KS active cysteine and some amino acid residues that in contact with the nascent molecule determine the moment when the polyketide has to be ejected (McDaniel et al. 1993; Tang, Tsai, and Khosla 2003a). The hydrophobic tunnel presumably protects the reactive radicals of the new polyketide against the cellular oxidizing environments avoiding in this way possible undesired reactions. An evidence of this can be that some of the genes responsible for cyclization arose at the same time as the KS-CLF dimers, enabling PKSs to diversify the chain length, oxidation state, and overall shape of their molecular products (Hillenmeyer et al. 2015; Ridley, Lee, and Khosla 2008). Figure I-24 shows a phylogeny of chain length factor genes where can be seen how they cluster together according to the number of carbon that their product have.



**Figure I-24.** Phylogeny of chain length factor proteins. Phylogenetic tree of 78 CLF protein sequences from selected genes. Accessory genes identified in the same gene cluster (within 30 kb) as the CLF are shown at each leaf. Leaf colors represent phylum of origin. Node support for the CLF phylogeny is shown as Bayesian posterior probabilities. CLF domains cluster according to the number of carbons the product of the cluster to which they belong produce. Figure taken from (Hillenmeyer et al., 2015).

KS-CLF heterodimers are composed of two KS domains whose folding is exactly the same as the rest of known KS. Both KS are faced together forming central interface in a vertical plane where the inter-domain interactions maintain the structure bound. The actinorhodin KS-CLF heterodimer whose structure is shown in figure I-25 is the only KS-CLF structure solved to date. KS-CLF are able to polymerize and control the backbone length in the synthesis in a way that is still not fully understood (Keatinge-Clay et al. 2004; Szu et al. 2011). The actinorhodin KS-CLF structure provided the first direct measurements of a CLF hydrophobic tunnel amplitude whose length was  $\sim 17$  Å in length at the heterodimer interface and is where the first cyclization of the polyketide occurs (Keatinge-Clay et al. 2004). In addition to cysteine and glutamate, two histidines are normally found in each of the active sites forming a catalytic triad and it was demonstrated that at least one of them is strictly necessary for a complete catalysis (Dreier and Khosla 2000).



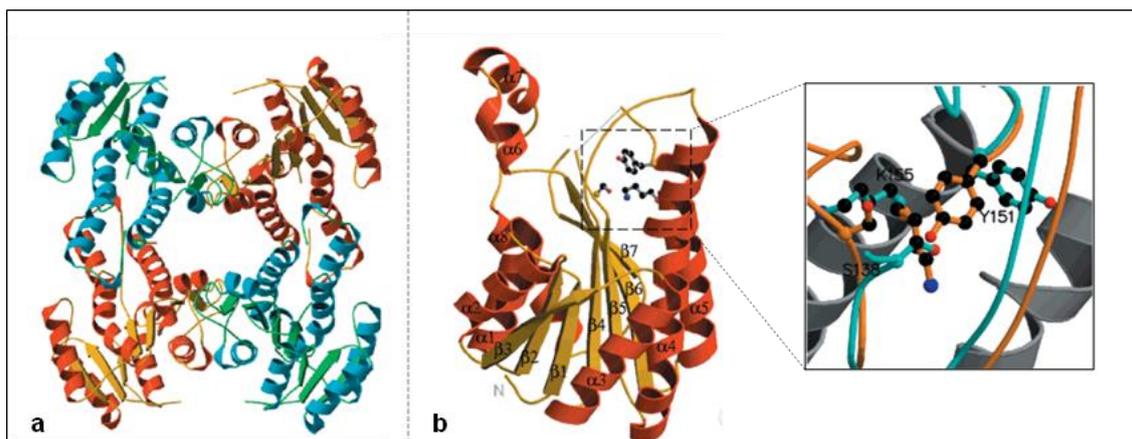
**Figure I-25.** The polyketide tunnel of keto synthase (KS)-chain length factor (CLF) proteins. The structure of the actinorhodin KS-CLF produced by *Streptomyces coelicolor* is shown in the figure. The thiolase fold can be observed in each of the two KS-like domains as in the case of similar domains of FAS systems. The KS domain is coloured in orange (beta sheets) and yellow (alpha helices) where the CLF domain coloured in blue (beta sheets) and green (alpha helices). Right panel zoom shows the residues that form the tunnel surface where the intermediate molecules presumably grow. Tunnel volume predictions suggest that the first cyclization could be carried out inside the cavity in this particular case. The catalytic triad is formed in a similar way as in KS domains of FAS systems: the cysteine, where the growing molecule is covalently bound and the two histidines. Figure adapted from (Keatinge-Clay et al., 2004).

## 1.5.4. The modifying domains

In addition to the condensation performed by the KS and AT domains in coordination with the ACP, a coordinated action of the chain-modifier domains keto reductase (KR), dehydratase (DH) and enoyl reductase (ER) is essential in both PKS and FAS systems. In FA synthesis, a complete reduction is necessary whereas in PKS the reduction step is variable. During the synthesis of FA the keto-group generated in each round of condensation first needs to be reduced by a KR domain (*FabG* in *E.coli*), that is typically an NADPH-dependent 3-ketoacyl-ACP reductase. Then, the hydroxyl group is removed by a DH domain (*FabZ* and *FabA* in *E.coli*), a 3-hydroxyacyl-ACP dehydratase that generate a trans or cis double bond. The final reduction is catalyzed by an ER (*FabI* in *E.coli*), to produce an acyl-ACP, which can serve as a substrate for another round of elongation or, if it has reached the expected chain length, the FA will be ready for being released. The enoyl reduction does not occur when the double bond is in a cis configuration and this is very important to be able to generate FA with unsaturations at different positions.

### 1.5.4.1. The Keto Reductase domain

During each condensation round the KS domain forms a covalent bond between one extender unit and the acyl-ACP. In this process, a keto group is introduced in the growing chain and will need to be reduced to a hydroxyl moiety. This reduction is performed by a NADPH-dependent keto reductase (KR) domain (*FabG* in *E.coli*) (Beld, Lee, and Burkart 2014). Surprisingly little has been published on FAS KR despite that they are essential genes for their implication in the synthesis of FA (Zhang and Cronan 1998). KR and enoyl-ACP reductase (ER) sequences show similarities and can be easily mistaken, but ER can be distinguished from KR by the presence of a Y-(X)<sub>6</sub>-K motif and the absence of the S-Y-K motif characteristic of the short-chain alcohol dehydrogenase (SDR) family (Jörnvall et al. 1995). It is known that KR are active on a large range of intermediate molecules with chain lengths varying between 4 and 10 carbons (Toomey and Wakil 1966; Weeks and Wakil 1968). In PKS systems, the KR domains are similar to the ones found in FAS. In PKS the stereospecificity is very important because chiral centers may be introduced by the use of branched extender units and by the reduction of ketone groups to alcohols (Caffrey 2003).

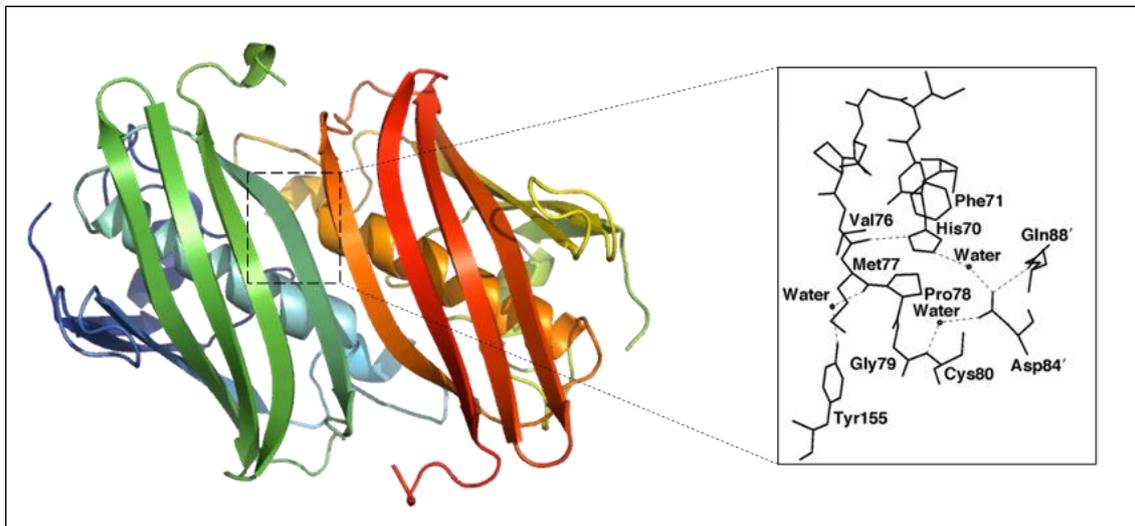


**Figure I-26.** Protein structure of the keto reductase FabG from *E. coli* and its active site. Panel (a) shows the FabG tetramer organization. Monomers were colored with alternating color schemes for clarity. All beta sheets are shown in yellow or green, and alpha helices in red or blue. Panel (b) show a single monomer with a zoom showing the catalytic residues, Ser 138, Tyr 151, and Arg 155. Figure adapted from (Ac, Ym, Co, & Sw, 2001).

The first structure of a KR solved was the KR of the plant *Brassica napus* (Fisher et al. 2000). Later on, the structures of FabG from *E. coli* (Ac et al. 2001) and *Mycobacterium tuberculosis* homologous MabA (Subramanian et al. 2011) were determined using X-ray crystallography. Their architecture is virtually identical to the one found in PKS like the one from the KR domain of the erythromycin synthase (Keatinge-Clay and Stroud 2006). All KR retain similar characteristics like the need of NADPH for being active (Alberts et al. 1964; Lai and Cronan 2004). KR are monomeric domains of about 25 kDa, but can exist as tetramers with two types of dimerization in the case of the FabG crystal forms (Ac et al. 2001; Fisher et al. 2000). KR domains belong to the SDR family of enzymes, whose members catalyze a broad range of reduction and dehydrogenase reactions using a nucleotide cofactor (Jörnvall et al. 1995). They show a typical Rossmann fold structure (Jörnvall et al. 1995), with a twisted, parallel  $\beta$ -sheet composed of seven  $\beta$ -strands that are flanked on both sides by eight  $\alpha$ -helices. The protein core consists of a repetition of a  $\beta\alpha\beta\alpha\beta$  motif and the active site residues (Ser, Tyr, and Lys) are clustered together near the loop connecting the  $\beta_5$  and  $\alpha_5$ . Figure I-26 shows the structural arrangement of the FabG tetramer as well as its monomeric form and the details of its active center. It was demonstrated that the binding of NADP<sup>+</sup> is associated with significant conformational changes (White et al. 2005). The tetramer from *B. napus* KR shows negative cooperativity in binding NADPH, which is further enhanced by the presence of ACP (Ac et al. 2001) but the structures of KR currently available do not explain the binding of ACP and how this influences cooperativity.

### 1.5.4.2. The Dehydratase/Isomerase domain

The next step in the elongation cycle of FA and PKS is the dehydration of the hydroxyacyl-ACP previously generated by the KR domain to the trans-2-enoyl-ACP. There are two genes, *fabA* and *fabZ*, encoding dehydratases (DH) that function in the type II FAS system of *E. coli* and can perform this reaction and, in some cases, the subsequent cis-isomerization of the double bond (FabA). FabZ functions in the metabolism of both saturated and unsaturated FA, whereas FabA most efficiently processes 10-carbon acyl chains to introduce unsaturations (Heath and Rock 1996). The trans-double bonds that FabZ generate can be subsequently reduced to single bonds by an enoyl reductase (ER), and in this way promote the synthesis of saturated fatty acids. In contrast, the cis double bond generated by FabA is retained through the further cycles of fatty-acid elongation. PKS DH domains have been poorly studied compared with other domains maybe because of the simple nature of the chemical reaction they catalyze and the lack of a convenient assay to measure substrate turnover. However some examples can be found in the literature that show their similarity with FAS DH domains in terms of their structure and the biochemical reactions they perform (Herbst et al. 2016; Li et al. 2015).



**Figure I-27.** Protein structure of the dehydratase dimer FabA from *E. coli* and its catalytic mechanism. Dehydratase is a symmetric dimer with an unusual  $\alpha+\beta$  'hot dog' fold. Each of the two independent active sites is located between the two subunits of the enzyme. The active sites are located in a tunnel-shaped pocket completely isolated from the general solvent. Right panel shows an scheme of the catalytic region and substrate-binding site. Only water molecules in hydrogen-bonding contact (dashed lines) with the catalytically important region of the pocket are shown here. Side chains of histidine from one subunit and aspartic acid from the other are the only potentially reactive protein groups in the active sites. Figure adapted from (Leesong, Henderson, Gillig, Schwab, & Smith, 1996).

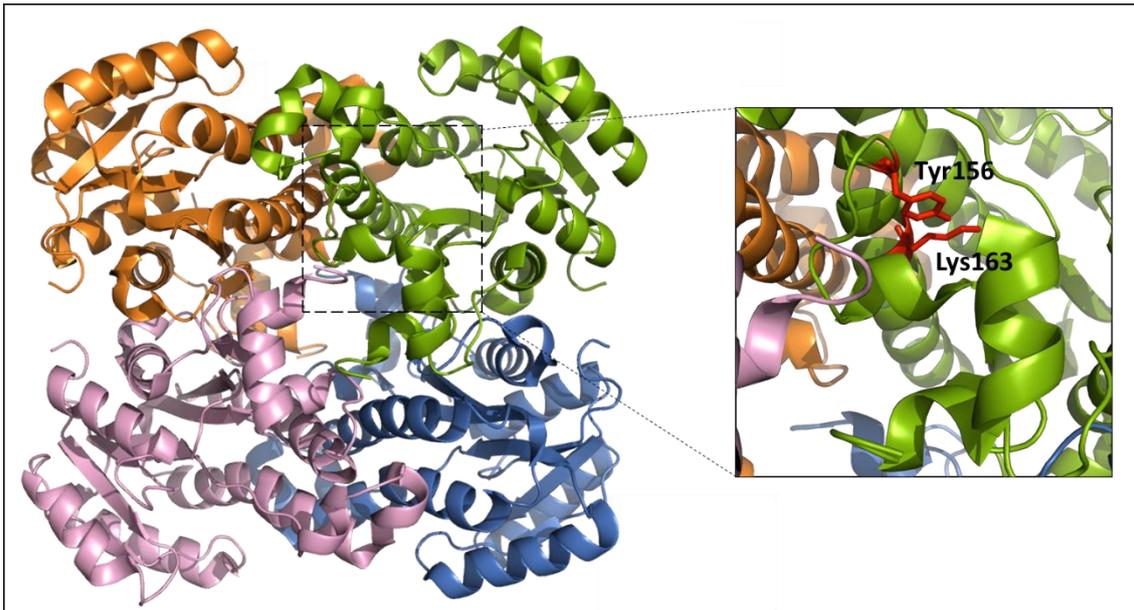
Structures from FabA from *E. coli* (Leesong et al. 1996) and FabZ from *Pseudomonas aeruginosa* (Kimber et al. 2004) have been solved. As can be anticipated from their homology at the primary sequence level, both adopt the same "hot dog" fold (Dillon and Bateman 2004) as is shown in a cartoon representation in figure I-27. In fact, FabA was the first example of the "hot-dog" fold in which each of the monomers have a long central  $\alpha$ -helix wrapped by a six-stranded antiparallel  $\beta$ -sheet (Leesong et al. 1996).

The monomers of both DH dimerize with the  $\beta$ 3-strand interacting with the last two turns of the central  $\alpha$ -helix by packing against each other in an antiparallel fashion. The active site is formed along the dimer interface with the critical His and Asp/Glu active site residues contributed by different monomers. These DH domains are symmetric dimers where each of the two independent active sites are located between the two subunits of the enzyme, and they have a tunnel-shaped pocket completely isolated from the general solvent. Both the dehydration and isomerization reactions seem to occur in the same active site. (Leesong et al. 1996)

#### **1.5.4.3. The Enoyl Reductase domain**

The enoyl reductase (ER) is the enzyme in charge of the last reduction step of the FA synthesis. ER removes the trans-double bond previously introduced by a DH domain and thereby generates a fully saturated moiety. The action of the ER is thereby necessary for the achievement of a correct pattern of double bonds. In FA synthesis most of the biochemical reactions described before are reversible but condensations performed by KS and the subsequent enoyl reductions are not. In this way the ER pull forward the reversible reactions of the FA condensed intermediaries for the final acyl chain moiety completion. The reduction performed by the ER domains are also fundamental in PKS synthesis to generate essential intermediaries.

The structures and sequences of all FAS and PKS domains have been conserved in nature. However, this conservation is not seen within the ER domains. An impressively wide diversity of proteins that catalyze this reduction step are found in nature (Massengo-Tiassé and Cronan 2009). In contrast to all other functional domains of the FA elongation cycle, the ER show different folds in the solved structures of both bacterial FAS system (type I and II). ER from type II FAS belong to the SDR family and have only one domain in their structure (Jörnvall et al. 1995) but mammalian and fungal FAS establishes a subfamily of medium-chain dehydrogenases/reductases (MDRs) that contain two subdomains instead of one (Maier, Leibundgut, and Ban 2008a; Nordling, Jörnvall, and Persson 2002).



**Figure I-28.** Protein structure of the enoyl reductase FabI from *E. coli*. The *E. coli* FabI is a tetramer made up of four single domains showing the Rossmann fold configuration identical to the one previously shown for FabG (the KR domain). Each subunit is represented in a different color for clarity. Although there are relatively extensive interactions between each of the chains of the tetramer, each active site is formed entirely by one domain as is shown in the right panel zoom for the green domain. The catalytic dyad Tyr156-Lys163 is part of the highly conserved SDR motif, Tyr-X6-Lys which is the fundamental difference between FabI and FabG.

FabI is the *E. coli* FAS II ER protein and its structure was determined by X-ray crystallography more than 20 years ago (Baldock et al. 1996; Baldock et al. 1998). Homologs of this structure were lately determined, like the one from *M. tuberculosis* (InhA) (Cohen-Gonsaud et al. 2002), demonstrating that all ER structures from FAS type II are virtually identical. They are homotetramers that have a highly conserved SDR motif, Tyr-X6-Lys. Each monomer of the tetramer is composed of a seven-stranded parallel  $\beta$ -sheet flanked on each side by three  $\alpha$ -helices with a further helix lying at the C terminus of the  $\beta$ -sheet. (Baldock et al. 1998). On the contrary, mammalian and fungal FAS type I ER domains are slightly different as they contain two subdomains, a nucleotide binding domain that shows a conserved Rossmann-fold and a substrate binding portion. In ER proteins with this arrangement the NADP<sup>+</sup> cofactor is bound between both subdomains (Maier, Leibundgut, and Ban 2008a). Figure I-28 shows a cartoon representation of the enoyl reductase FabI from *E. coli* (with a structure typical of the type II FAS ER domains) with a detailed zoom in its active site.

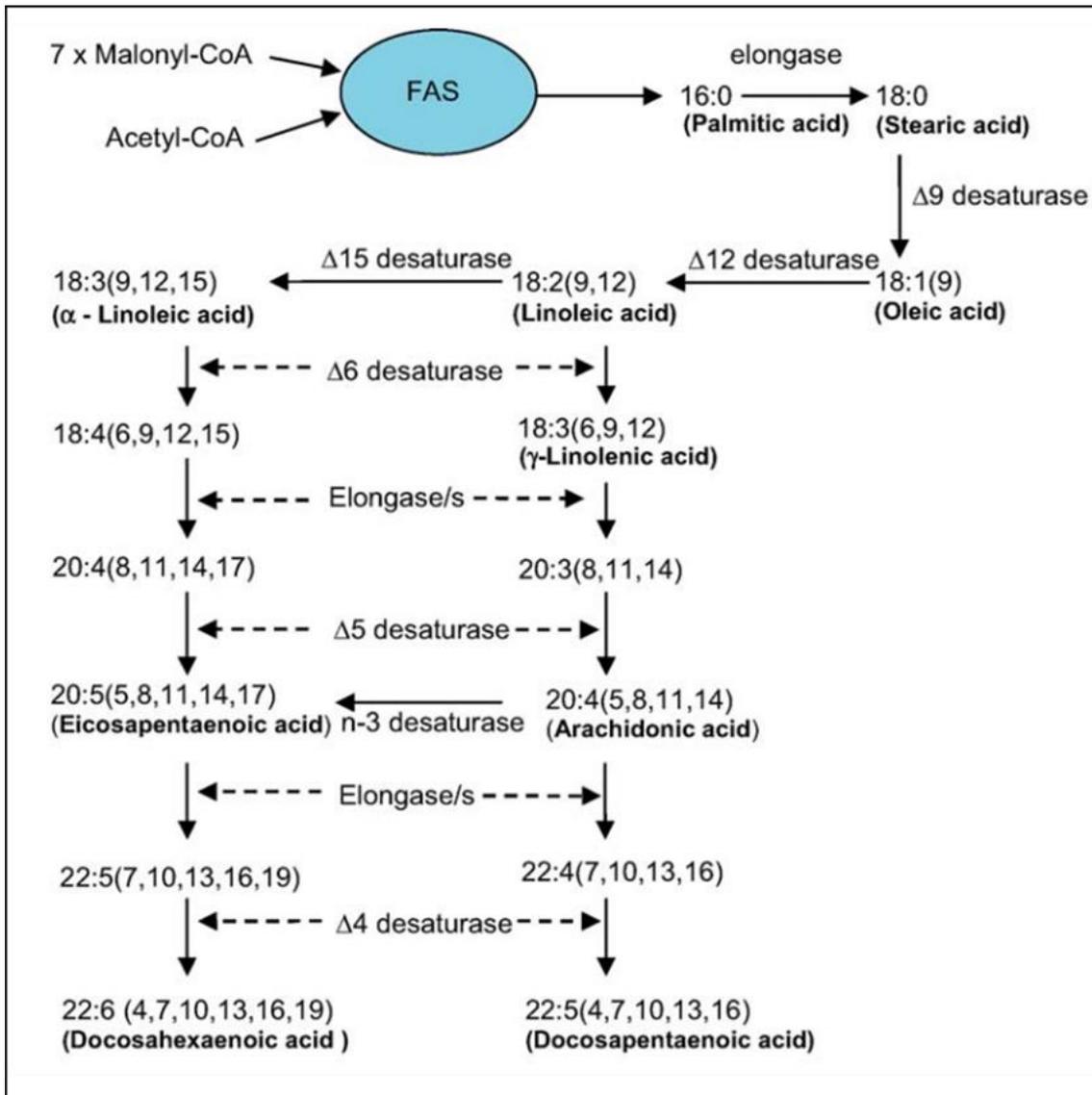
A tyrosine and a lysine have been proposed as the catalytic residues for ER from FAS type II systems, but an asparagine is replacing the tyrosine in ER from mammal FAS. Here we will just exemplify the mechanism of action of the bacterial ER since they have been better described in the literature. The lysine of the catalytic dyad was proposed to stabilize the binding of the NADP<sup>+</sup> cofactor through hydrogen bond interactions with the hydroxyl groups of the nicotinamide ribose moiety. The tyrosine acts as the proton donor for the second carbon of the trans-2 unsaturated chain. The reduction of the double bond is thought to proceed by conjugate addition of a hydride ion from the NADPH to the carbon 3 of the acyl group with the intermediate formation of an enzyme-stabilized enolate anion on the carbonyl oxygen of the first carbon.

## 1.6. Omega-3 Fatty Acid synthesis

Omega-3 FA are polyunsaturated FA (PUFAs) whose chain have 18 carbons or more and contain at least two methylene-interrupted double bonds in the cis position. FAS production systems are not able to synthesize these large and reduced molecules and for that reason they have been traditionally included into the general term of “secondary lipids” as they are not present in the classical metabolism pathways. The term “omega-3” refers to the position of the first unsaturation that is closer to the functional carboxylic acid group. Thus an omega-3 FA always have a site of unsaturation between the third and fourth carbon from the omega end. There are three major types of omega-3 FA (that were previously introduced in the first section of this manuscript) that are ingested in foods and needed for the correct function of the body: alpha-linolenic acid (ALA, 18:3n-3), eicosapentaenoic acid (EPA, 20:5n-3), and docosahexaenoic acid (DHA, 22:6n-3).

### 1.6.1. The aerobic pathway: desaturases and elongases

The best-known pathway of PUFA synthesis is found in plants and lower animals and is carried out by special enzymes called desaturases and elongases. They modify the regular FA produced by the classical lipid routes (FAS synthases) elongating their chain and introducing new double bonds to generate longer FA with more unsaturations. Elongases are condensing enzymes that incorporate malonyl units to the previously generated FA. Desaturases introduce double bonds at specific positions on the acyl chain, thereby influencing the biophysical properties of the fatty acid itself. The action of several desaturases that introduce double bonds in different positions is necessary to produce a fatty acid with a structure like DHA or EPA. Mammalian cells for instance express  $\Delta 9$ ,  $\Delta 6$  and  $\Delta 5$  desaturases (where the  $\Delta$ -number indicates the position at which the double bond is introduced) but they lack  $\Delta 12$  or  $\Delta 15$  activities that are necessary for the *de novo* production of some essential PUFAs. For that reason, mammalian production of DHA and EPA is deficient so that they need to consume them in the diet (Lee et al. 2016). Figure I-29 is a scheme of the aerobic pathways for the formation of polyunsaturated fatty acids (PUFAs).

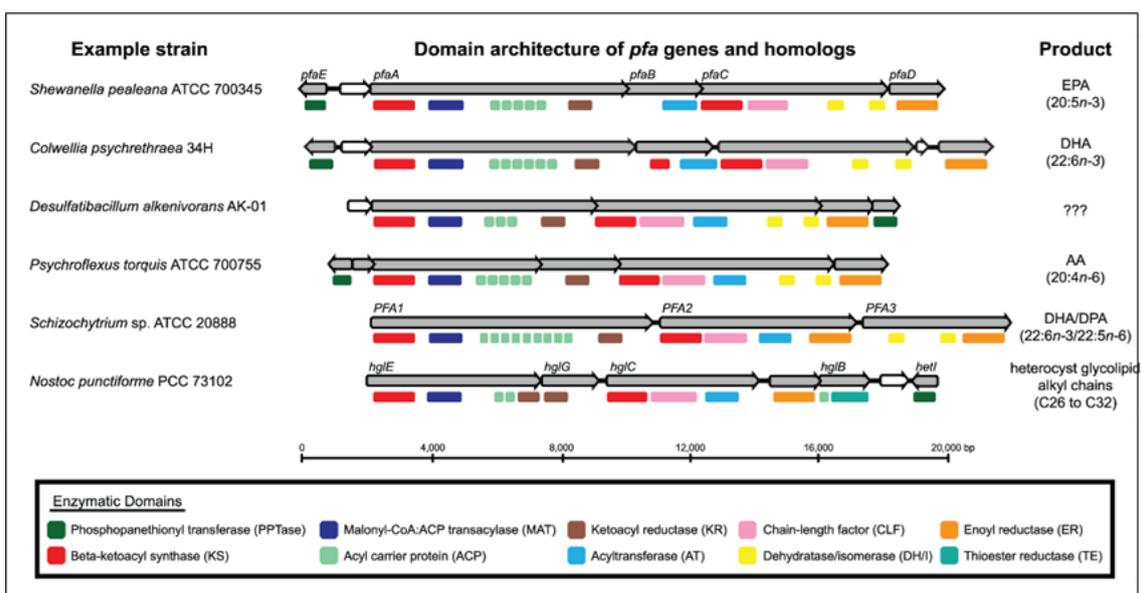


**Figure I-29.** Pathways for the formation of polyunsaturated fatty acids (PUFAs) using the aerobic FAS route. Fatty acids are synthesized from acetyl-CoA and malonyl-CoA using the FAS complex of enzymes. The saturated fatty acid, stearic acid C18:0 is then successively desaturated and elongated through a series of reactions leading to the formation of various PUFAs (the n-3 and n-6 series). Figure taken from (Ratledge, 2004).

Desaturases function as oxygenase enzymes using activated molecular oxygen to remove hydrogens from an acyl chain and for that reason the route is known as the aerobic route for PUFA synthesis. To have an idea of the huge complexity of this aerobic route it should be remarked that the synthesis of DHA from acetyl-CoA requires approximately 30 distinct enzyme activities and nearly 70 biochemical reactions (Metz et al. 2001). Due to the amount enzymes and regulators that come into play in this biosynthesis it is difficult to manage them to make genetic constructions to perform bioengineering experiments and develop suitable platforms to produce PUFA in the laboratory.

## 1.6.2. The anaerobic pathway: pfa synthases

The previously described pathway of PUFA synthesis that involves the processing of saturated products of FAS enzymes by elongation and aerobic desaturation is not the only way by which they can be produced in nature. Other pathways that are able to produce Omega-3 FA starting from basic molecules have been discovered. It is known that some subset of marine gamma-proteobacteria like *Shewanella pealeana* or *Moritella marina* are able to produce very long PUFAs (Nichols et al. 1999) using the proteins generated with gene clusters that are totally independent from genes of the aerobic PUFA synthesis. These gene clusters possess high homology with enzymes involved in FA synthesis. Analysis of the first genomic cluster of this type from *Shewanella pealeana* identified five open reading frames (ORFs) that are necessary and sufficient for EPA production in *E. coli* (Yazawa 1996a). The recombinant production of DHA in *E. coli* has also been lately reported using a homolog gene cluster obtained from *M. marina* (Orikasa et al. 2006). These alternative anaerobic pathways for the biosynthesis of PUFAs share common characteristics with most of the domains found in iterative PKSs (type II) and FAS systems, including gene cluster organization, sequence homology and structural biology. These genes responsible for *de novo* bacterial PUFA biosynthesis are usually designated as the “pfa” cluster (Nishida et al. 2006; Orikasa et al. 2004). Figure I-30 shows a scheme of the genetic organization of the *pfa* genetic clusters in different organisms.



**Figure I-30.** Diversity of *pfa*-like gene clusters. Genetic organization of the *pfa* genetic clusters in different organisms. A legend is included in the figure with a color code that indicates the different domains. Figure adapted from (Shulse & Allen, 2011).

Pfa synthases described to date are huge multifunctional enzyme complexes consisting of three to four subunits usually named as PfaA, PfaB, PfaC and PfaD. These proteins contain all the classical domains necessary for the synthesis of a regular FA and some domains that are only present in PKSs: acyltransferase (AT), ketoacyl synthase (KS), ketoacyl reductase (KR), dehydratase (DH), enoyl reductase (ER), chain length factor (CLF), and acyl carrier protein (ACP) domains. There is a fifth gene called

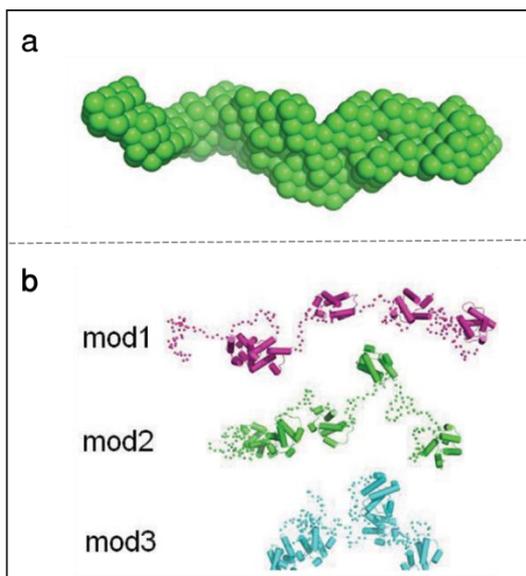
pfaE that encodes phosphopantetheine transferase (PPTase) and is in charge of activate the ACPs with the PPT arm to initiate the synthesis. Normally, each FAS or PKS condensing module contains a single ACP domain whereas PUFA synthases are characterized by the presence of multiple ACP domains (Jiang, Rajski, and Shen 2009), ranging from two to over ten in some *Thraustochytrids* (Aasen et al. 2016). An increased number of ACP domains in PUFA synthases was described to have an additive effect of PUFA synthesis yield (Hayashi et al. 2016; Jiang et al. 2008). Some of the modules mentioned above (KS, AT and DH modules) are duplicated in pfa-like clusters but the biological role of these duplications remains a mystery.

The sequential reactions of PUFA anaerobic biosynthesis are speculated based on similarities between the functional domain structures of the pfa gene clusters and the very well-known FAS and PKS systems, however, confirmation of the individual enzymatic reactions of the gene products have not yet been fully determined (Yoshida et al. 2016). The LC-PUFA biosynthesis reaction is presumably initiated by activating the ACPs of PfaA from their apo- to their active holo-form by pfaE (Chan and Vogel 2010). This enzymatic activation can be replaced with an heterologous gene from *E.coli*, indicating that the reaction is not specific (Sugihara, Orikasa, and Okuyama 2010). As in the FA synthesis, the activated ACP domains would be ready to accept acetyl and malonyl groups from acetyl- and malonyl-CoA and each reaction should be respectively catalyzed by an AT domain. These two molecules on the PfaA (ACP) would be condensed with the release of one molecule of CO<sub>2</sub> by a Claisen condensation reaction catalyzed by KS. Subsequently, the keto group would be reduced by a KR domain forming a hydroxyl moiety. The latter undergoes a dehydration reaction catalyzed by DH to form a trans-double bond that would be reduced by an ER to form a saturated C-C bond or isomerized by DH forming a cis double bond that will remain in the final FA chain (Yoshida et al. 2016).

Pfa gene clusters are involved in the synthesis of molecules with long carbon chains like EPA, DHA or ARA, but the mechanism by which PUFA synthases determine the final product length is not yet fully understood. *Shewanella pneumatophori* and *M. marina* produce EPA and DHA respectively. A few years ago it was described the role of the KS-AT domain codified by pfaB, the second gene of their pfa cluster. It seems to be involved in the determination of the final chain length of the pfa product, deciding whether the protein complex produce EPA (20 carbons) or DHA (22 carbons) (Orikasa et al. 2009a). This gene codify for a protein that have two predicted domains: an N terminal KS-like domain that lacks its canonical catalytic cysteine but could be fold as a regular KS domain, and a regular AT domain at the C terminal part (Shulse and Allen 2011). Even if pfaB substitution leads to the production of a 2-carbon-shorter or longer molecule in the case of *M. marina*, other experiments suggest that other components can influence the final carbon-chain length in close related systems, such as the introduction of new starter units in the case of iterative PKS (Shen et al. 1999). It therefore seems possible that chain-length determination is a function of the entire PKS complex and the mechanism by which this protein produces this effect is still unknown.

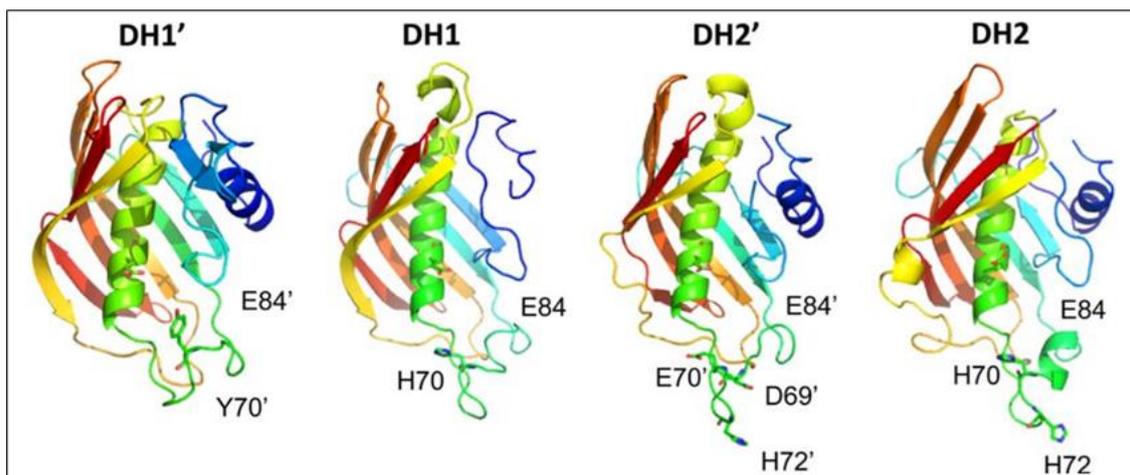
Other important domains present in pfa systems that were extensively studied are the tandem ACPs. The crystallization of proteins with flexible linkers like the ones found in tandem ACPs is often problematic due to their higher flexibility in solution, making it difficult to analyze their structure by X-ray crystallography. To understand the role of these multiple ACP domains, Abel Baerga from the University of Puerto Rico and co-workers reconstructed three-dimensional models of the five ACP domains tandemly

arranged in the PUFA synthase of *P. profundum* using small-angle X-ray scattering (SAXS) analysis to predict their structure (Trujillo et al. 2013a). The most plausible model extracted from the SAXS data suggested a monomer form and an elongated beads-on-a-string structure. In addition, the SAXS analysis revealed that the tandem ACP could adopt several conformations in solution, indicating that the structural flexibility of the tandem ACP would have additive and parallel effects on PUFA production. These results, despite having a great interest since they reveal the apparent absence of supra structures, do not explain why the presence of more modules increases the yield of the synthesis. Figure I-31 shows the solution structure of tandem ACP solved using NMR.



**Figure I-31.** Solution structure of tandem ACP by using nuclear magnetic resonance (NMR). The three-dimensional bead model shown in panel a reveals a molecular volume of 96,600 Å<sup>3</sup>. Simulation of the scattering data based on structural models reveals that an extended and flexible overall configuration can sufficiently account for the observed data as is shown in panel b. Both panels are on a different scale. Figure adapted from (Trujillo et al., 2013).

Baerga's laboratory have also recently reported a prediction of the tertiary structures of the DH domains from the PUFA synthase of *Photobacterium profundum* (Oyola-Robles et al. 2013a). A sequence analysis of the C-terminal region of the *pfaC* gene revealed two hidden pseudo-DH domains, which show relatively high sequence similarity to each other but lack a conserved active site His residue. Each of the pseudo-DH domains (DH') is located immediately upstream of two regular DH domains, that resemble FabA from *E.coli*. A three-dimensional model of these DH-DH' didomains showed that not only the DH domains but also the DH' pseudo domains were predicted to show a "hotdog" folds despite lacking the conventional active center. As was introduced before, this conserved structural motif was recognized in most of the DH proteins (Dillon and Bateman 2004). It is suggested that the DH-DH'-DH-DH' motif is probably forming two dimers, each of which assembles as a DH-DH' dimer in which residues of both subdomains are used for the catalysis. This arrangement of domains and pseudo-domains is generally conserved among Omega-3 fatty acid synthases.



**Figure I-32.** 3D models for DH domains and DH' pseudo-domains. A comparison of the 3D models obtained using the FabA homology regions (DH1 and DH2) and for the uncharacterized pseudo-domains (DH1' and DH2') anticipated that the pseudo-domains will likely consist of a hotdog fold with some differences in conserved amino acids in key positions. While the DH models contain a conserved His in the active site which has been implicated in DH activity in other DH domains, the DH' pseudo-domains do not. While in FAS/PKS proteins the pseudo-domains are located C-terminal to the DH domain, in the PUFA synthases, the pseudo-domains are located in an N-terminal position. Where in FAS/PKS proteins the DH domains are always didomains, the PUFA DH complex invariably consists of a tetradomain (two FabA-homology plus two pseudo-domains). Figure taken from (Oyola-Robles et al., 2013).



## **Aims and Scope**

---



## 2. Aims and Scope

---

Structural biology and protein engineering provide us with opportunities to improve enzymes whose functions could be interesting to develop some application or simply to improve our specific knowledge about a particular biological system. In this work two families of bacterial proteins that synthesize lipids will be studied independently: A WS/DGAT protein from *Thermomonospora curvata* that is able to produce TAG and WEs when is heterologously expressed in *Escherichia coli* and the *pfa* proteins from marine gamma proteobacterias that synthesize Omega-3 fatty acids. Here we focused on two specific objectives each of which involves the study of one synthesis process carried out by each of the protein systems:

- Improve TAG production in *E.coli* evolving WS/DGAT from *T. curvata*. To carry out this task, experiments of directed evolution using the WS/DGAT protein will be conducted in order to improve its catalytic mechanisms and stability in its heterologous host. The ability to rapidly produce TAG and WEs in a few hours when it is expressed in *E.coli* was previously demonstrated in our laboratory, but we believe that the production system has the potential to be improved. Besides that, direct evolution experiments can also contribute to improve the structural knowledge and maybe point out key residues for the accumulation process.
- Understand the mechanism by which the proteins of the *pfa* cluster from marine bacteria naturally produce omega 3 fatty acids. To do this we will focus the study on trying to solve the structure of each of the domains involved in the synthesis process by using X-ray crystallography techniques. Protein-substrate binding biochemical assays will be also performed to try to define the enzymatic contribution of each of the domains present in the complex. With the information obtained in these experiments we will try to postulate a preliminary model for the omega-3 production in marine bacteria.



## **Materials and Methods**

---



## 3. Materials and Methods

---

### 3.1. Bacterial Strains

Dh5 $\alpha$  strain has been routinely used for obtaining and maintaining constructs because it lacks the natural *E. coli* recombinase system which improves the stability of plasmids. DE3 Strains Rosetta or BL21 were used for protein expression. They have a chromosomal copy of the phage T7 RNA polymerase gene controlled by IPTG-inducible promoter that facilitates protein expression. For activity assays and directed evolution of tDGAT, BW27783 strain was used because it allows the protein expression controlled by arabinose concentration, allowing the tuning of the amount of protein produced.

### 3.2. Microbiological techniques

#### 3.2.1. General culture conditions

For standard bacterial growth LB culture medium (10 g tryptone, 5 g yeast extract, NaCl 5 g/L) (Pronadisa, Spain) was used. For growth on solid medium plates LB was supplemented with 1.5% (w / v) agar (Pronadisa). All medium were sterilized in an autoclave at 120 ° C for 20 minutes. Whenever a strain carrying a plasmid with antibiotic resistance is grown, the medium is supplemented with the following antibiotic concentrations: Ampicillin (Amp): 100  $\mu$ g/ml, Chloramphenicol (Cm): 25  $\mu$ g/ml, Kanamycin (Kn): 50  $\mu$ g/ml.

#### 3.2.2. Culture conditions for neutral lipid production for TLC analysis

For the direct evolution of tDGAT and selection of the different mutants, TAG production is needed to be controlled as finely as possible. *Escherichia coli* BW27783 strain with a pBAD expression vector was used for recombinant protein expression due to its ability to tune protein expression depending of to the amount of arabinose present in the medium at induction time, controlling in this way how fast TAG are produced in the experiment. Cells were grown in Luria–Bertani (LB) medium and chloramphenicol concentration in LB medium was 25  $\mu$ g/ml. An overnight culture of *Escherichia coli* BW27783 cells harboring the PBAD::tDGAT constructs (wild type or mutants) was diluted 20-fold in a 50 ml flask. When OD<sub>600</sub> reached 0.6, arabinose (0.0001% v/v) was added to induce protein expression and start neutral lipid production for the following 24 hours at 37 °C, with shaking at 170 RPM. Cells pellets were collected by centrifugation on an Eppendorf 5810R centrifuge (4000 rpm, 20 °C, 15 min) and were stored at -20 °C for further lipid extraction and TLC analysis.

#### 3.2.3. Culture conditions in microtiter plates

To perform high throughput screening and detect differences in neutral lipid production in tDGAT mutants, small-scale cultures were made in Thermo Scientific™ 96-Well microtiter microplates containing 200  $\mu$ l of LB supplemented with chloramphenicol (25  $\mu$ g/ml) and grown at 37°C overnight with shaking (170 RPM) in a PerkinElmer's VICTORX3 Multilabel Plate Reader. Cells were induced with arabinose (0.00001% v/v).

The screening protocol is described in detail in the “Fluorescence-based High-Throughput Screening” section in materials and methods.

### 3.2.4. Bacterial glycerol stock

Bacteria on an LB agar plate can be stored at 4°C for a few days/weeks. However to store bacteria for a longer time, it is needed to establish glycerol stocks. The addition of glycerol stabilizes the frozen bacteria, preventing damage to the cell membranes and keeping the cells alive. To preserve the strains used in this work, stationary phase 19 ml cultures were centrifuged and resuspended in 1 ml of peptone-glycerol [peptone 1.5 % (w/v), glycerol 50% (v/v)]. Strains were stored at -80°C.

### 3.2.5. Preparation of competent cells

For the production of electrocompetent cells the desired strain of *E. coli* a colony from a LB-agar fresh plate was inoculated on a flask with LB and grown overnight at 37 ° C in agitation. Cells were diluted to 1:20 in LB and incubated with shaking to an O.D.<sub>600</sub> of 0,8. At this point cells were incubated on ice for 30 minutes and recovered by centrifugation at 4000 rpm for 10 minutes. Pellet was washed with cold sterile water and centrifugated again in the same conditions. Supernatant was poured and the washing process was repeated once. Finally cells were washed with cold sterile water supplemented with 10% glycerol and frozen in 50 µl aliquots at -80 ° C.

### 3.2.6. Electroporation transformation

*E. coli* cells were transformed by the standart electroporation protocol. The DNA that is going to be electroporated must be salt-free. Therefore when this DNA was the product of an enzymatic reaction (ligations, isothermal assembly and megawhop protocols) the protocol included a previous microdialysis step. A drop with the DNA was deposited on a 0.05 µm pore size nitrocellulose filter (Millipore GS), in a petri dish filled with sterile ultra-pure water. After 30 minutes, the drop was collected.

For transformation the amount of DNA that is mixed with the competent cells is important, especially when high efficiency is needed. For regular electroporation of high concentrated DNA samples (≈100ng/µl) 1 µl of DNA is enough. For electroporations where the concentration of DNA is small (products of ligations or isothermal assembly) of 5-10 µl were mixed with 50 µl of competent cells. This mixture was poured on an electroporation cuvette 0.2cm Gene Pulser (BioRad) and cooled in ice. The electric pulse was done with the following conditions: 2.5 kV, 25 µF and 200 Ω using a MicroPulser TM electroporator (BioRad). Immediately after the electric pulse, 1 ml of sterile LB previously preheated at 37 °C was poured over the cells that were incubated at 37 ° C for 1 hour. Cells were plated on a selective medium according to the plasmid resistance.

## 3.3. Oligonucleotides

Primers used in this work were purchased from Sigma and are summarized in table M-1. Primers used for isothermal assembly are represented without the pET29c or pET3a homology tail of 40-base pairs needed. This tails are the same for pET29c vector in all constructions (forward tail: tagaaataattttgttaactttaagaaggagatatacat and reverse tail: tagcagccggatctcagtggtggtggtggtggtgctcgag) but for pfaE that was cloned into pET3a using the follow complementary tails (forward tail: caactcagcttccttccgggctttgtagcagccggatcc and reverse tail: caactcagcttccttccgggctttgtagcagccggatcc)

tagaaataatttggtaacttaagaaggagatatacat). ACP domains from *Colwellia psychrerythraea* were the only construction cloned with restriction enzymes and is indicated in the table.

**Table M-1.** Description of the oligonucleotides used for each genetic construction.

Primer Name	Constructions	Sequence	Primer Name	Constructions	Sequence
T7_primer	Multi cloning site in pET plasmids	taatacgactcactataggg	6ACPc_S	ACP from Colwellia psychrerythrae.	aaaaacatgatgatgtagcaactat
pT7_primer		gctagttattgctcagcgg	6ACPc_E		aaaaaggatcctaaacaccactgccttgcttg
pBAD_mcsF	Multi cloning site in PBAD plasmid	ctgtttctccataccggtt	KR_S1	KR 1, KR-DH 1	atgagtggtgttactgtaaatgtagtggtagccc
pBAD_mcsR		ctcatccgccaaaacac	KR_S2	KR 2, KR-DH 2	atggctaacgccttaactgtagtgaggcac
tDGAT_mutN_F	tDGAT-Nt domain random mutagenesis	gaaggagatatacatatg	KR_S3	KR 3, KR-DH 3	atggttgatgaaactgaaatcaacaacattattac
tDGAT_mutN_R		aaccggatgtgccagtg	KR_E	KR (1-3)	ctgtgcactgcatcaagtggaataatgtaaac
tDGAT_mutC_F	tDGAT-Ct domain random mutagenesis	actgggtcgtgatggtgtgttcc	DH_S	DH (pfaA)	atgagctctttatcaacaagaactagtc
tDGAT_mutC_R		gtgctcgagaccatcaatcagacc	pfaB_S	pfaB (1-2)	atgacggaattagctgttattggtatggatg
pet29c_iso_S	pet29c for isothermal assembly	atgtatatctccttcttaagtaaacaaaaattatttc	pfaB_E1	pfaB 1	gctatcctaattcaaatcacgcatgac
pet29c_iso_E		ctcgagcaccaccacc	pfaB_E2	pfaB 2	tttgtcgtgttctatgatgcctgc
pet3a_iso_S	pet3a for isothermal assembly	ggatccggctgctaacaagc	pfaC_S	pfaC, KS-KS (1-3), KS 1M	atggaaaaattgcagtagtaggtattgc
pet3a_iso_E		atgtatatctccttcttaagtaaacaaaaattattctag	pfaC_E	pfaC, DH2M (1-2)	cgctcaacaatactaaaacgatgttttaac
pfaA_S	pfaA, KSAT (1-5)	atggctaaaaagaacaccacatcg	KS-KSE1	KS-KS 1	cgctgaactcgataagataagatg
pfaA_E	pfaA, KRDH (1-3), DH	tgacatatcgttcaaatgtcactgacactgac	KS-KSE2	KS-KS 2	cgactgttaacaatcgctgtgtaattaac
KSAT_E1	KSAT 1	tgcgacaccaagcactgccattg	KS-KSE3	KS-KS 3	ggctgctgtagcttgcattttgaac
KSAT_E2	KSAT 2	gccatcagtcattgcatcagcaaacg	KS-1M_E	KS 1M	agacacagtcgtggtggcattgctc
KSAT_E3	KSAT 3	accagtaagcatggtgttcatgttgcctcg	DH_S1	DH1M 1, DH2M 1	atggtgactgtgaaacataaagcaag
KSAT_E4	KSAT 4	tgcaaccgccgcttattactgacactag	DH_S2	DH1M 2, DH2M 2	atgcgtataataaacgaaaaacgtgatttacgatc
KSAT_E5	KSAT 5	agcaacatgcgcaacaatcacaggttcag	DH_E	DH1M (1-2)	cagtgttacagggtaatctgaatgctcatc
5ACP_S	5ACP	atgatcgatgtagcaacttaacaaagtaatg	pfaD_S	pfaD	atgtcgagtttaggttttaacaataacaacgc
5ACP_E		catgcaaaagcaaggcagtggtgtt	pfaD_E		atcactcgtacgataactgccaattctgtac
6ACPc_S	ACP from Colwellia psychrerythrae.	aaaaacatgatgatgtagcaactat	pfaE_S	pfaE (pet3a)	atgtacagcggcgtaaaagataagctcacc
6ACPc_E		aaaaaggatcctaaacaccactgccttgcttg	pfaE_E		ctatttagcgtcaggtttaaaattagtctcagg

### 3.4. Plasmids

Plasmids used in this work were constructed by cloning the relevant PCR-generated fragments into pET29c (Km), pET3a (Ap) or PBAD33 (Cm) expression vectors (Novagen). Desired DNA fragments were amplified using oligonucleotides described in the previous “oligonucleotide” section using the genetic cluster pfa:ABCD (pDHA construction) of *Moritella marina* and genomic DNA of *Colwellia psychrerythraea* as the template and prepared to perform the isothermal assembly protocol described in the “DNA engineering” section. The identity of constructed plasmids was checked by DNA sequencing. Table M-2 shows information about the plasmid used for each construction.

**Table M-2.** Description of the plasmids used for each genetic construction.

Plasmid	Constructs
pET29c (Km)	pDHA full length proteins and domains described in table R-1
pET3a (Ap)	pfaE
PBAD (Cm)	tDGAT wt and mutants

### 3.5. DNA engineering

#### 3.5.1. DNA extraction and purification

Plasmid DNA was extracted using Gene JET Plasmid Miniprep Kit (Thermo Scientific). PCR products were purified with the kit GenElute PCR Clean-Up (Sigma-Aldrich, USA) or extracted from electrophoretic gels by employing the kit GenElute Gel Extraction (Sigma-Aldrich, USA) as described by the manufacturer. The DNA concentration was determined by measuring the absorbance at 260 nm in a Nano-Drop ND-1000 spectrophotometer.

#### 3.5.2. Polymerase chain reaction (PCR)

All DNA fragments were amplified by polymerase chain reaction (PCR) with oligonucleotides purchased from Sigma (Sigma-Aldrich, EE.UU.) and commercial polymerases. For routine testing and simple amplifications of small DNA regions or colony analysis, that do not require high fidelity, a homemade-purified version of original taq polymerase was used.

High fidelity enzymes Vent DNA polymerase (BioLabs, UK) or Phusion pol enzyme (ThermoScientific, USA) were employed for cloning, using PCR reactions with final volumes of 50  $\mu$ l or 70  $\mu$ l. For the amplification of complicated templates longer than 8 kb, Herculase II Fusion DNA Polymerase (Agilent) was used due to its ability to amplify genomic DNA targets up to 23kb. In every case the manufacturer's specifications of the commercial polymerase were followed.

Here a standard protocol for PCR is detailed. PCR mix containing both primers at a final concentration of 10 pmol/ $\mu$ l (10 $\mu$ M), 1 to 10 pg of template DNA, 200  $\mu$ M of the deoxynucleotide (dNTPs) mixture and 1X of the commercial

polymerase buffer that included MgSO<sub>4</sub> 20 mM, 1U of DNA polymerase and distilled water. The following general program was set in the thermocycler: 5-10 min of initial denaturation at 95°C; 25 cycles of amplification, including steps of denaturalization for 30 seconds at 94°C, annealing for 30 seconds at the corresponding annealing temperature (depending on the primers) and elongation for the appropriate time (1 minute for each Kb of target DNA to be amplified) at 72°C; and a last step of 10 min of final elongation at 72°C. After completion of the reaction, samples were maintained at 4°C for short-term conservation.

### **3.5.3. Agarose electrophoresis**

To analyze the amplified fragments of a PCR for cloning or check the quality of a plasmid DNA extraction, electrophoresis in agarose gel is routinely used. Agarose is dissolved in TBE (Tris-HCl 45 mM, boric acid 45 mM, EDTA 0.5 mM and pH 8.2) to a final concentration of 1% (w/v). For staining, 0.05 mg/mL SYBR Safe (Invitrogene, Life Technologies, USA) were added to the agarose solution before preparing the gel. Loading buffer (bromophenol blue 0.25 % (w/v), glycerol 30 % (w/v) in TBE 0.5x] was added to DNA samples in a 5:1 ratio. Hyperladder I (Bio-labs, UK) or o 100 bp ladder (Amersham), was used as molecular weight marker. A horizontal electrophoretic running system (Bio-Rad, USA) with TBE buffer 0.5 X was employed, with a constant voltage of 80-120V. Finally, a Quantity One software (BioRad, USA) served to visualize and analyse images taken by a Gel Doc 2000 (BioRad, USA) UV system.

### **3.5.4. Restriction enzymes cloning**

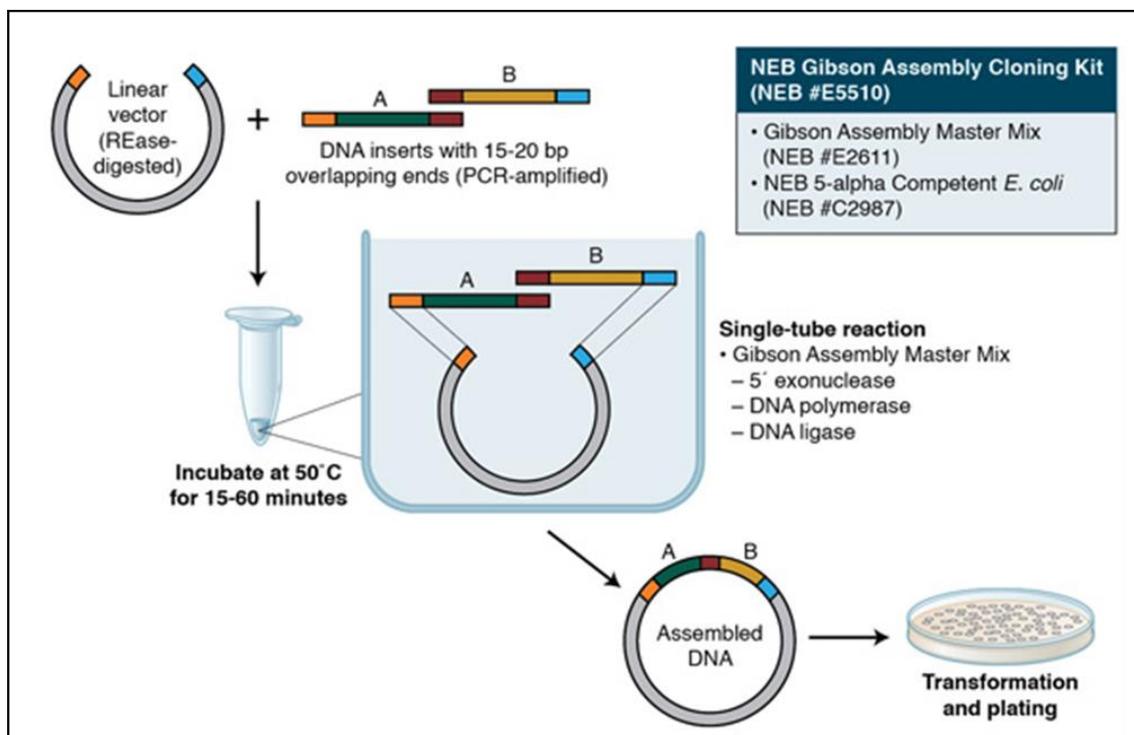
Only a few clones in the initial stage of this work have been cloned by the classical method of restriction enzymes. NdeI and XhoI restriction sites were used to introduce the DNA fragments into a pET Bacterial Expression Vector (pET29c; Novagen, EEUU). Restriction enzymes, Shrimp Alkaline Phosphatase and T4 DNA ligase, were purchased from Thermo Scientific (EEUU).

Digestion reactions were performed following the manufacturer's indications, generally in 20 µl as final volume and incubation at 37°C for 2 hours. Inactivation was carried out according to the specific recommendations of each enzyme, usually by incubation at 65 or 80°C during 10 minutes. When several enzymes were required for the same DNA, either the double digestion instructions from the manufacturer were followed or both digestions were carried out with a DNA purification step in between.

Insert fragments for digestion were obtained by PCR with primers that carry restriction sites for ligation with the plasmid vector. A 5:1 (insert/vector) molar ratio was normally used for ligation reaction that was generally performed with 0.5-20 ng of DNA and 2.5 U of T4 DNA ligase (Fermentas) in a final volume of 20 µl. The ligation reaction was incubated overnight at 22°C. For each ligation, the same reaction without insert DNA was used as negative control to quantify the number of false positive colonies due to self-ligation events. Inactivation was carried out by incubation at 65°C during 10 minutes and the ligation sample was subjected to microdialysis. A drop of the ligation reaction was deposited on a 0.05 µm pore size nitrocellulose filter (Millipore GS), in a petri dish filled with sterile ultra-pure water. After 30 minutes, the drop was collected. As a final step, 10 µl of the ligation product was electroporated into competent cells

### 3.5.5. Isothermal assembly cloning

Besides the classic cloning methodology based on restriction enzyme cleavage and ligation in this thesis a new method to rapidly make constructs has been implemented. This method is usually known as isothermal assembly or Gibson assembly (Gibson et al. 2009) and is a molecular cloning method which allows joining of multiple DNA fragments in a single, isothermal reaction. The method can simultaneously combine more than ten DNA fragments based on sequence identity. It requires that the DNA fragments contain 20-40 base pair overlap with adjacent DNA fragments. The method is based on the following sequence of reactions (Figure M-1): T5 Exonuclease removes nucleotides from the 5' end of two double stranded DNA fragments carrying overlapping terminal sequences. Thus, these two molecules will have complementary sticky ends. DNA Phusion polymerase then close the gaps and Taq ligase seal the nicks present in the DNA. Three reactions are performed at 50 ° C (isothermal). The 3 enzymes present in the mixture function correctly at this temperature, but at a certain point at the end of the reaction the exonuclease is denatured so preventing DNA degradation.



**Figure M-1.** Gibson Isothermal Assembly cloning scheme. Taken from New England Biolabs website.

The isothermal assembly protocol was done as follows: 6 ml of reaction buffer were prepared by mixing 3 ml of 1M Tris-HCl pH 7.5, 150 µl of 2M MgCl<sub>2</sub>, 60 µl of 100 mM dGTP, 60 µl of 100 mM dCTP, 60 µl of 100 mM dTTP, 60 µl of 100 mM dATP, 300 µl of 1M DTT, 1.5 g of PEG 8000, 300 µl of 100 mM NAD and ultra-pure water to full volume. Buffer was aliquoted into 18 aliquots of 320 µl, and frozen at -20°C for later use. Using one of these aliquots the reaction mixture was prepared by adding 1.2 µl of T5 exonuclease (Epicentre), 20 µl of Phusion polymerase (Thermo), 160 µl of Taq ligase (BioLabs) and 700 µl ultra-pure water. Aliquots of 15 µl were then prepared in PCR tubes and stored at -20°C (It is recommended to use them before two months). The next step

was mixing in a volume of 5  $\mu$ l a 1:1 molar ratio of the DNA mixture with fragments designed for the construction (normally the insert and the PCR-amplified and linearized vector). These 5  $\mu$ l were added to the reaction aliquots previously frozen and the mixture was incubated for one hour at 50°C. Finally, the resulting construction was introduced into Dh5 $\alpha$  by electroporation with a previous microdialysis step to get rid of salts.

Note that in the “oligonucleotide” section it is indicated that the first 40 nucleotides of each primer designed for isothermal assembly are homology tails to its correspondent vector used in the cloning procedure. These DNA tails are detailed in table M-1.

### **3.5.6. DNA sequencing**

The DNA sequence of all cloned PCR fragments was determined by Sanger automated sequencing in the company Stabvida (Caparica, Portugal). Samples were prepared following the recommendations of the company and consisted of 1000 ng of template DNA, 25 pmol oligonucleotide and MiliQ water to a final volume of 13  $\mu$ l.

### **3.5.7. DNA quantification**

Absorbance at 260 nm was measured with Nano-drop ND-1000 spectrophotometer to estimate the amount of DNA. It was also helpful to measure the absorbance at 280 nm to estimate the amount of protein present in the DNA samples and then have a measure of the quality of the DNA purification. In some cases, electrophoresis in agarose gels were helpful also to have an approximation of the actual DNA concentration

## **3.6. Protein purification techniques**

### **3.6.1. Recombinant protein production in *E. coli***

All the protocols intended for the production of recombinant proteins in this work were based on IPTG-inducible systems. A starter 50 ml culture of a DE3 *E.coli* strain for protein expression (BL21, C41 or Rosetta), carrying the genetic construction to overexpressed was grown overnight at 37°C with shaking (160 RPM) in LB medium supplemented with the appropriate antibiotics. A 1/20 dilution was prepared in 1L flasks containing fresh LB with selective medium. In order to carry out the induction of the transcription, isopropyl b-D-thiogalactoside (IPTG) to a final concentration of 0.5 mM was added to the culture when it reached OD<sub>600</sub>=0.6. When the constructions were difficult to overexpress at 37°C, temperature was set at lower temperatures (15-25°C), chilling the flasks on ice before induction. In the results section the actual temperature for induction and overexpression of each construct is specified. When protein overexpression is performed at 37°C, 4 hours is normally enough for a good yield, but in special cases, when the temperature has to be lowered, is better to let the bacteria produce protein overnight (12-16 hours). For testing protein expression, aliquots of 1 mL of the cell cultures were harvested by centrifugation at 13,200 rpm for 10 minutes in an Eppendorf centrifuge 5427R and the pellets were frozen at -20°C for posterior electrophoresis analysis. For protein purification, cell cultures were harvested by centrifugation at 5,500 rpm at 12°C for 20 minutes in a JA10 rotor (Beckman Coulter, USA) for Beckman coulter avanti j-26 centrifuge. Supernatants were discarded and pellets were frozen at -80°C.

### 3.6.2. Protein purification

In this work 3 basic methods for protein purification have been used: affinity chromatography, ion-exchange chromatography and size exclusion chromatography. As a general rule, all constructs for protein expression were made using the pET-29 vector that carry a polyhistidine-tag in the C-terminal domain, with a few exceptions that are indicated in the results section. This made straightforward the use of an affinity chromatography Nickel column as the first purification step. In most of the cases this purification step was enough to obtain a purity of around 90% in some elution fractions. For certain cases where higher purity was required (as some crystallization experiments), the protein was subjected to a second purification step, generally using ion exchange columns. For cases in which the ion exchange was not effective the method of choice was size exclusion chromatography.

#### 3.6.2.1. Nickel column affinity chromatography

The pellet was thawed at room temperature to facilitate cell lysis. Once thawed, pellet was resuspended in 50 ml of buffer A (50 mM Tris pH = 7.5, 300 mM NaCl) to which 1mM protease inhibitor phenylmethylsulfonyl fluoride (PMSF) have been added. Lysis was carried out by sonication, by 3 cycles of 1 minute each, normally set at 80% power, leaving a certain amount of time between each cycle (5 minutes) to avoid overheating of the sample. The cell lysate was centrifuged for 40 minutes at 40000 rpm. The supernatant that contains the soluble protein fraction was loaded onto a 5ml Ni-NTA column (GE Healthcare) previously equilibrated with buffer A. Proteins bound to the column were eluted with a gradient of buffer B (50mM Tris pH = 7.5, 1M NaCl, 0.5M Imidazole) using an AKTA system (GE). After elution, fractions containing the protein were concentrated by centrifugation at 4000 rpm at 4°C using Amicon Ultra Centrifugal device (Millipore) until the needed concentration was reached.

#### 3.6.2.2. Ion-exchange chromatography

Ion-exchange column chromatography was especially useful to purify omega-3 synthase domains. Due to it characteristic low isoelectric point they were negatively charged at pH 7.5. For that reason it was easy to improve previous affinity-based purifications using ion exchange columns. The solution containing the protein was loaded into a 5 mL HiTrap Q HP column (GE Healthcare) equilibrated with buffer A (50 mM Tris pH = 7.5, 0,2 M NaCl). Proteins bound to the column were eluted with a gradient of buffer B (50mM Tris pH = 7.5, 1M NaCl) using an AKTA system (GE).

#### 3.6.2.3. Size Exclusion Chromatography

Size Exclusion Chromatography is useful to separate a wide range of molecules according to size and shape. In the case of proteins, it can help in the last steps of purification and gives a good approximation of the size and the oligomeric state of the proteins that are being purified.

Gel filtration or size exclusion chromatography was performed on a GE Healthcare Life Sciences Superdex 200 10/300 column on an AKTA system. The column was equilibrated with 50ml of buffer A (50 mM Tris pH = 7.5, 0.3 M NaCl). Proteins were injected onto the column in 0.5ml aliquots and eluted at a flow rate of 0.5 ml/minute with buffer A. The elution profile was monitored by absorbance at 280nm. Samples of each elution fraction were taken and analyzed by SDS-PAGE.

### 3.6.3. Buffer exchange

Proteins to be crystallized need to be in a buffer with low concentration of tris buffer (20 mM) and with lower amount of salt (100mM of NaCl). Purified protein was concentrated to between 5 to 10 mg/ml in a centricon tube (Millipore). Fractions of protein to be subjected to buffer exchange were diluted in the new buffer to reach 1/10 of the initial protein concentration. The new volume of protein was concentrated again in the same centricon tube 2-3 times to remove the initial buffer and reach the desired concentration for further experiments.

### 3.6.4. Protein electrophoresis

Protein electrophoresis was carried out in 12% polyacrylamide SDS gels, employing as loading buffer a solution containing 50 mM Tris-HCl pH 6,8, 4 % SDS, 4 % glycerol and 0,02 % bromophenol blue. Low Range Protein Marker (BioRad, EEUU) or PageRuler Plus Prestained Protein Ladder (ThermoFisher) were loaded onto the gels as molecular weight markers, Coomassie Brilliant Blue R-250 (Merck, Germany) in methanol/acetic acid/deionized water (4.5/4.5/1, v/v/v) was used for staining and a solution of methanol/acetic acid/deionized water (1:1:8, v/v/v) for destaining. Electrophoresis was run for 55 min at 185 V and 400 mA.

### 3.6.5. Protein quantification

Absorbance at 280 nm has been the method of choice for protein quantitation because of their simplicity and cost-effectiveness relation. Absorbance at 280 nm was measured with a Nano-Drop ND-1000 spectrophotometer. The relation between the theoretical extinction coefficient and de molecular mass of the protein was used to estimate the actual concentration.

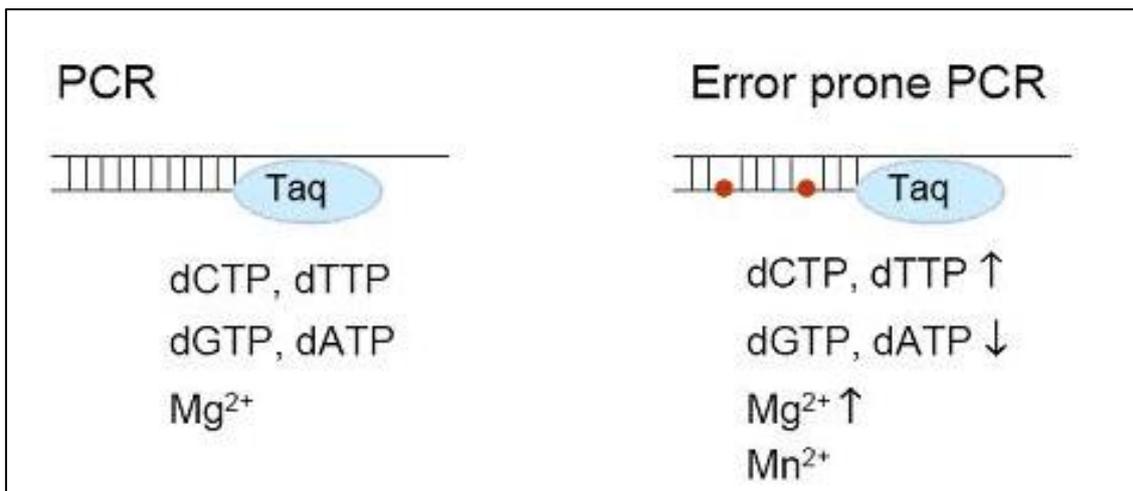
### 3.6.6. Mass spectrometry analysis

Molecular masses of protease-digested proteins were determined by MALDI-TOF mass spectrometry. Both preparations were desalted using ZipTip C4 microcolumns (Millipore) (2- $\mu$ l sample of Ma2 at 0.8 mg/ml) with elution using a 0.5- $\mu$ l SA (sinapinic acid [10 mg/ml] in [70:30] acetonitrile-trifluoroacetic acid [0.1%]) matrix on a GroundSteel massive 384 target (Bruker Daltonics). An Autoflex III MALDI-TOF/TOF spectrometer (Bruker Daltonics) was used in linear mode with the following settings: 5,000- to 200,000-Da window; linear positive mode; ion source 1, 20 kV; ion source 2, 18.5 kV; lens, 9 kV; pulsed ion extraction of 120 ns; and high gating ion suppression up to 1,000 Mr. Mass calibration was performed externally with Bruker's Protein 1 standard calibration mixture (Bruker Daltonics) in the same range as the samples. Data acquisition was performed using the FlexControl 3.0 software program (Bruker Daltonics), and peak peaking and subsequent spectral analysis were performed using FlexAnalysis 3.0 software (Bruker Daltonics).

## 3.7. Directed evolution and protein engineering

### 3.7.1. Error prone PCR

Error prone PCR was used to generate genetic diversity as a first step for the direct evolution of tDGAT using the tDGAT\_mut primers described in table M-1. This is a method by which mutations are randomly inserted into any piece of DNA. The difference with respect to regular PCR is that the fidelity of the Taq DNA polymerase is modulated by alteration of the composition of the reaction buffer. In these conditions, the polymerase introduces some mistakes in the base pairing during DNA synthesis that results in the introduction of errors in the newly synthesized complementary DNA strand (figure M-2).



**Figure M-2.** Schematic representation of the error prone PCR technique.

pBAD plasmid DNA carrying wild type DNA sequence for tDGAT was subjected to this Error prone PCR in order to generate new variants of the protein. Here two mutant libraries corresponding to each of the two natural domains of the protein were generated separately. In this way mutation rate could be adjusted more accurately and significantly reduce the mutational spectrum, making the further selection easier. The protocol was performed using the GeneMorph II Random Mutagenesis kit (Stratagene) according to the manufacturer's protocol and using this PCR cycle (95°C for 2 min, 1 cycle; 95°C, 30 sec/55°C, 25 sec/72°C, 1,5 min, 20 cycles; 72°C, 10 min; with 1000 ng of template DNA). Primers used for mutagenesis are listed in the "oligonucleotides" section in materials and methods and were used for mutagenesis of the N terminal domain from the aa Arg2 to Arg189 and the C terminal domain from Arg189 to Phe261. PCR products were separated by gel electrophoresis and purified using a gel-extraction kit (Thermo Scientific)

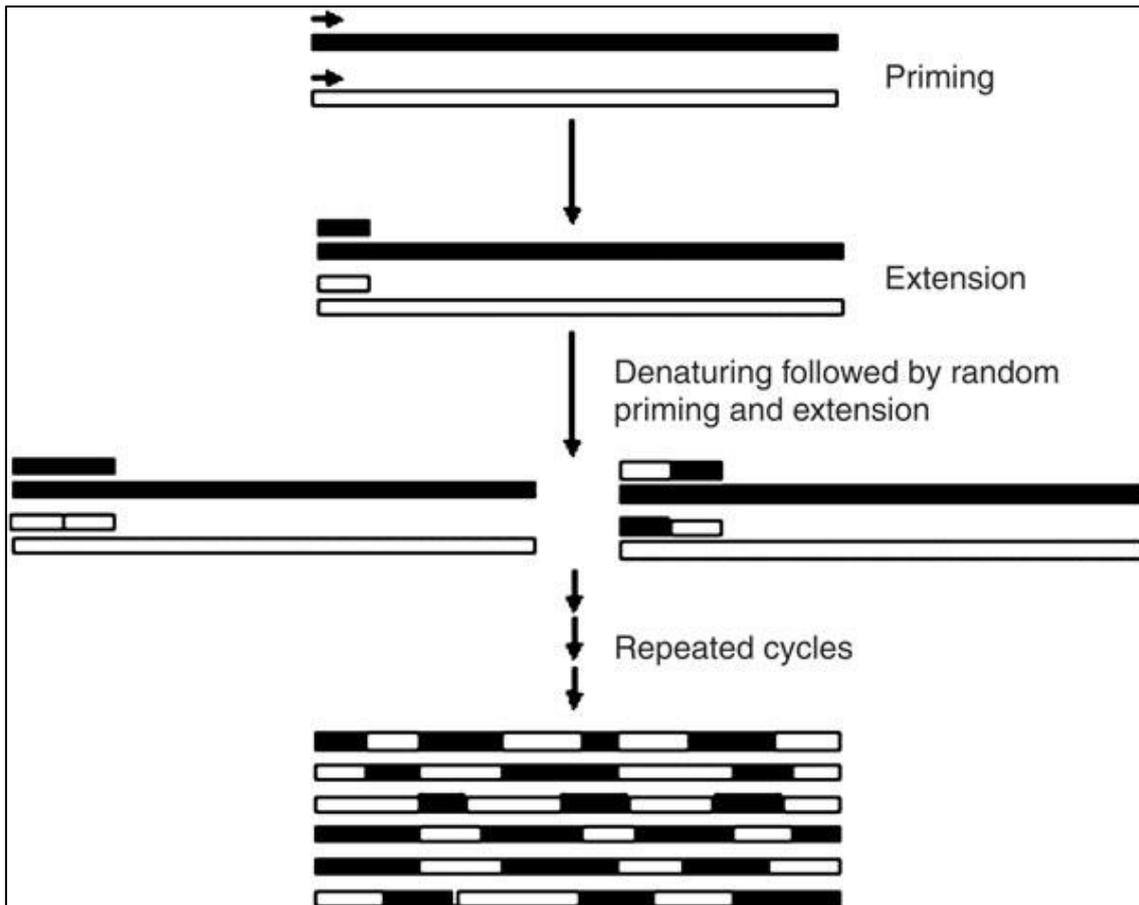
### 3.7.2. Site directed mutagenesis

In this thesis it was necessary to introduce point mutations in some omega-3 fatty acid synthases domains like KSAT of pfaAm. For this purpose an adaptation based on the quickchange method of Stratagene mixed with the megawhop protocol of Miyazaki K (K and M 2002) was developed. This technique is useful for introducing single base pair, single amino acid (codon), or multiple changes into a gene residing on a plasmid.

In short, one mutagenic primer is paired with a standard primer to create a megaprimer product that contains the mutation(s) of choice. This PCR product has to be longer than 200 bp. The desired point mutation is introduced in the mutagenic primer with at least 15 bp of homology downstream of the mutation. A regular PCR was run using these two primers so that an amplification of around 200-400 bp is generated, using the wild type gene as the PCR template. 2x50 µl reactions normally yielded enough product to be able to run an agarose gel and purify at least 1000 ng of DNA from the band of the PCR product. After this, the DNA purified was used as a megaprimer for a second PCR using the plasmid construct that is going to be mutated as the template. Phusion polymerase was used for both PCR reactions in the protocol. PCR conditions for the second megaprimer-like PCR (that uses the megaprimers previously generated): (98°C for 4 min, 1 cycle; 98°C, 30 sec/65°C, 30 sec/72°C, 5 min, 30 cycles; 72°C, 5 min). After the PCR, 1 µl of DpnI was added to digest the template plasmid. Finally, the resulting PCR product was introduced into Dh5α by electroporation with a previous microdialysis step.

### **3.7.3. Staggered extension process (StEP) *in vitro* recombination**

*In vitro* polymerase chain reaction (PCR)-based recombination methods are used to shuffle segments from various homologous DNA sequences to produce highly mosaic chimeric sequences. StEP recombination is based on cross hybridization of growing gene fragments during polymerase-catalyzed primer extension. This process is easily controlled introducing very short extension times (1-5 s) in the PCR reaction using the genetic variants that are wanted to recombine as the templates. As a final step for the direct evolution of tDGAT, the 4 variants that were selected during the process were recombined to check if it was possible to generate even better variants.



**Figure M-3.** Schematic representation of the in vitro recombination PCR technique. Taken from (Zhao and Zha 2006)

In vitro recombination was performed as described in (Aguinaldo and Arnold 2003) with minor modifications. Biotaq DNA pol (Bioline) was used with the following PCR adapted protocol (94°C for 5 min, 1 cycle; 94°C, 30 sec/ 55°C, 5 sec/, 20 cycles; 72°C, 7 min) using plasmid DNA of 4 selected tDGAT mutants with improved TAG production as the template and the following primers: Forward: 5'-TAATACGACTCACTATAGGG-3' and Reverse: 5'-GCTAGTTATTGCTCAGCGG -3'. Figure M-3 shows an schematic representation of the in vitro recombination technique.

### 3.7.4. Megawhop PCR

Megawhop is a PCR-based method that allows the cloning of DNA fragments into a vector. The DNA fragment to be cloned is used as a set of complementary primers that replace a homologous region in a template vector through whole-plasmid PCR. In this work, megawhop-PCR was used to generate the mutant libraries of the tDGAT gene using the DNA pool extracted from random mutagenesis as primers and the tDGAT wild type gene as template.

DNA-mutated fragments coding for the two predicted domains of tDGAT were used as megaprimers according to the megawhop-PCR protocol established by Miyazaki K (K and M 2002). PCR conditions were tested with different amounts of megaprimers. Best conditions for improving electroporation efficiency of megawhop product were found

to be as follows (98°C for 4 min, 1 cycle; 98°C, 30 sec/65°C, 30 sec/72°C, 5 min, 30 cycles; 72°C, 5 min; with 50 ng of pBAD:tDGAT as template, and 150 ng of megaprimers in 70 µl PCR reaction tubes). Products were digested with DpnI restriction enzyme for 2 hours. After that, a second 2-hour digestion has been done in order to completely remove biological DNA. Random mutagenized DNA constructs were transformed into competent *E.coli* BW27783.

## 3.8. Biochemical assays

### 3.8.1. Lipid extraction and thin layer chromatography

Thin layer chromatography was used to experimentally quantify the changes in neutral lipid production of the tDGAT variants analysed in this work. Lipid extraction were performed as follows: pellets from 50 ml cultures were thawed and resuspended in 1 ml of hexane:isopropanol (3:2 vol/vol), with shaking for 5 hours. 500 µl of each supernatant was collected by centrifugation on an Eppendorf 5415R centrifuge. Extracts were dry at 45°C using Eppendorf concentrator plus and analyzed by TLC.

TLC analyses were carried out using 20x20 cm DC-Fertigfolien POLYGRAM®SIL G pre-coated TLC-sheets (0,2mm) and developed using hexane:diethyl ether:acetic acid (8:1:1 vol/vol/vol) as mobile phase. Dry samples were resuspended in 20 µl of hexane:isopropanol (3:2 vol/vol) and applied on the TLC plate. TLC plates were stained with iodine vapor for 40 minutes. Tripalmitoyl-glycerol (Sigma) or olive oil were used as standards.

### 3.8.2. Fluorescence-based lipid detection

For the mutant library screening around 10.000 individual mutants were analysed for each domain of the protein. In both cases, colonies were transferred to Thermo Scientific™ 96-Well microtiter microplates containing 200 µl of LB supplemented with chloramphenicol (25 µg/ml) and grown at 37°C overnight with shaking (160 RPM). To avoid the hand-picking of the individual colonies, a dilution was carried out after electroporation in order to reach an initial bacterial concentration of 1.5 per well. This was doing by plating dilutions of the transformed cells to exactly know how many cells we had per ml. 8 hours after inoculation, arabinose (0.00001% v/v) induction was performed directly on the 96-well plates and cells were checked for the accumulation of neutral lipids for the next 24 hours under the same growing conditions. Red Nile (stock solution prepared at 0.5 mg/ml in DMSO, Sigma, EEUU) was added to the plates to a final concentration of 0.5 µg/ml. 10 minutes after staining fluorescence and OD<sub>600</sub> were measured (590<sub>nm</sub>-620<sub>nm</sub>) using PerkinElmer's VICTORX3 Multilabel Plate Reader. Fluorescent data was normalized by the OD<sub>600</sub>. Mutants were selected using the average of the normalized fluorescence in each plate as a reference (figure R-1.B). Best neutral lipid producers and deleterious mutants must differ 2 times the average of the plate to be selected for further analysis.

As bacterial concentration was calculated for being around 1.5 cells per well, selected mutants had to be plated to pick single colonies that were subjected to a second fluorescence-based screening. DNAs of selected mutants were sequenced using the oligonucleotides T7 and pT7 after TLC analysis.

### 3.8.3. Radioactivity assays: Substrate-protein complexes

To investigate the enzymatic activity of the different pfa domain *in vitro*, selected protein combinations were incubated with radiolabeled substrates and analyzed via radio-SDS-PAGE. For example, a reaction mixture (20  $\mu$ L) contained 20  $\mu$ M [ $^{14}$ C]-malonyl-CoA (55 Ci/mol) and 10  $\mu$ M of each of the protein/s tested, in a buffer containing 50 mM Tris, pH 7.5, 0.5 mM EDTA. Following incubation at room temperature for 30 min, the reaction was quenched with 10  $\mu$ L of protein loading buffer (0.1% bromophenol blue, 1% SDS, 40% glycerol), and loaded directly onto a 12% SDS gel. Following staining, dried gels were subjected to autoradiography.

## 3.9. Biophysical methods

### 3.9.1. Ultracentrifugation analysis of tDGAT-TAG lipoprotein complexes

An overnight 50 ml culture of *E. coli* C41 (DE3) cells that expressed the tDGAT protein for 3 hours were pelleted, resuspended in buffer (Tris 50mM pH 7.5; NaCl 150mM) and sonicated. The broken cells were subjected to a series centrifugation cycles to pellet and separate different subcellular parts. In this way the cell fractions where tDGAT protein and neutral lipids remain associated were identified. For this purpose the sample was subjected to consecutive centrifugations at 2,000 g., 10,000 g. and 55,000 g., harvesting the supernatant in each centrifugation step for using it in the next one. In each step a sample of the supernatant and a sample of the pellet were analyzed by SDS-PAGE and lipid extraction followed by TLC.

Lipid bodies were isolated essentially as described by Ding et al (Ding et al. 2012). An overnight 1 L culture of *E. coli* C41 (DE3) cells that expressed the tDGAT protein was harvested by centrifugation at 3,000 g for 10 min, and resuspended in 80 ml buffer A (25 mM tricine, 250 mM sucrose, pH 7.8). After 20 min incubation on ice, cells were lysed by a constant cell disruptor (constant systems) treatment twice at 40,000 psi. The sample was centrifuged at 3,000 g for 10 min to collect cell debris and inclusion bodies. Supernatant (around 80 ml) was loaded into ultracentrifuge tubes with 2 ml buffer B (20 mM HEPES, 100 mM KCl, 2 mM MgCl<sub>2</sub>, pH 7.4) on top, and was then centrifuged at 180,000 g for 1 h at 4°C. The fraction on top of the sucrose gradient was collected and the lipids were extracted and analysed by TLC.

## 3.10. X-ray Crystalization

The crystallization of a protein requires high purity protein samples of at least 90% of purity, being the major bottleneck for protein crystal structure determination the obtention of diffraction-quality crystals. To obtaining proteins or protein fragments for crystallization our main strategy was to generate a pool of constructs big enough that gave us a good chance to be able to produce soluble and stable proteins that were good candidates for crystallization trials.

In order to find initial crystallization conditions, the protein of interest was screened using various commercially available screening kits (Hampton research crystal screen 1 and 2. These kits contain reagents that cover a large range of precipitants, pH and organic compounds. The solutions were pipetted into the reservoirs of 96 wells plate (sitting drop) in 50  $\mu$ l aliquots. A 0.7  $\mu$ l drop of the protein wanted to crystallize was then placed in the well of each chamber and 07.  $\mu$ l of precipitant from the reservoir was added

to these drops without further mixing. Each tray was sealed with Crystal Clear Tape (Manco, Inc) and placed in an incubator at 22°C.

If these screens produced crystals in any condition after a few days, a second screening was performed in order to improve crystal quality. Initial promising conditions were emulated in a 24 wells plate, varying the concentrations of precipitants and also the pH of the buffer solution.

### 3.10.1. Crystallographic data collection

At the beginning of the X-ray diffraction experiment the measured crystal has to be mounted on a goniometer head and slowly rotated during X-ray beam. For this purpose, the crystal is fished out from the crystallization drop by a nylon or plastic loop. Mounted crystals are then flash-frozen with liquid nitrogen. To prevent cracking of crystals during freezing process, crystals are generally plunged into a cryoprotectant solution that in our case was always ethylene glycol 10% (final concentration). After that, the mounted crystal are diffracted by an X-ray beam. Reflections are recorded at different positions of the crystal, providing the information of the molecular arrangement of the crystal. Diffraction experiments were performed at ALBA synchrotron (Barcelona).

To collect all the necessary data, the crystal has to be rotated gradually through 0-360°, depending of the symmetry detected. In the case of crystals with higher symmetry, a smaller angular range such as 90° or 45° can be used but it is worth to collect as much data as the crystal can withstand. Usually multiple data sets are necessary to solve certain phasing problems in the case where no similar structures were solved in the past, but here we were able solve the structure of PfaC of *Moritella marina* by molecular replacement so it was enough with one complete data set.

### 3.10.2. Crystallographic data processing

In the following steps the X-ray data are interpreted computationally. The two-dimensional images of the reflections were collected at different rotation angles and will led to construct and refinement of a 3D model of the electron density and later on to arrange atoms within a crystal. This extensive approach could be separated into the following parts: indexing, merging, scaling, and phasing. Data processing was accomplished using iMOSFLM and merged and reduced with SCALA from the CCP4 package.

Next step involves solving the phase problem. The x-ray detector can only record intensities but not phases of the electromagnetic waves. Each reflection on the diffraction pattern led to the formulation of a structure factor that correspond to a wave consisting of an amplitude and phase. The amplitude is easily calculated by taking the square root of the intensity, but the phase information is lost during the data collection. In this work, we were able to solve the phase problem only using the molecular replacement method that rely on having homologous structures previously published to solve the phase problem.

### 3.10.3. Molecular replacement

The collected diffraction data from protein crystal is a reciprocal space representation of the crystal lattice. The crystal symmetry and the unit cell size and shape dictate the arrangement of the diffraction reflections during data collection experiment. The structure factor is complex number containing information about the amplitude and phase of the X-ray wave. Both amplitude and phase must be known to construct the

electron density map that allows a crystallographer to build a starting model of the molecule. Molecular replacement is the most commonly used method of a phase problem solution. In this approach a related 3D structure has to be known with more than 30% of sequential identity to determine the orientation and position of the molecules within the unit cell. The phase of the structure factor of the search model can be used for the refinement.

KSCLF structure was solved by molecular replacement using the structural core of curacin polyketide synthase as model (Protein Data Bank accession: 4MZ0) and phasing was performed with PHASER (McCoy et al. 2007) software of the ccp4 suite (Winn et al. 2011:4).

#### **3.10.4. Model building and refinement**

After solution of the phase problem an initial model can be built. This initial model can be used to refine the phases. Every cycle of this process leads to the improvement of the model; therefore, a new model would be applied for further structure development. The refinement cycle is fitting atomic positions of the model and their corresponding B-factors (parameter that reflects the thermal motion of the atom) to the observed electron density map, usually yielding a better set of phases. The number of refinement cycles depends on the difference between the diffraction data and the model that has to be low enough. Stereochemistry, hydrogen bonds and allocation of bond lengths and angles are corresponding quantities, which describe the quality of the refined model. But the main parameters that have to be controlled after every round of model improvement are R- and R-free factors. Both factors depend on the resolution of the data. R-factor is a criterion of agreement between X-ray diffraction data and constructed crystallographic model, showing how the refined model predicts the measured data. For macromolecules R-factor usually varies from 0.6 to 0.2. Model refinement was accomplished using PHENIX (Adams et al. 2002) and Coot software (Emsley and Cowtan 2004).

### **3.11. Bioinformatic tools**

#### **3.11.1. Densitometry analysis of TLC**

The TLC spots densities were quantified using a scanning densitometric analysis and Fiji (Image J) software program (NIH, Bethesda, MD) (Schindelin et al. 2012). Normalization using spots unaffected by tDGAT expression is required in order to avoid positional differences in staining. Densitometric data for TAGs and WEs production are expressed as a percentage of the wild type control spot density.

#### **3.11.2. Structural modeling**

Structure predictions used in this work were predicted by homology modeling using the Phyre2 server (Kelley et al. 2015). Images of the resulting 3D models were generated using the Pymol software (Delano 2002).

### **3.11.3. Bioinformatic softwares**

Vector NTI and Bioedit software suites were used to work with genetic and protein sequences for oligonucleotide design and visualization. Blast online server was used for finding protein and DNA homology sequences. The translate tool (ExpASy) was used for translating DNA code into amino acid code. The ProtParam tool (Expasy) was used to estimate protein parameters such as the isoelectric point, the molecular mass or the extinction coefficient. Swiss model and Phyre2 online servers were the tools of choice to find structural homologs and generate structural models of our proteins. Pymol software was used for visualization. Jpred was used to predict protein secondary structures. InterPro online tool was used for prediction of protein motifs and foldings. T-Coffee was used for multiple sequence alignment.

## **Results and Discussion**

---



## 4. Results and Discussion

---

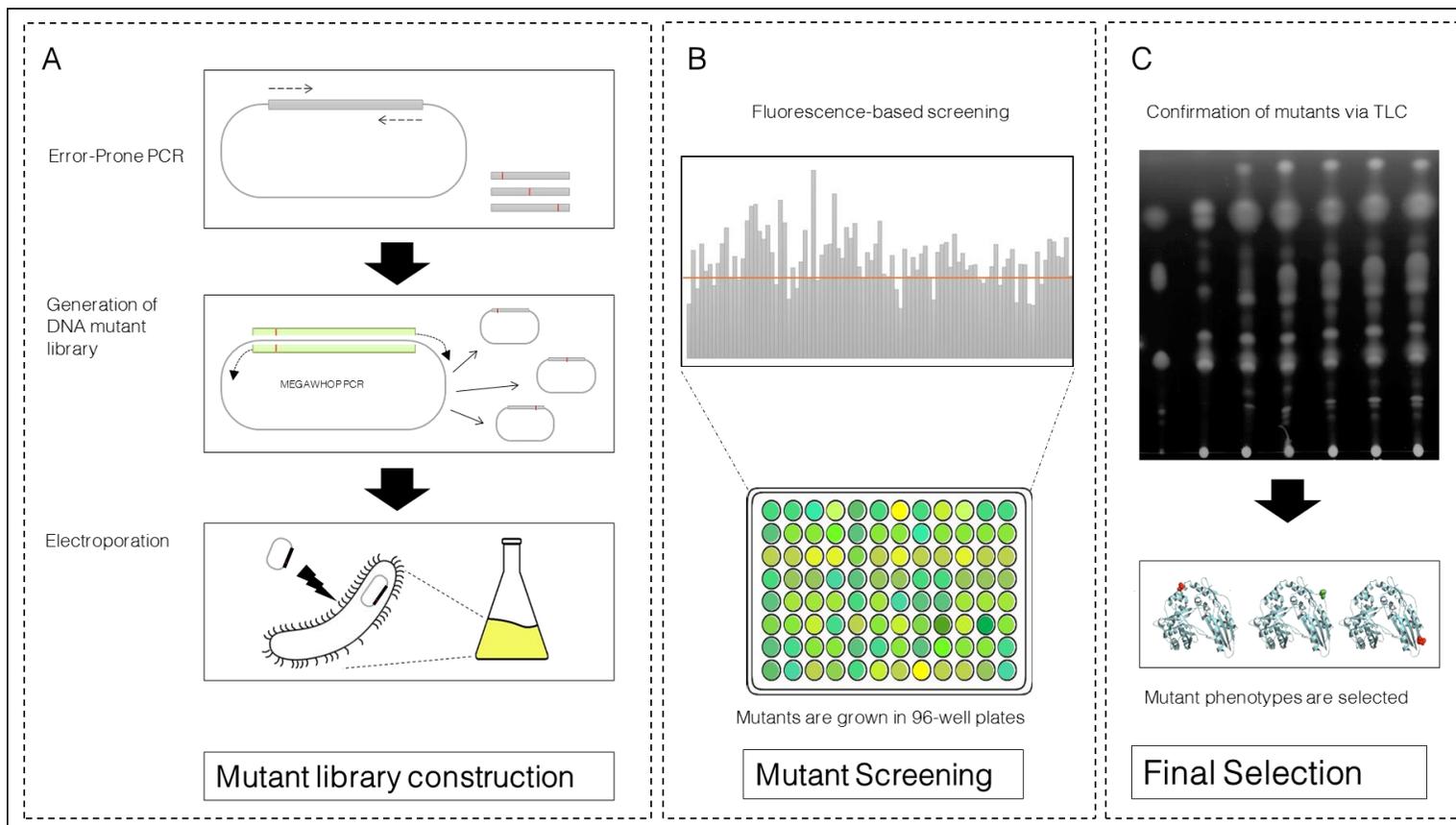
### 4.1. Optimization of TAG production in *E. coli*

As is explained in the “introduction” section, our group have recently reported that *E. coli* is able to rapidly accumulate TAGs and WEs when *tdgat* gene from *Thermomonospora curvata* is expressed (Lázaro et al., 2017, *in press*). As we are interested in the biological production of high-added value lipids, we decided to evolve *in vitro* the tDGAT protein to tune the protein and improve the productivity of the system. The main objective of this part of the work was to improve the neutral lipid accumulation observed in *E. coli* subjecting the enzyme to a directed evolution process. Through this process, we could also be able to learn some structural aspects of our protein.

#### 4.1.1. Directed evolution

##### 4.1.1.1. Evolving tDGAT using a fluorescence-based method

An experimental screening method based on fluorimetry is here optimized for the directed evolution of the WS/DGAT protein tDGAT from *T. curvata*. Genetic diversity is generated through a random mutagenesis process based on PCR, as described in detail in the “materials and methods” section. Constructs carrying mutations are transformed into competent *E. coli* BW27783 cells and subsequently, mutants showing changes in final TAGs/WEs production are selected in 96-well plates using the lipophilic stain Nile red. We were interested in mutants showing an increased TAGs accumulation, but using this technique, we could also be able to select mutants with a decreased activity that will help us in the understanding of the tDGAT esterification process. Therefore, proper absorption-emission wavelengths for detection of intracellular TAGs were experimentally established as 590nm-620nm. However, these parameters could vary significantly depending on the solvent used and lipid concentration. It was also experimentally established that the use of bacterial strains with a vast membrane surface as C41 or C43 (Bl21) can lead to problems due to a significant background that can mask the signal. Thus, we decided to use a different strains considering this characteristic. BW27783, our strain of choice, showed a very good correlation between the quantity of lipids detected in TLC and the fluorescence signal detected.



**Figure R-1.** Schematic representation of the experimental workflow for tDGAT direct evolution. This methodology was implemented in our group for the modification of WS/DGAT enzymes in order to change its neutral lipid production. Box A. Mutant library construction using the megawhop protocol after an error-prone PCR step where the genetic diversity is generated. The DNA library is subsequently transformed into electrocompetent *E.coli* cells. Box B. Promising mutants are pre-selected by a fluorescence-based screening process developed to detect the binding of red Nile dye using 96-well plates. Box C. Mutants showing significant changes in TAGs or WE accumulation are confirmed by TLC and selected.

Usually one of the major bottlenecks in protocols for mutant selection is the hand-picking of the colonies to transfer them to a suitable platform for the screening. To avoid this, in this work we made a dilution of the bacterial sample after electroporation of the megawhop product in order to reach an initial bacterial concentration of approximately 1.5-2 cells per well in the 96-well plates. For doing that, it was necessary to first plate serial dilutions of the electroporation product to estimate the concentration of cells of the vial and dilute them within the 96-well plates reaching the desired concentration. Using this system, the adjusted dilution rate allows detection of mutants that show different phenotypes in coexistence with mutant expressing wild type proteins. After plating the selected dilutions at the end of the first selection round, a second round of manual screening is needed in order to isolate single mutants due to the possibility of having more than one genotype in each well. Thin layer chromatography is also used as a final step to confirm the phenotype of the selected mutants. The entire process is illustrated in figure R-1.

Structural predictions suggest that tDGAT has two distinct domains connected by a linker as it is shown in figure R-2 panel (a). Our group has recently reported the absence of conserved sequences in the loop within the WS/DGAT structure from *marinobacter hydrocarbonoclasticus* (Villa et al. 2014). Moreover, there is no evidence of its implication in catalysis or protein folding. Thus, we excluded this connecting loop for mutagenesis focusing the evolution process on the two natural domains separately. In this manner, the screening for mutant selection was simplified and the mutational spectrum reduced.

#### **4.1.1.2. tDGAT mutants with increased in vivo activity are mainly located in the protein surface**

Two mutant libraries of approximately 100.000 mutants each, corresponding to the N and C terminal domains of tDGAT, were created. Individual sequencing of 10 plasmids of the library showed a media of 1,5 non-synonymous amino acid substitutions per mutant. The plasmid library was transformed into *E. coli* BW27783 as described in “Materials and Methods” and subsequently screened following our methodology. In a first round of selection we have screened 10,080 colonies in 3,360 wells (35x96-well plates). We have found 30 wells with a 2-fold or higher increased in fluorescence (compared with the average fluorescence of the entire plate) in the presence of Nile Red. These colonies were isolated and were further analyzed by TLC. As a result, 7 mutants with an enhanced TAG accumulation were selected. Table R-1 shows the amino acid substitutions of the selected mutants that showed higher *in vivo* activity together with positional prediction for all of them.

**Table R-1.** Information of the selected mutants of the direct evolution experiment. Mutants highlighted in bold letters were selected for further analysis.

Amino Acid Substitution	TAG production	WEs production	Mutant Position prediction
<b>TAG and WE Positive mutants</b>			
<b>P35L</b>	+++	++	loop connecting $\alpha 2$ and $\beta 2$ , <b>surface</b>
<b>V87I</b>	++	++	$\alpha 1$ , <b>surface</b>
<b>D114G</b>	++	++	loop, connecting $\alpha 4$ and $\beta 6$ , <b>surface</b>
<b>G17C/V32L/A94G</b>	++	++	loop connecting $\alpha 1$ and $\beta 2$ , <b>surface</b> / loop connecting $\alpha 2$ and $\beta 2$ , <b>surface</b> / loop connecting $\alpha 4$ and $\beta 5$ , <b>surface</b>
V124L	+	+	$\beta 6$ , beta-sheet
L166P	+	+	loop connecting $\alpha 5$ and $\alpha 7$ , <b>surface</b>
D80E	+	+	loop connecting $\beta 4$ and $\beta 5$ , <b>surface</b>
<b>TAG Negative mutants</b>			
P426L	-	+	loop connecting $\alpha 14$ and $\beta 11$ , hydrophobic pocket entrance
D71Y	-	+	loop connecting $\beta 4$ and $\beta 5$ , <b>surface</b>
L29P	-	+	$\beta 2$ , beta-sheet
T5I	-	+	$\alpha 1$ , near active center

Each of the mutants that showed an improvement in neutral lipid accumulation present single amino acid substitutions except for one triple mutant (G17C/V32L/A94G). Surprisingly, all mutations were found at the N-terminal domain, suggesting that it plays a key role in the overall performance of the enzymatic system. In fact, this domain is hosting the conserved catalytic motif (HHxxxDG). According to structural predictions made with Phyre2, all beneficial mutations but one, V124L (that is located in the middle of the sixth beta-strand, closer to the protein core) are predicted to map at the protein surface. The location of these mutations within the secondary structure is quite characteristic since most of them are located in flexible loops that connect structured elements (P35L, D114G, D80E, 166P G17C, V32L, A94G) and only two of them are found in the middle of structured elements. Mutation V87I is located in the middle of the alpha helix  $\alpha 1$  and V124L is in the beta-strand  $\beta 6$ . This observation probably indicates an intrinsic lower capacity of the structured elements to accept amino acid changes.

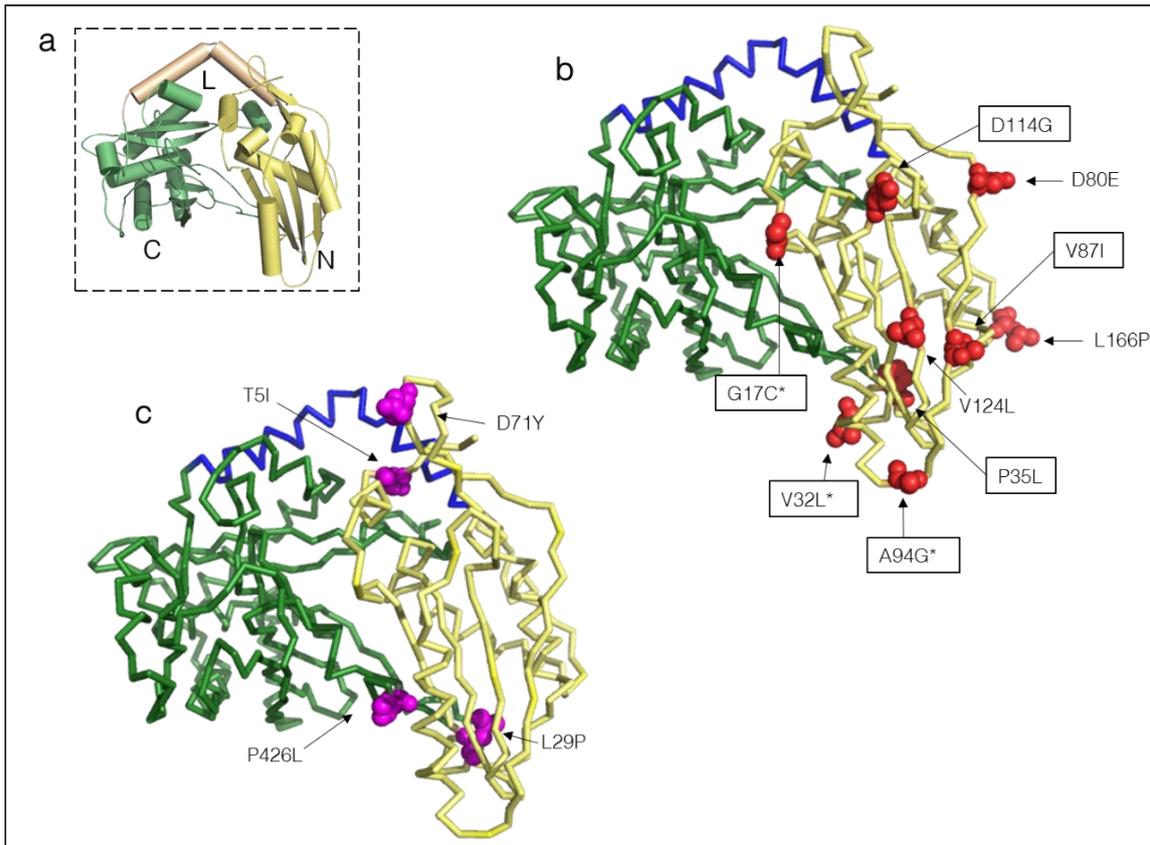
The location of each beneficial mutation for TAG accumulation is shown within the tDGAT modeled structure in Figure R-2 panel (b). This result, together with predictions that describe WS/DGAT as proteins attached to cell inner membranes by electrostatic interactions (Stöveken et al. 2005) suggest the importance of these residues for interactions that lead to proper system operation. All beneficial mutations, especially the ones that lead to major changes in lipid accumulation, the ones highlighted in bold letters in table R-1, lower the hydrophobicity index of the protein. This corroborates the hypothesis that the surface hydrophobicity may play a fundamental role within the protein-protein contacts or the protein contacts with cell membranes that and can be affecting the overall process of lipid accumulation.

#### **4.1.1.3. tDGAT mutants with monofunctional synthesis that produce only waxes were found**

Although the main goal of this study was to select tDGAT variants with improved TAG production, this selection method also provides the ability to select deleterious mutants. Mutants that exhibited diminished fluorescence were selected in order to find deleterious amino acid substitutions for lipid production. In this way we will be able to identify tDGAT residues that are essential for neutral lipid production. Some mutants showing decreases in lipid accumulation were selected and further analyzed by TLC (figure R-3). The Nile red staining protocol cannot distinguish between TAGs and WEs accumulation. The experimental controls that have been used to tune the fluorescence screening procedure were made with the wild type tDGAT that produce more TAG than WEs. Then, what is expected is that when the ranges of ultraviolet light were experimentally established for this experiments the product that mainly contributed to fluorescence shift were TAG. For that reason, these mutants that showed less fluorescence in the primary screening showed a drastic reduction in TAGs accumulation but they still had WEs production. As is shown in the TLC of the figure R-3, the four mutants selected for the analysis only showed traces of TAGs production, but they still produce WEs. Mutants **\_P426L** and **L29P** showed a slightly reduced synthesis capacity but mutants **D71Y** and **T5I** yielded more WEs than wild type tDGAT. Table R-1 shows the amino acid substitutions of the selected mutants deleterious for TAG production together with positional prediction for all of them.

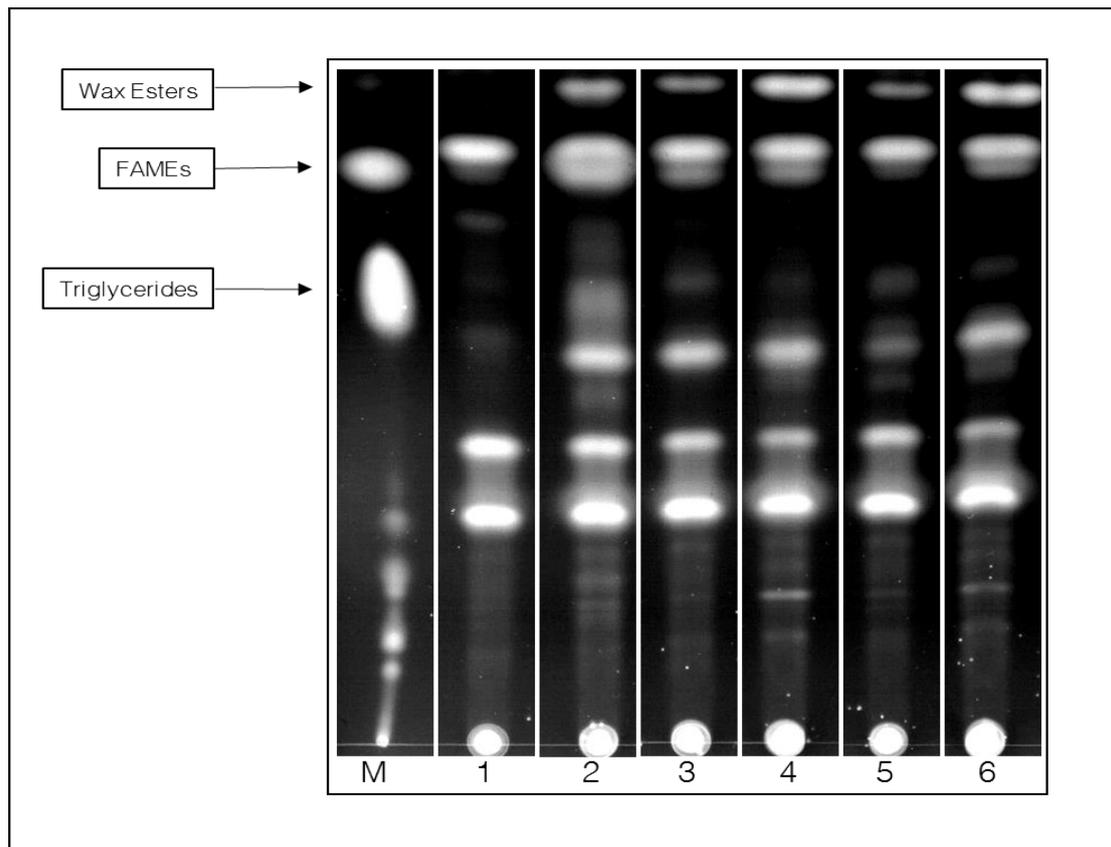
Structural prediction for the localization of these mutations are shown in figure R-2., panel (c) Mutation **P426L** is located in the C-terminal domain while the other three, **D71Y**, **L29P** and **T5I** are located at the N-terminal domain. As in the case of mutations

affecting TAGs production positively, these four single mutations are not located in any of the known catalytic residues and they are not exclusively located in superficial areas. Mutation P426 is predicted to be in a loop connecting the alpha helix  $\alpha 14$  and the beta strand  $\beta 11$  nearby hydrophobic pocket entrance. Substrates with big and flexible moieties as the diacylglycerol could be in contact with this area. Mutation T5I maps in the alpha helix  $\alpha 1$ , just nearby the active center. These both mutations were found in residues that could be involved in catalysis or even in protein substrate contact and recognition processes that can explain the phenotypes observed. Mutation D71Y is located in the disordered loop that connects  $\beta 4$  and the  $\beta 5$  in the protein surface area, and mutation L29P that maps in the middle of the beta strand  $\beta 2$ . Proline is an amino acid that is usually found at the beta hairpins because its dihedral angle allows structural turns, but its present in the middle of a structured beta strand may disrupt their native conformation. Because there is no resolved structure of any WS/DGAT and therefore substrate-protein interaction zones are still unknown, it is difficult to relate the mutations to the associated phenotypic changes. Structural biology studies are needed to further elucidate the structural implications of the selected mutations that strongly affect enzymatic catalysis in a way that remains unknown.



**Figure R-2.** tDGAT structural model showing the location of each of the selected mutations. a) predicted 3D structure of tDGAT with the secondary elements represented as cylinders ( $\alpha$ -helices) and ribbons ( $\beta$ -sheets). N-terminal domain is shown in green, C-terminal domain in yellow and a central linker that connects them in blue. This scheme is also used in panels (b) and (c). b) selected mutations where production of neutral lipids was increased are highlighted with red spheres. Mutations marked with a frame were selected for further TLC-based quantitative analysis. Mutations marked with asterisk (\*) belong to a unique-triple mutant. c) mutations (purple spheres) that showed decreases in TAG accumulation but preserved WE accumulation.

We have observed a direct correlation between the solubility of a WS/DGAT enzyme and its inability to produce TAG. Enzymes of this family that show a good solubility when overexpressed normally are able to *in vivo* synthesize only WEs and lack TAG accumulation. Examples of this are WS/DGAT enzymes from *Alcanivorax* sp., *Marinobacter* sp. or *Acinetobacter* sp. that are able to accumulate only residual amounts of TAG. On the other hand the tDGAT protein seems to be insoluble but able to generate TAG at least when heterologously expressed in *E.coli*. This phenomenon could be related with the tendency that tDGAT shows to form inclusion bodies. In this new lipophilic environment the TAGs can remain bound to the tDGAT proteins and protected from degradation forming lipoprotein complexes with the insoluble proteins. It would be necessary to check the solubility of these mutants with changes in the final production to verify this hypothesis and probe that changes in solubility can affect to the accumulation and stability of the lipids produced.



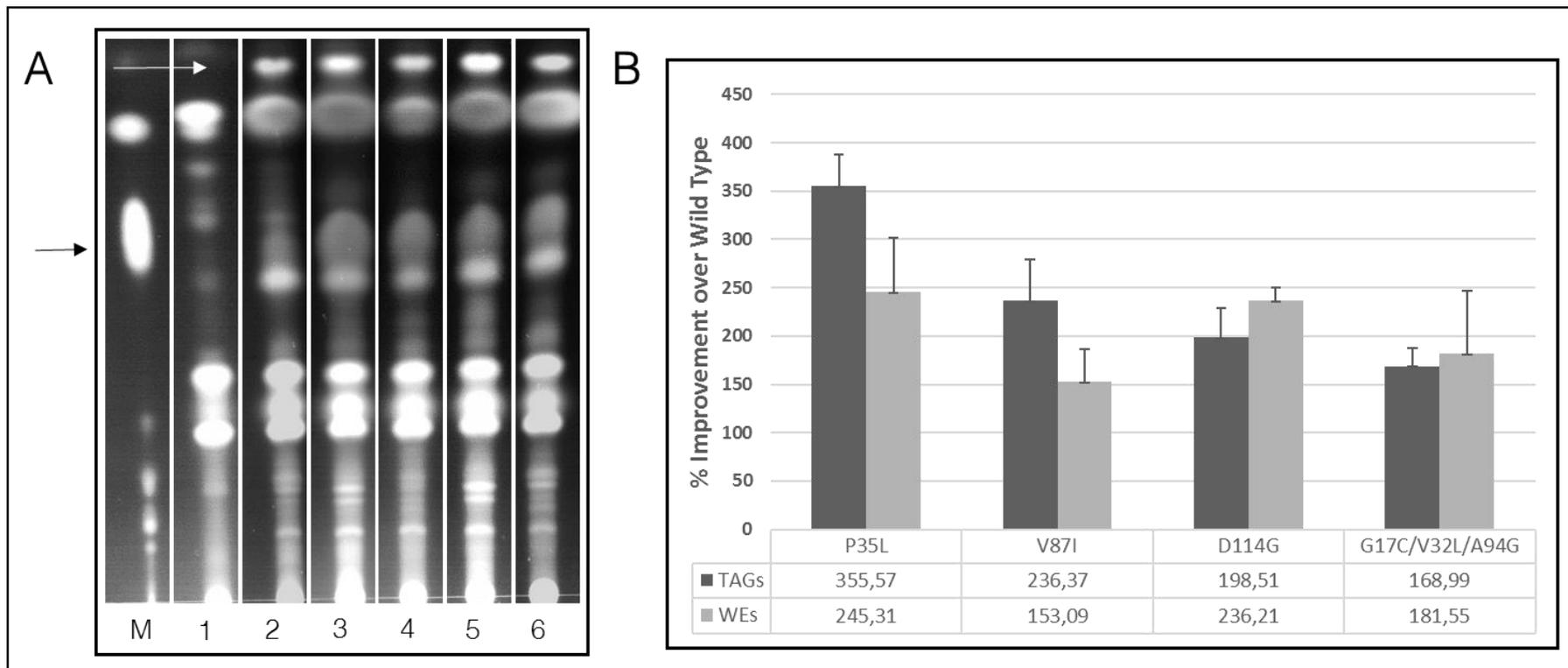
**Figure R-3.** Thin layer chromatography analysis of deleterious mutants for TAG accumulation. M: Marker (triolein and biodiesel); TAG were isolated from BW strains transformed with 1: PBAD; 2: PBAD:tDGAT; 3: PBAD:tDGAT\_P426L; 4: PBAD:tDGAT\_D71Y; 5: PBAD:tDGAT\_L29P; 6: PBAD:tDGAT\_T5I. Mutants selected for the analysis show a strong change in final TAG accumulation, clearly shifting the substrate selectivity to produce WEs in the cases 4 and 6.

#### 4.1.1.4. In vitro recombination of positive mutants did not lead to improved TAG accumulation

In other evolutionary experiments or protein design approaches the first screening and further selection process are successfully coupled with *in vitro* recombination techniques. In this work several attempts have also been made in order to obtain recombinants with additive improvement effects on TAG production. The 4 more productive mutants were subjected to the staggered extension process of *in vitro* recombination (StEP) (Aguinaldo and Arnold 2003). Approximately 1000 recombinant variants were analyzed but unfortunately none of them showed significant improvement (Data not shown).

#### 4.1.1.5. Quantification by densitometric analysis of TAG accumulation

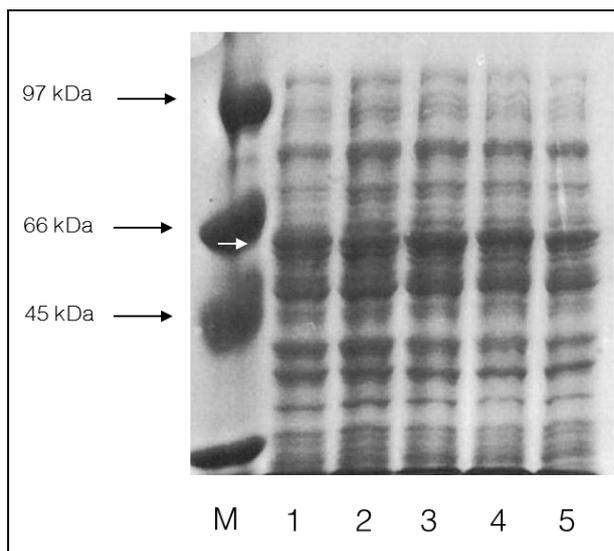
In order to confirm that the mutants produced more neutral lipids than the wild type protein, analysis of the mutants obtained by high throughput screening was further performed by TLC. According to analytical TLC results, four mutants (highlighted in bold in table R-1), three single mutants (P35L; V87I; D114G) and a triple mutant (G17C/V32L/A94G), were selected as the best TAG producers. These mutants were subjected to TAG quantification through densitometry analysis. A representative TLC showing neutral lipid production of these four mutants is shown in figure R-4, panel (a). As it is common to find positional differences after staining the TLC plates that can interfere the quantification, the integrated densities of each TAG spot were normalized with spots in the same region of the TLC (that presented similar background noise) unaffected by tDGAT expression. Final data was calculated as the relative percentage of TAG and WEs production compared with the production of the wild type tDGAT protein. As shown in figure R-4, panel (b), a significant improvement in neutral lipid production was detected in all mutants tested. In particular, the mutant P35L seems to achieve the best performance in TAG production (up to 355 % more than wild type tDGAT). Since we did not find any mutated residue near the hydrophobic pocket where the substrates bind to perform the enzymatic reaction, we should expect similar behavior in the production of both products, TAG and WE. In fact, we have found that these mutations enhance the enzymatic production of both products in the four mutants tested.



**Figure R-4.** TLC and relative quantification of neutral lipid content of the four positive mutants tested. A: Thin layer chromatography from the different mutants in the analysis. M: Marker (triolein and biodiesel); neutral lipids produced by *E. coli* BW27783 transformed with PBAD (lane 1); PBAD:tDGAT (lane 2); 3: PBAD:tDGAT\_P35L (lane 3); 4: PBAD:tDGAT\_V87I (lane 4); 5: PBAD:tDGAT\_D114G (lane 5); 6: PBAD:tDGAT\_G17C/V32L/A94G (lane 6). White and black arrow point to WEs and TAGs, respectively. B: TAGs and WEs were quantified by densitometry and compared with TAGs and WEs produced by the wt tDGAT protein. Values are given as a percentage of improvement over tDGAT wild type.

#### 4.1.1.6. Solubility of highly productive tDGAT mutants has not change

tDGAT tends to aggregate when is purified and *in vitro* activity could not be detected by using Ellman's reagent (DTNB) as we have done with other WS/DGAT enzymes such as Ma2 from *Marinobacter hydrocarbonoclasticus* (Villa et al. 2014). For this reason, it was impossible to characterize the protein mutants obtained by a direct *in vitro* measurement of the kinetic constants with different substrates. We also cannot quantify changes in thermal/pH-range stability, that could provide us more information about how these amino acid changes affect to the protein. Despite this fact, it is known that one of the main characteristic that can affect to the catalysis is the protein stability (Bloom et al. 2004). For this reason, we wanted to detect changes in protein solubility to know if selected variants of the protein presented same intracellular levels as wild type tDGAT. Overexpression experiments were performed under the same conditions for all protein variants with beneficial mutations tested here. Samples were sonicated in order to analyse the soluble fractions as described in materials and methods. Figure R-5 shows a 10% SDS-PAGE gel with the soluble fractions of cell cultures carrying tDGAT mutants. It is shown that a band with a molecular weight consistent with the one predicted for tDGAT is present with a similar intensity in all cases. Thereby the protein expression and solubility of the different mutants selected are similar in our work-conditions, allowing us to exclude these factors as a cause for the phenotype changes detected.



**Figure R-5.** SDS-PAGE analysis of the soluble fraction of cultures expressing the positive mutants found. Soluble fraction obtained after sonication and centrifugation of overnight cultures of *E.coli* cells carrying the different four mutant constructions previously selected for TLC quantification of neutral lipid content. Line M: Marker; 1: Wild type tDGAT; 2: P35L tDGAT; 3: V87I tDGAT; 4: D114G tDGAT; 5: G17C/V32L/A94G tDGAT. White and black arrow point to tDGAT protein and three bands of the protein ladder (Low Range, Bio-Rad).

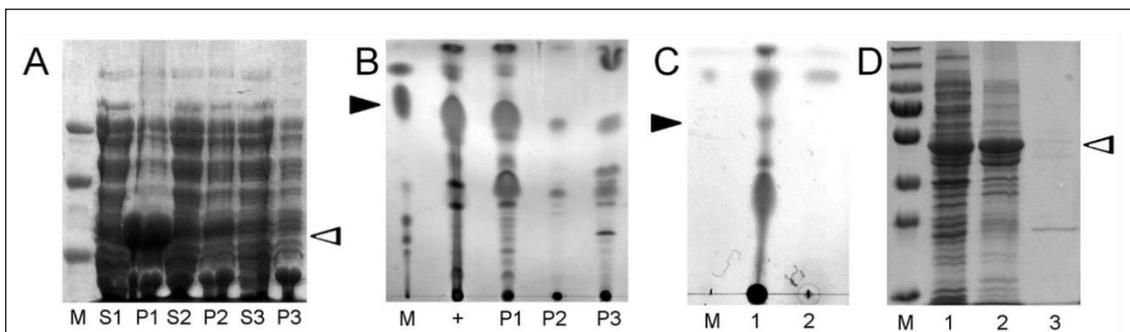
#### 4.1.2. Ultracentrifugation analysis of tDGAT-TAG lipoprotein complexes

As described before, proteins of the WS/DGAT family are prone to bind to lipid membranes. In their natural environment, they are attached to the inner membrane of the cell where the lipid droplets are being formed. When a protein of the WS/DGAT family, like tDGAT, is expressed in an unnatural host some unexpected biophysical interactions may occur. In the case of tDGAT, when we tried to overexpress and purified it in *E. coli*, strong precipitates were always found, regardless of the buffer used for purification. This seems to indicate that the protein is not stable when is out of the cell. Moreover, *E. coli* doesn't have the proper protein pool that is needed to correctly form the lipid droplets and maintain the protein in a natural-like environment into the cytosol.

For that reason, we designed an experiment to answer the question of how this lipid accumulation occurs when the protein that is being expressed cannot be working as in a natural host.

It is widely known that particles of different densities or sizes in a suspension will sediment at different rates, with the larger and denser particles sedimenting faster. These sedimentation rates can be increased by using centrifugal forces. A suspension of broken cells subjected to a series of increasing centrifugal force cycles will yield a series of pellets containing subcellular parts with different sedimentation rates. We wanted to use this technique to separate the different subcellular components of *E. coli* when tDGAT is overexpressed, and check where we were able to detect the protein in the different pellets obtained and also the different lipids.

For that purpose, we centrifugated the sonicated cells 3 times, gradually increasing the centrifugal force set at 2,000 g., 10,000 g. and 55,000 g. where the inclusion bodies, cell debris and cell membranes respectively sediment. At the end of each centrifugation step we collected the supernatant and the pellet and we ran an acrylamide gel and a TLC, both shown in figure R-6.



**FigureR-6.** Ultracentrifugation analysis of tDGAT-TAG lipoprotein complexes. A and B, cell lysate of *E. coli* C41 cells expressing tDGAT were sonicated and separated into phases by centrifugation. Samples from 1 to 3 correspond to consecutive centrifugations of the same sample at 2,000 g, 10,000 g and 55,000 g respectively. (A) SDS-PAGE gel analysis of the supernatants (S1-S3) and pellets (P1-P3) of the different centrifugation fractions. Lane M: Low Range Standard (BioRad) (B) Thin layer chromatography (TLC) of lipid fractions extracted from 50 ml cultures expressing tDGAT. Lane +: Full pellet centrifuged at 55,000g. Lanes 1-4: Centrifugation pellets from P1 to P3. C and D, lipid bodies purification by sucrose gradient. (C) TLC of lipid fractions extracted from a cell lysate of *E. coli* C41 (DE3) cells expressing tDGAT: Lane M, triolein; lane 1, pellet obtained after centrifugation at 3,000 g and lane 2, the supernatant of the sucrose gradient centrifugation at 180,000 g (D) SDS-PAGE gel analysis of the cell lysate (1), proteins in the pellet obtained at 3,000 g (2) and proteins in the supernatant of the sucrose gradient centrifugation (3). Lane M shows the PageRuler Plus Ladder (ThermoFisher). Position of TAGs are marked with black arrows. White arrowheads corresponds to the protein tDGAT.

As is presented in the figure R-6, after the first centrifugation at 2000 g, almost all tDGAT protein falls to the pellet (lane P1 of the acrylamide gel), also happening the same with the neutral lipids (lane P1 of the TLC). The sedimentation of inclusion bodies is characteristic when this centrifugal force is applied, being often associated with unstable or insoluble proteins in the cytoplasm. If neutral lipids of our sample were

associated with the membrane fraction, they should fall in the next centrifugation steps. This is quite surprising because it is indicating the formation of lipoprotein complexes that are clearly forming some kind of aggregation with biophysical properties very similar to the one that inclusion bodies have. Hence, this experiment revealed that the lipids accumulated in the recombinant strain are somehow associated to the acyltransferase responsible for its biosynthesis. As the previous experiment does not have the sensitivity to isolate the lipid bodies from the inclusion bodies fraction, an alternative experiment was designed to be able to tell these fractions apart and accurately identify whether the TAGs remain associated with the aggregated protein or, on the contrary, *E. coli* can produce proper lipid droplets where TAGs can be protected from the cytosol polar environment. We have isolated the lipid bodies by sucrose gradient as described by Ding et al (Ding et al. 2012). As observed in figure R-6 only a residual amount of TAGs was obtained from the top of the sucrose gradient while this fraction does not contain tDGAT. Hence, this experiment revealed that most of the lipids accumulated in the recombinant strain are somehow associated to tDGAT responsible for its biosynthesis. This results indicates that TAG accumulation observed in *E. coli* when expressing tDGAT occurs through an aberrant process where the cell does not have proper control for coupling the lipid production and its storage. The TAGs and WEs produced just remain attached to the protein after synthesis. Apparently, *E. coli* does not have the ability to generate lipid droplets as other organisms that are natural lipid producers like *Rhodococcus* do, probably because it lacks the proteins associated to the process or due to the intrinsic instability of the tDGAT protein. As the tDGAT proteins are not stable or soluble in these conditions they are prone to form inclusion bodies together with the lipids produced. However, these lipoprotein inclusion bodies protect TAGs and WEs from degradation and allows the accumulation that in the case of soluble WS/DGATs (as Ma2 from *Marinobacter hydrocarbonoclasticus*) is not observed.

The new peculiarities of this lipid production system just described here may be actual barriers to fully understand the biological process behind. Somehow the system itself is showing us the the apparent inability of *E. coli* to form proper lipid droplets. Small changes in the protein surface hydrophobicity produced by the amino acid substitutions could just make the lipoprotein complexes aggregation slightly different allowing then to tolerate higher amounts of TAG and WEs into the lipoprotein interface. Despite this, we are in front of a good biological example in which the phenotypes observed from a more macroscopic perspective can be the result of an artifact and not part of a controlled biochemical process.

#### 4.1.3. General Discussion

Here we have applied a new method to evolve WS/DGAT-like enzyme tDGAT towards the improvement of TAG production, achieving an increase of productivity of up to 355%. With this method we were also able to find mutants with different specificity, that mainly produce WEs but lack TAG accumulation. We believe that this method can be universally applied to any neutral lipid producing enzyme with minimal adjustments of some parameters like the wavelengths at which fluorescence are measured, that have to be in accordance to the final enzymatic product, or the host used for protein expression that will depend of the protein characteristics.

It is known that the WS/DGAT family is formed by promiscuous enzymes that naturally have a broad range of substrates. This characteristic is a consequence of having large unspecific hydrophobic cavities around the active site where a different

variety of lipophilic molecules can fit. The low requirements of lipid substrates in terms of specificity to settle in the hydrophobic pocket could explain the difficulties found to select mutants with phenotypic changes mapping in this region nearby the catalytic site, probably due to an intrinsic facility to accept small sequence changes without significant changes in enzyme capabilities.

In this study, seven mutants that improve TAG accumulation and four mutants that showed no TAG but WE accumulation were selected. It is difficult to speculate on whether the phenotypic changes observed in the different selected mutants are due to changes produced by the amino acid substitutions (that could affect enzymatic activity) or due to physical impairments that affect lipid accumulation prompted by the poor structural adaptation of *E. coli* for the formation of lipid droplets. Structural models of the protein as the ones we presented here can provide basic information of domain organization and overall positioning of each amino acid. However, real structures are needed to explain how the structure of the protein is affected with each mutation found and especially how these mutations affect to the relations of the protein with its environment. Unfortunately, crystallization attempts were unsuccessful for any bacterial WS/DGAT protein.

It is widely known that a complex pool of protein is needed to correctly form lipid droplets as was described for bacteria (Ding et al. 2012) an eukarya (Brown 2001). In a natural environment these enzymes are associated with lipid membranes, and then should be fully coordinated with the machinery that generates lipid droplets in organisms like *Rhodococcus opacus* (Wältermann et al. 2005). The protein pool necessary to do that is not present in *E. coli* due its natural inability to accumulate lipids, and for that reason is difficult to obtain high yields when is used as a “lipid factory”. When a strange lipid environment is formed in the cytosol with no tight protein and genetic control, biological aberrations can occur. In the case of tDGAT protein, we always have observed precipitation phenomena during purifications, probably due to the hydrophobic surface domains of the protein that in the natural lipid producers interact with the cell membranes. In these work we have demonstrated that the lipid accumulation observed *in vivo* is not a proper coordinated process as the formation of lipid droplets in *Rhodococcus*, but the result of a lipoprotein aggregation forming inclusion bodies.

Models that attempt to explain how WS/DGAT-mediated lipid droplet formation in bacteria occurs try to explain how these proteins could be associated with inner faces of bacterial membranes during neutral lipid accumulation (Wältermann et al. 2005). It is reasonable to assume that the hydrophobic amino acid residues of the protein surface should contact and interact with the inner membrane of the cell and with other proteins to correctly perform its physiological role. In our directed evolution experiment of tDGAT we have observed that the selected mutants showing improvement in the TAGs accumulation mainly display mutations that increase surface hydrophobicity. This trend is especially strong in the most productive mutants specifically the mutant P35L, the one that produce more neutral lipids. This trend could be indicating that a marked hydrophobicity on the surface of the protein may help in a yet unknown manner to establish membrane-protein or protein-protein interactions that result beneficial for lipid accumulation. This increase in hydrophobicity could stabilize the interactions of inclusion bodies and allow the storage of a greater amount of lipoprotein aggregates. In contrast, amino acid substitutions found in mutants defective for TAGs accumulation do not follow a defined pattern and is difficult to theorize about their meaning. Their capacity to use diacylglycerol as a substrate has been seriously diminished with respect to its ability to capture fatty alcohol. This may be because its amino acid substitutions prevent proper

placement of large substrates such as diacylglycerol into the catalytic center but allow the entry of smaller substrates such as fatty alcohols. Unfortunately to confirm this hypothesis it is necessary the resolution of the three-dimensional structure of the protein to specifically define the protein-substrate contact regions. As was introduced before there is a direct correlation between the solubility of a WS/DGAT protein and its capacity to produce TAGs. Soluble proteins as Ma2 from *Marinobacter hydrocarbonoclasticus* are able to produce traces of TAGs but they accumulate WEs. Experiments to detect changes in solubility in tDGAT mutants only able to produce WEs are necessary to confirm that this phenomena can be conditioning the storage capacity of the metabolic system.

Ultracentrifugation experiments revealed an association between the protein tDGAT and the neutral lipids produced. This connection is probably related both with the hydrophobicity of the amino acid sequence and the high amphipathic nature of tDGAT. This way, the union protein-lipid droplets is likely to form insoluble lipoprotein inclusion bodies. The amino acid modifications observed in tDGAT after the direct evolution experiment and the observation of the lipoprotein complex formation point out the importance of the surface hydrophobic residues in the overall TAG synthesis that had to be modified to improve the process.

Here we have developed a robust method for the evolution of tDGAT, a protein of the WS/DGAT family, which might be applicable to other enzymes with similar characteristics. Protein variants that produce modified phenotypes have been successfully generated and selected according to the initial expectations. However, the interpretation suggested to explain changes in lipid accumulation at molecular level needs more experiments to be confirmed. Biophysical and structural studies are needed to clarify the reality of molecular interactions that WS/DGAT proteins are part of during lipid droplet formation.

## 4.2. Structural and Biochemical analysis of PUFA synthases

Long-chain PUFA production in deep sea bacteria is a process that is not very well understood. Organisms like *Moritella marina* or *Colwellia psychrerythraea* carry *pfa* genetic clusters that are translated into iterative large, multi-domain protein complexes that are able to produce different PUFAs. These complexes resemble bacterial PKS or FAS complexes of mammals and fungi. To date there is no evidence to explain how these enzymes are able to produce molecules like DHA or EPA with a specific double bond pattern. One of the reasons for that is the lack of structural and biochemical studies available in the literature. More research in these fields will shed light and explain the differences between FAS and PUFA synthesis.

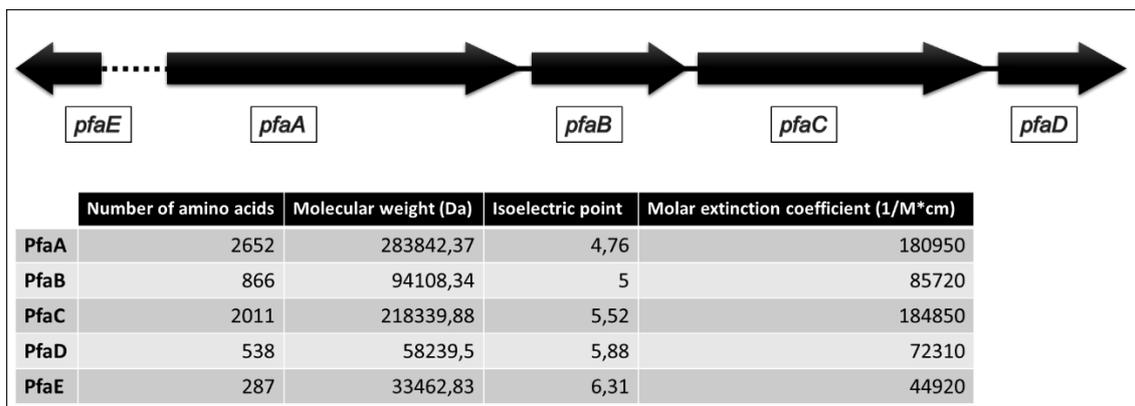
Thus, our goal in this work was to study how the PUFA synthases perform the iterative condensation and reduction cycles in order to produce the final omega-3 FA molecules. Of course, this is a very ambitious project that could be approached from different experimental procedures and techniques. We decided to study by protein crystallography the *pfa* genetic cluster of *M. marina*, one of the better known DHA PUFA synthases. Crystallization of the protein complex or any of its domains would improve our understanding of the system architecture. Apart from that we performed biochemical experiments that helped us to understand how some of the domains of the complex are using and transforming their substrates.

### 4.2.1. *M. marina* Pfa functional domains

The marine bacteria *M. marina* MP-1 is able to produce the polyunsaturated fatty acid docosahexaenoic acid (DHA) using the proteins translated from the *pfa* genetic cluster. Pfa is naturally divided in 5 different polypeptides (PfaA, PfaB, PfaC, PfaD, PfaE). The first 4 proteins (from A to D) are found together in the same DNA region while PfaE is out of the cluster as an independent polypeptide. These proteins are related to FAS and to bacterial PKS proteins, so they also have a modular organization, which means that are formed by a series of individual domains that carry out different biochemical reactions. Due to the lack of structural and biochemical studies that can define how this modular organization is working, there is still controversy to define how the different enzymatic blocks can form these protein complexes. Although there are some approximations in the literature where the organization of these clusters is predicted, we wanted to start from a blind perspective and carry out a bioinformatic study through which we will define the domains that form our Pfa system. This will help to define the genetic constructs that we are going to use for the further experiments.

The first approach to understand Pfa organization was to predict the molecular weight, isoelectric point and the molar extinction coefficient of each of the polypeptides using the computation algorithm of ProtParam (ExPaSy). In figure R-7, the genetic organization of the cluster as well as the predictions obtained for each polypeptide is shown. PfaA and PfaC are the largest proteins with 2652 and 2011 amino acids respectively followed by PfaB that contains 866 amino acids. PfaD and PfaE are the smallest proteins of the system, composed of 538 and 287 amino acids respectively. The isoelectric point of the proteins PfaA, PfaB, PfaC and PfaD are relatively low so they

probably behave as acidic proteins at neutral pH. This characteristic can help to design purification protocols based on anion exchange column chromatography.

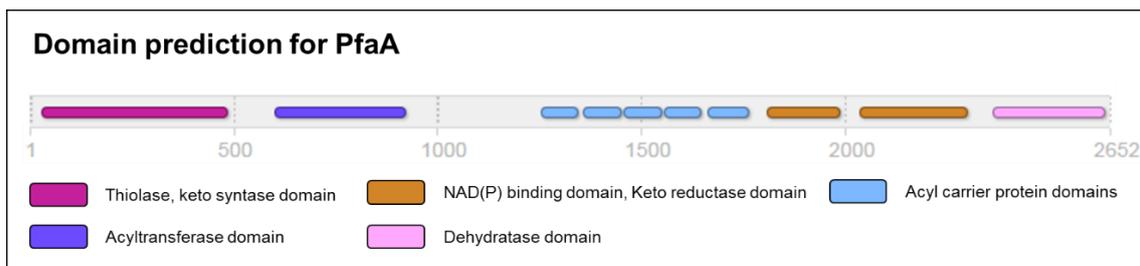


**Figure R-7.** *pfa* genes for biosynthesis of DHA from *Moritella marina*. The functional domain structure is not represented as it will be subject of study in this work. The relative position of *pfaE* in *M. marina* is not represented in the scheme as it is found downstream from the genetic cluster. Table below the representation describe some predicted characteristics of each of the coded proteins (Number of amino acids, molecular weight, isoelectric point and molar extinction coefficient).

After having acquired a general perspective of the *Pfa* genetic organization we will carry out a detailed analysis of each of the polypeptides in which we will identify their functional domains and compare them with other known genetic organizations. PfaE contains a phosphopantetheinyl transferase (PPTase) domain whose function is to activate the ACP domain by adding a phosphopantetheine arm that will be the linker between the ACP protein and the fatty acid moiety. PfaE has been excluded from this analysis since it was demonstrated that is not an essential protein for PUFA synthesis.

#### 4.2.1.1. PfaA is formed by ten individual domains that show five different structural folds

To obtain an overview of the organization of the PfaA protein, which seems to be the most complex one due to its length, various bioinformatic tools have been used. To analyze the protein sequence and predict possible functional domains the InterPro online server (Finn et al. 2017) was used. Figure R-8 shows the prediction for the genetic organization of PfaA where five different folds were found.

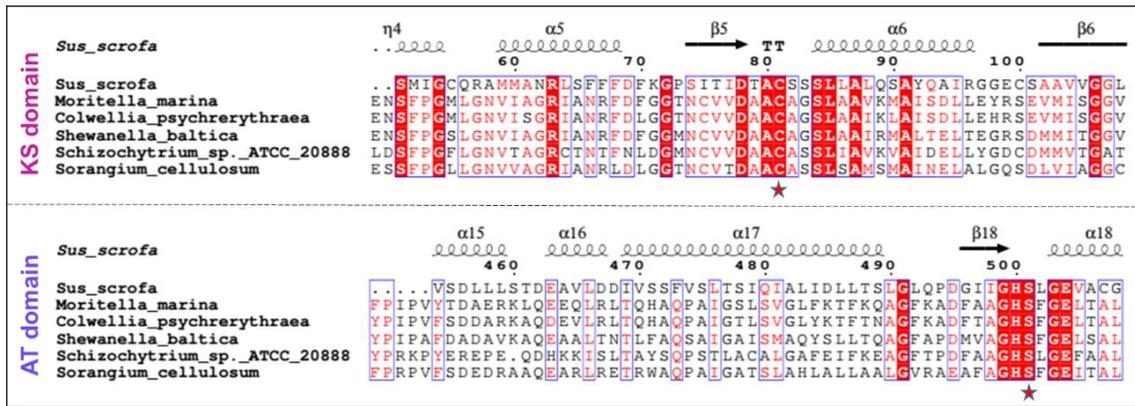


**Figure R-8.** Domain prediction for PfaA. InterPro online tool prediction of protein motifs and folds within the PfaA protein sequence. Each domain predicted is represented in a different color.

In the N-terminal domain a thiolase-like folding that resemble a canonical keto synthase domain is found (aa 29-487). This domains is very close to an acyltransferase domain (aa 601-923) with which should be in close association as in the case of the mammalian FAS previously described (Maier, Leibundgut, and Ban 2008b). In the middle of PfaA, five acyl carrier protein folds are predicted to be in a tandem arrangement (aa 1255-1768) as the ones found in organisms like *Schizochytrium*. These ACPs are followed by two NAD(P) binding domains (aa 1810-1992 and 2037-2303) and a dehydratase domains (aa 2366-2642) very similar to the ones found in some PKS proteins. A blast search for the sequence of the NAD(P) binding domains allows us to identify them as regular keto reductase domain.

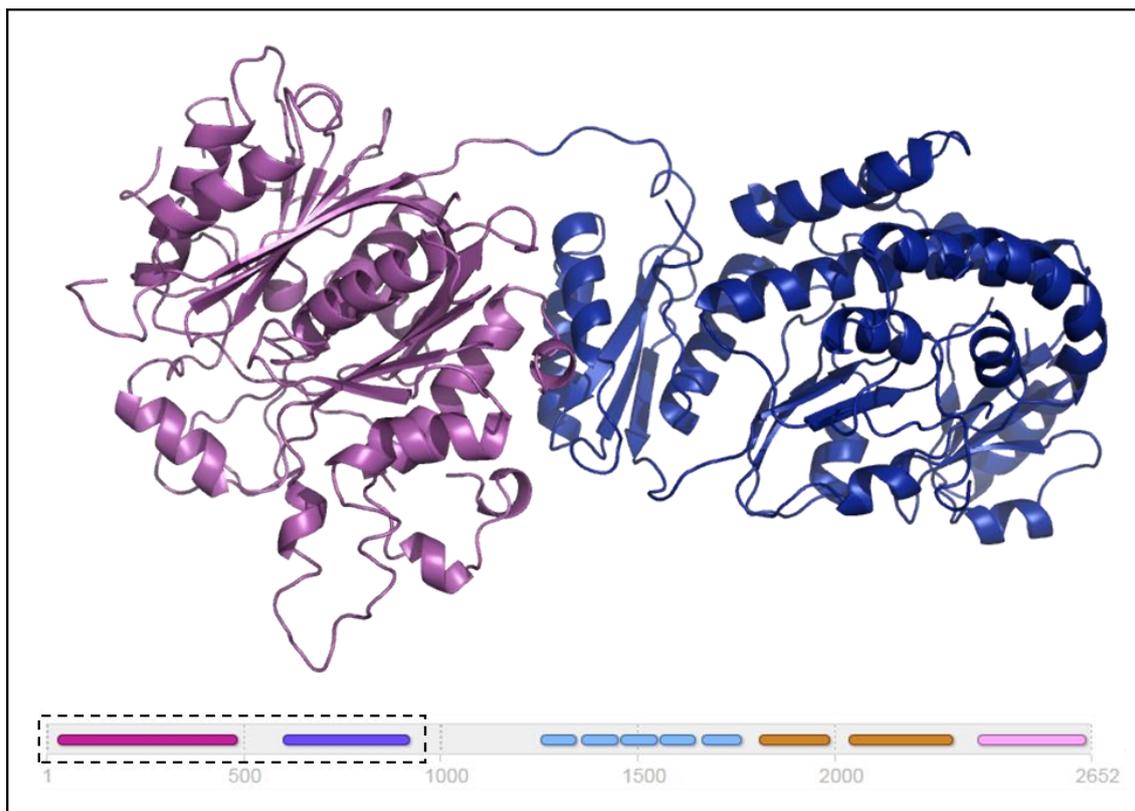
To verify the existence of these five different domains within PfaA protein we have used the Phyre2 server with which structural predictions can be made by comparison with previously solved protein structures. In this way, we have obtained structural models of each of the PfaA domains. In addition to this, a sequence alignment study was performed to compare our sequences with those of other omega-3 fatty acid producing organisms like the marine bacterias *C. psychrerythraea* and *Shewanella baltica*, the eukaryotic organism *Schizochytrium* or the *Myxobacteria Sorangium cellulosum*.

The first two domains, corresponding to the motifs KS and AT were treated as a single di-domain due to the number of examples in the literature (Khosla et al. 2007; Maier, Leibundgut, and Ban 2008b) in which similar domains have crystallized together. Figure R-9 shows an alignment of representative sequences including the KS-AT domain from *Sus scrofa* that will be used as a template for the structural modeling as the most representative mammal FAS structure was solved using the protein complex from this organism. Only the region of the alignment corresponding to the active sites of both domains, KS and AT is shown. Protein cores of both domains are similar in all sequences analysed, being both active sites highly conserved, specially the catalytic residues, a cysteine for the KS (C229) and a serine (S703) for the AT domain that are marked with red stars in the figure R-9. The complete sequence alignment can be found at supplementary material, figure sm-1.



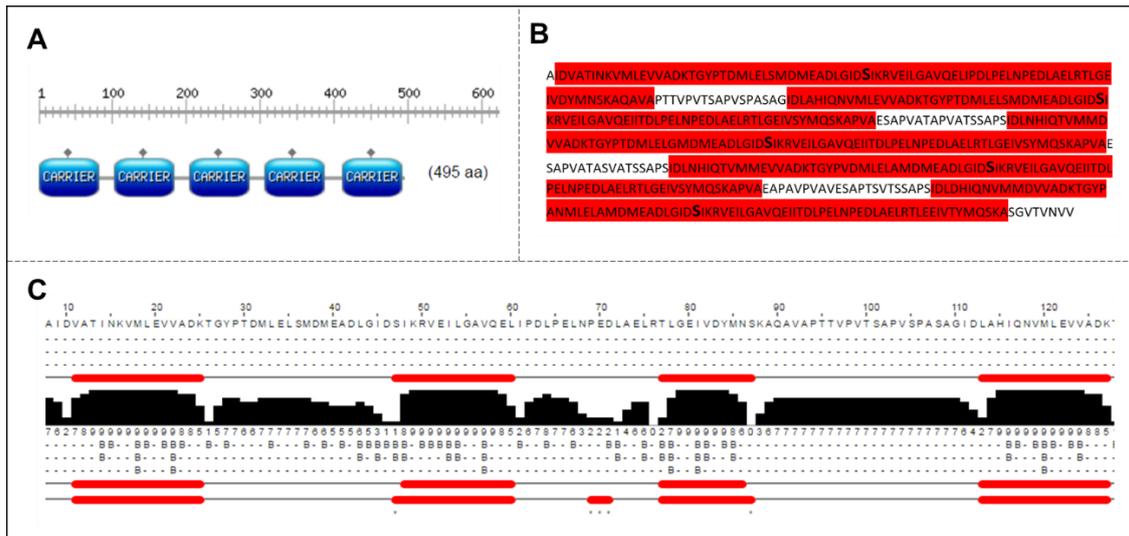
**Figure R-9.** Structure-based multiple sequence alignment of KS-AT di-domain of PfaA. An alignment was made in order to show evolutionary conservation of amino acid residues in active centers of both KS and AT domains of the PfaA protein with other representative omega-3 producers. Top and bottom panels show the alignment of the protein cores for the KS and AT domain, respectively. Red starts at the bottom of the alignments indicate the conserved active residues, a cysteine for the KS domain and a serine for the AT domain. Mammalian FAS (accession number: 2VZ8) was used as a model for secondary structure prediction showed above the alignment.

A structural model for the KS-AT didomain protein was done with Phyre2 using the mammalian FAS structure as template (accession code: 2ZV8) and the result is shown in figure R-10. The model predicted the folds of the two natural KS and AT domains that were colored in purple and blue respectively. KS domain show a duplication of a regular thiolase fold as the rest of the KS known were AT domain showed a  $\alpha/\beta$ -hydrolase core fold with a central, four-stranded parallel  $\beta$ -sheet surrounded by  $\alpha$ -helices and a ferredoxin-like subdomain. This prediction, together with the alignments of the figure R-9 indicate as highly probable the existence of a KSAT didomain at the N terminal part of the PfaA protein.



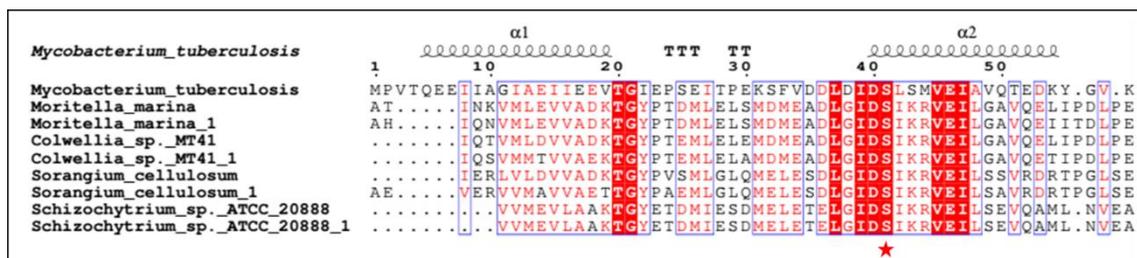
**Figure R-10.** Structural model for the KS-AT di-domain of PfaA. The N terminal part of PfaA was modelled using the Phyre2 online server. Domain prediction by InterPro is represented at the bottom of the figure with a dotted rectangle indicating the sequence region used for the structural modelling. The KS-AT-like domain cartoon representation was generated with Pymol and painted in purple (KS-like) and blue (AT-like) respectively. Mammalian FAS (accession number: 2VZ8) was used as the structural template.

Next conserved motifs that are predicted to be in the PfaA protein are the five tandem ACP domains. Since they contain a very short sequence and are arranged very close to each other, it is difficult to define exactly the positions of each of them. In order to solve this problem a detailed analysis of their sequence has been carried out using the ScanProsite and the Jpred web tools to predict the positions of their active sites and the secondary elements present in each of the individual domains. Panel A of figure R-11 shows the general arrangement obtained from ScanProsite for the five ACP modules with a diamond mark pointing out the serine active sites of each of them. Panel B shows highlighted in red the five ACPs predicted motifs within the entire protein sequence with the serine active sites in bold. Panel C at the bottom of the figure R-11 shows the Jpred estimation for the secondary elements for the first ACP of the polypeptide where four alpha helices are differentiated. Each of the ACP domains consists of about 80 amino acids in length and their amino acid sequences are highly conserved among the Pfa-like cluster of Omega-3 producer organisms.



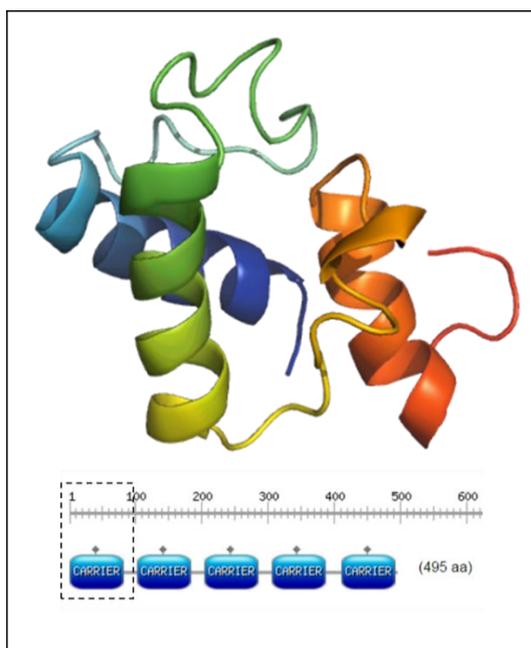
**Figure R-11.** Domain secondary structure prediction for tandem ACP proteins. Panel A shows ScanProsite results together with ProRule-based predicted intra-domain features where each individual ACP domain is represented as a blue boxes, with the O-(pantetheine 4'-phosphoryl) serines highlighted with grey diamonds on the top. Panel B shows the complete sequence of the five tandem ACPs with the predicted individual domains highlighted in red and the serine residues in bold. Panel C shows the secondary element prediction given by the Jpred server where the four alpha helices of the first ACP repetition can be differentiated.

ACP single domains retain a very similar structure and sequence in all species analysed. Since these proteins consist of tandem repetitions of the same motif with small variations, only two of them were selected from each organism to be included in the sequence alignment shown in figure R-12. It can be seen how the active serine, where the PPT motif is anchored in the holo- form, is conserved. The complete sequence alignment can be found at supplementary material, figure sm-2.



**Figure R-12.** Structure-based multiple sequence alignment of two of the five ACP domains of PfaA. An alignment was made in order to show evolutionary conservation of amino acid residues in the two firsts ACP domains of the PfaA protein with other representative omega-3 producers. Red stars at the bottom of the alignments indicate the conserved active serine residues. Mycobacterium tuberculosis ACP (accession number: 1KLP) was used as a model for secondary structure prediction showed above the alignment

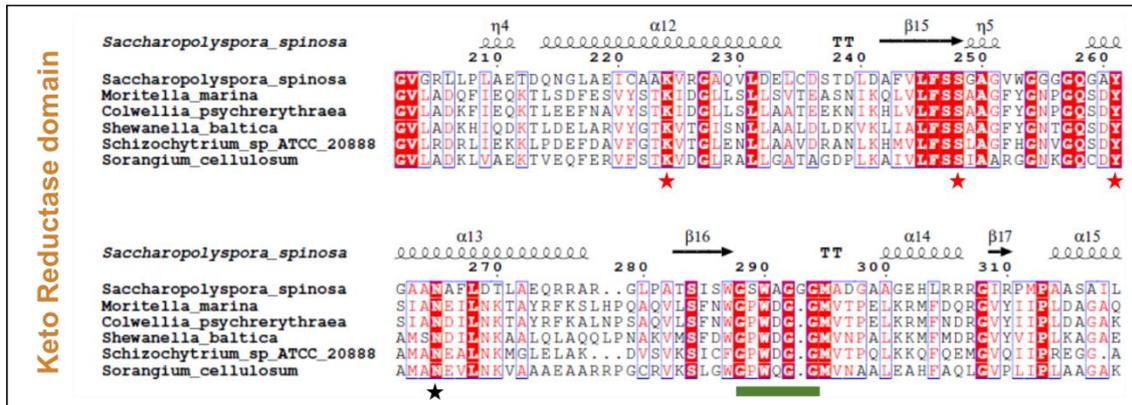
A structural model for the first of the tandem ACPs was done with Phyre2 using the ACP from *Mycobacterium tuberculosis* as template (accession code: 1KLP) and the result is shown in figure R-13. The structure is similar to all ACPs known structures and show four  $\alpha$ -helices with a long but structured loop connecting the first and the second one. These helices are oriented in an up-down-down topological arrangement to form a helical bundle plus a short fourth helix. The supra structure formed by the tandem ACPs of the organism *Photobacterium profundum* has recently been solved by the Abel Baerga and his group in Puerto Rico using nuclear magnetic resonance (Trujillo et al. 2013b). They suggested a monomer form and an elongated beads-on-a-string structure. In this way, the ACPs could adopt several conformations in solution, indicating their structural flexibility.



**Figure R-13.** Structural model for one of the ACP domains of PfaA. The first ACP repetition modelled using the Phyre2 online server. Structure prediction was represented as cartoon using the Pymol software. Below the model, scanProsite domain prediction is shown with the first ACP highlighted with a dotted rectangle indicating the sequence region used for the structural prediction. The ACP from *Mycobacterium tuberculosis* (accession number: 1KLP) was used as a structural template.

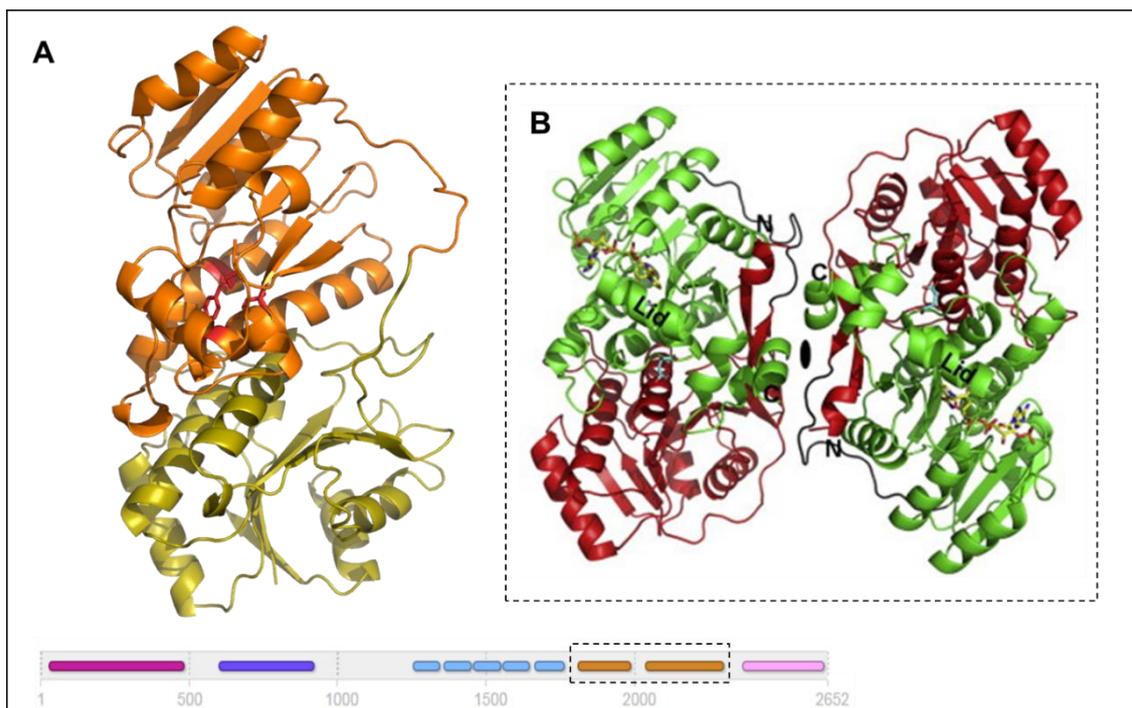
Downstream the PfaA sequence, two KR domains were predicted as NADP(H) binding domains. These domains usually have a conserved S-Y-K triad near the loop connecting the  $\beta$ 5 and  $\alpha$ 5 that differentiates them from ER domains that show a Y-(X)<sup>6</sup>-K motif. When we tried to perform sequence alignment it was impossible to find this typical S-Y-K motif within the first KR repetition. The lack of a known active center and the poor general sequence conservation is quite bad in this region, indicating that this domain can be a structural element rather than a catalytic domain. This domain arrangement where a catalytic domain is located next to a structural pseudodomain is consistent with the described for mammalian FAS (Maier, Leibundgut, and Ban 2008b) and bacterial KR domains from PKS systems like the one described from *Streptomyces nodosus* (Zheng et al. 2010). Figure R-14 shows an alignment of the second KR domain that we predict to be catalytic since the typical catalytic triad can be found. As in the case of mammalian FAS the asparagine and lysine involved in proton replenishment are swapped which indicates a strong conservation between these two protein systems. Catalytic residues were marked with red stars and asparagine with a black one. Template used for structural annotation was the KR domain from *Saccharopolyspora spinosa*

(accession code: 4IMP). The complete sequence alignment can be found at supplementary material, figure sm-3.



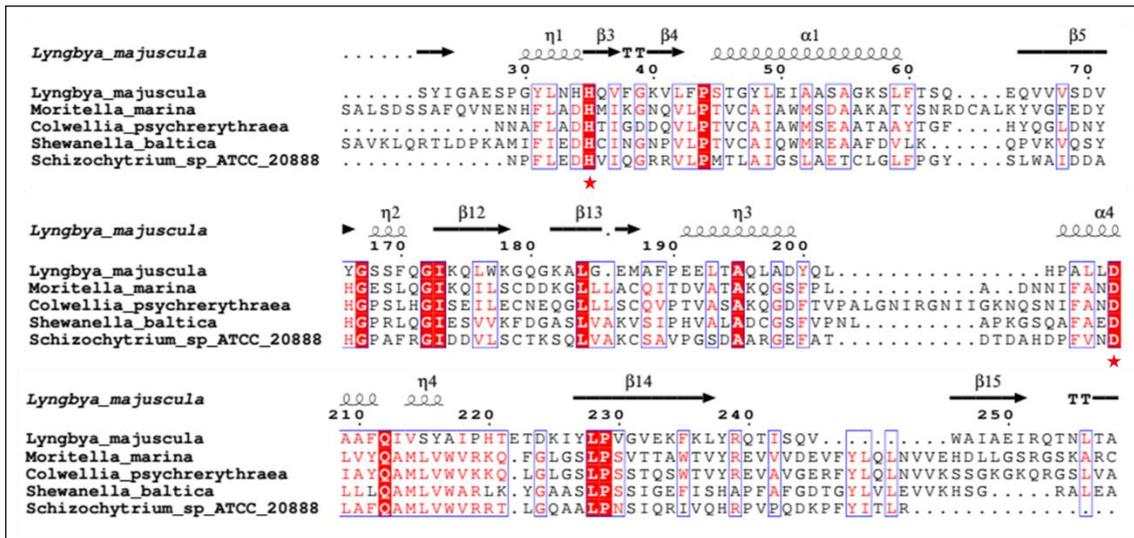
**Figure R-14.** Structure-based multiple sequence alignment of the KR domain of PfaA. An alignment was made in order to show evolutionary conservation of core and active site amino acid residues of the second KR domain of the PfaA protein with other representative omega-3 producers. Red starts at the bottom of the alignments indicate the conserved active residues, a K-S-Y triad and a black star show a conserved Asn that together with the active site Lys form a proton-wire. The green bar shows the conserved GXGXXG dinucleotide binding motif. A keto reductase domain from *Saccharopolyspora spinosa* (accession number: 4IMP) was used as a model for secondary structure prediction showed above the alignment.

Structural prediction of both domains, the catalytic and the structural one was performed using the same template (4IMP) as in the structural alignment and is represented in figure R-15, panel A. The model shows a conserved Rossmann fold with a twisted, parallel  $\beta$ -sheet composed of seven  $\beta$ -strands flanked on both sides by eight  $\alpha$ -helices. As the mammalian FAS or the KR domain from *Saccharopolyspora spinosa*, that is shown in panel B of figure R-15, the KR didomain of PfaA showed the typical KR arrangement with a structural element on the bottom (colored in yellow) and a catalytic domain on the top of the figure R-15 panel A (colored in orange). Within the catalytic domain, the active site is perfectly conserved (the catalytic residues are shown in red in the predicted model) as well as the “lid” formed by an alpha helix in whose proximity the conserved GXGXXG residues that were highlighted with a green rectangle in the alignment of the figure R-14 form the dinucleotide binding motif.



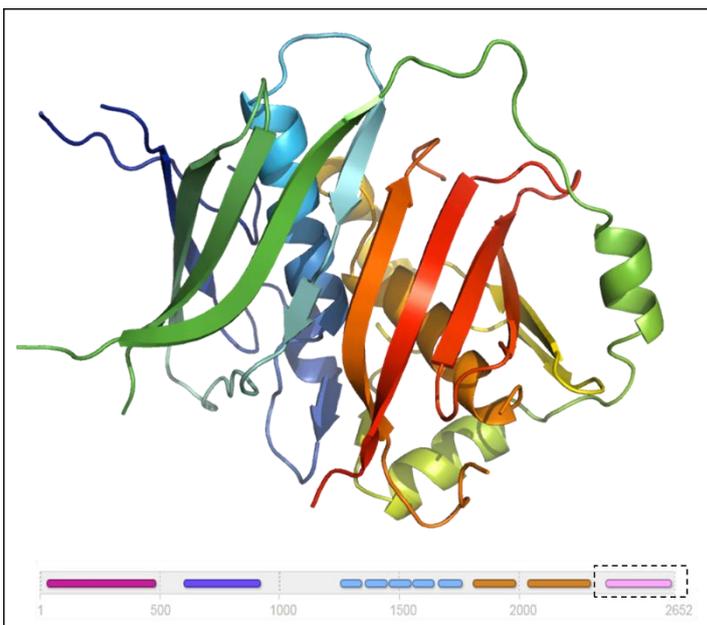
**Figure R-15.** Structural model for the KR'-KR di-domain of PfaA. The zone predicted to fold as two KR domains of the C-terminal part of PfaA was modelled using the Phyre2 online server. Domain prediction by InterPro is represented at the bottom of the figure with a dotted rectangle indicating the sequence region used for the structural prediction. The KR'-KR-like domain cartoon representation of panel A was generated with the Pymol software and painted in orange (KR-like) and yellow (pseudo KR-like) respectively. A KR protein from *Saccharopolyspora spinosa* (accession number: 4IMP) was used as a structural template. Panel B shows the AmpKR2 protein structure that was crystallized as a dimer. The structure shows two monomers related through a twofold axis that represents the dimeric organization of some KRs within intact PKS modules. The structural subdomain is colored in red, and the catalytic subdomain is colored in green.

Last domain that was predicted to be in PfaA is a DH domain. The active site of these domains is normally formed by a His and Asp or Glu residues. A sequence alignment was performed using the same representative sequences of the previous domains but adding the structural template that we will use for the modeling (the dehydratase domain from CurF; accession code: 3KG6) in order to find the His-Asp/Glu catalytic motif as well as to check the general conservation of the protein. There was a lack of defined homology for the organism *S. Cellulosum* so we decided to exclude it from the analysis. As can be observed in figure R-16 there is a full conservation of both active site residues, a histidine and a aspartic acid, marked below the sequence with red stars. The complete sequence alignment can be found at supplementary material, figure sm-4.



**Figure R-16.** Structure-based multiple sequence alignment of the DH domain of PfaA. An alignment was made in order to show evolutionary conservation of core and active site amino acid residues of the C-terminal DH-like domain of PfaA protein with other representative omega-3 producers. Red stars at the bottom of the alignments indicate the conserved active residues, a His-Asp diad. The DH domain from CurF module of curacin polyketide synthase from *L. majuscula* (Accession number: 3KG6) was used as a model for secondary structure prediction showed above the alignment.

The structural prediction for the DH domain that is located at the C-terminal position of PfaA is shown in figure R-17. A repetition of a hot dog motif can be observed forming two differentiated domains. The model reveals then a homodimeric enzyme that contains both active sites residues contributed by the two subunits of the structure like in the FabA and FabZ proteins from *E.coli*.

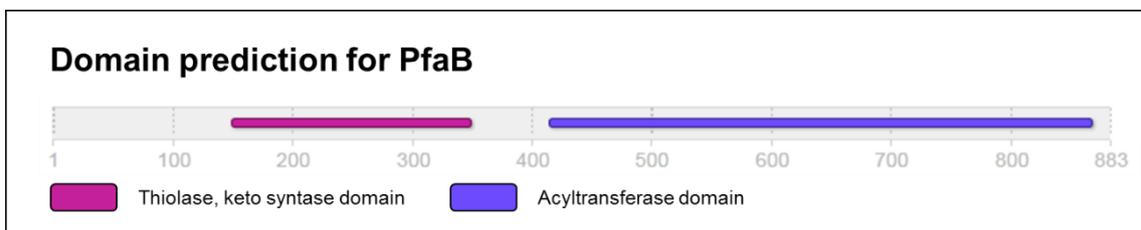


**Figure R-17.** Structural model for the DH domain of PfaA. DH domain of PfaA was modelled using the Phyre2 online server. Domain prediction by InterPro is represented at the bottom of the figure with a dotted rectangle indicating the sequence region used for the structural prediction. The DH-like domain cartoon representation was generated with Pymol. The DH domain from CurF module of curacin polyketide synthase (Accession number: 3KG6) was used as a structural model for the alignment.

The modular organization of PfaA, that is named PFA1 in *Schizochytrium* or Pfa2 in bacteria of the *Myxobacter* genus like *S. cellulosum*, is strictly conserved among the omega-3 producer organisms. All different PfaA-like proteins show a KS-AT didomain followed by a series of ACP domains in tandem whose number of repetitions can vary from five to nine in some representatives of the *Schizochytrium* genus. These tandem ACPs are always followed by a KR'-KR di-domain and a DH domain similar to the ones normally found in PKSs. This conservation among nature can be indicating that its structural conformation is essential for the proper function of the PUFA synthase protein complex.

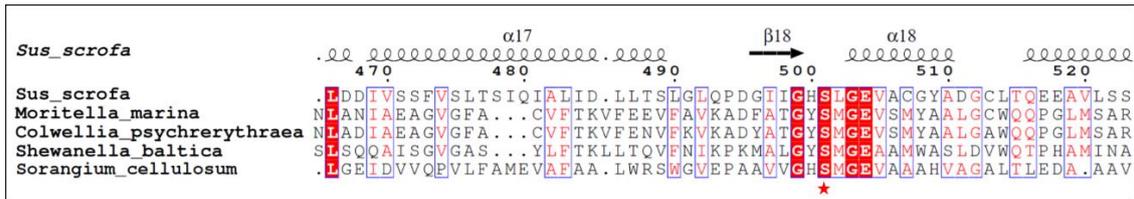
#### 4.2.1.2. PfaB has an acyltransferase and a pseudo keto synthase domain

To analyze the protein sequence and predict possible functional domains the InterPro online server was used. Like in the case of PfaA Figure R-18 shows the prediction for the genetic organization of PfaB where two different folds were found.



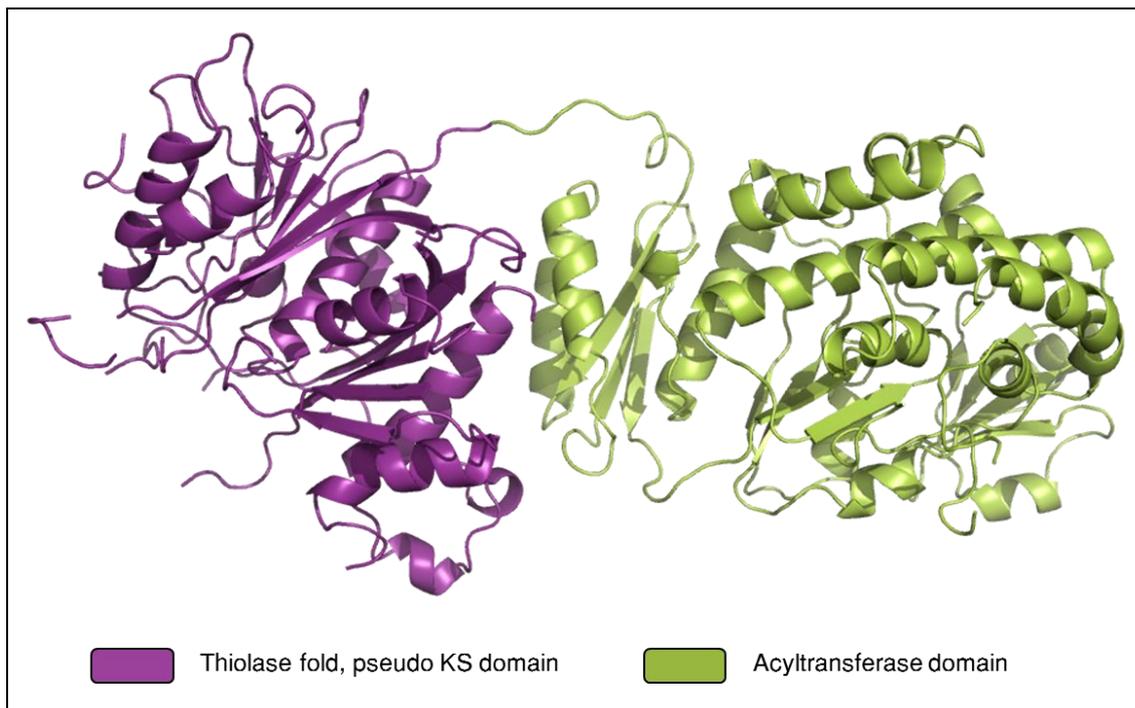
**Figure R-18.** Domain prediction for PfaB. InterPro online tool prediction of protein motifs and folding within the PfaB protein sequence InterPro. Each predicted domain is represented in a different color.

PfaB is a protein that is predicted to have an AT domain found at the C-terminal position. In some organisms like *C. psychrethraea* and *M. marina*, both DHA producers, the structural predictions suggest that it may also have an adjacent KS-like domain at the N-terminal part of the protein. This domain is not very well defined as it does not have the Cys-His-His catalytic triad thus its catalytic role remains unclear. PfaB was proposed to play a role in the determination of the final product of the synthesis in some marine bacteria (Orikasa et al. 2009a), but the mechanism by which this happens is not fully understood. To investigate the sequence conservation of PfaB-like proteins within omega-3 producers an alignment was done. *S. cellulosum* was excluded from the analysis as *Myxobacteria* only have one AT domain that previously showed higher homology with PfaA. Sequence conservation of the N-terminal domain for the rest of the organisms analyzed was practically null, especially in EPA producers like *Shewanella baltica* where the sequence is shorter or inexistent. Thus, we decided to show only the conservation of the C-terminal domain that maintained the AT serine residue as well as most of the important residues to maintain the protein core (figure R-19). The AT domain of the mammalian FAS synthase (accession code: 2VZ8) that will be used as template to generate the structural model, was also included in the analysis. The complete PfaB sequence alignment can be found at supplementary material, figure sm-5.



**Figure R-19.** Structure-based multiple sequence alignment of AT domain of PfaB. An alignment was made in order to show evolutionary conservation of amino acid residues in active centers of the AT domains of the PfaB protein with other representative omega-3 producers. The red star at the bottom of the alignments indicates the conserved active residue, a serine. Mammalian FAS (accession number: 2VZ8) was used as a model for secondary structure prediction showed above the alignment.

PfaB structural model was build using the same template as in the previous sequence alignment and the result is shown in figure R-20. The prediction suggests a folding similar to the one found for the N-terminal domain of PfaA. After the sequence analysis we can conclude that although the structural predictions indicate the existence of a thiolase-like folding similar to a canonical KS domain, PfaB does not have at least the same catalytic mechanism due to the lacking of the Cys-His-His catalytic triad. When looking in the PfaB structural model a Leu-Ser-Glu triad can be found instead but this motif is not conserved through evolution. It is possible that this N-terminal domain is only a structural element and therefore lacks catalytic activity. In contrast, the C-terminal AT domain showed a typical  $\alpha/\beta$ -hydrolase core fold with a ferredoxin-like subdomain like the one showed for PfaA. This result together with the conservation of the active serine involved in the catalytic transference of acyl groups suggest the existence of an actual AT domain whose catalytic activity has to be proved.

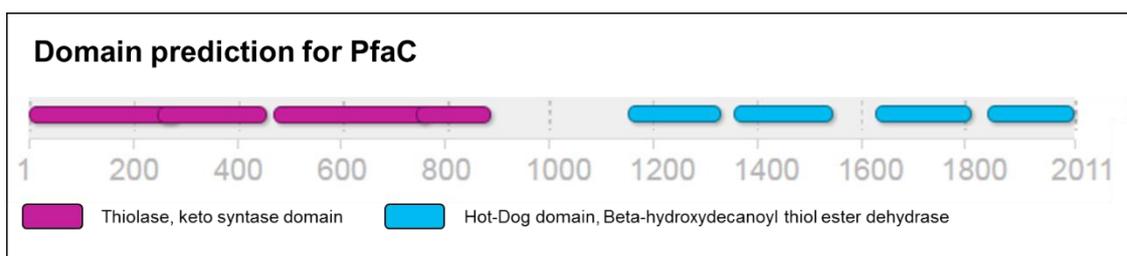


**Figure R-20.** Structural model for the KS-AT di-domain of PfaB. The PfaB protein sequence was modelled using the Phyre2 online server. The KS-AT-like domain cartoon representation was generated with Pymol software and painted in purple (KS-like) and green (AT-like) respectively. Mammalian FAS (accession number: 2VZ8) was used as a structural template.

In *Schizochytrium*, PfaB-like proteins are not individual polypeptides but part of a multi modular protein named as PFA2, that is similar to the PfaC protein from *Marinobacter*. In bacterias of the *Myxobacter* genus homologous of this AT domain can be found at *pfa3*. Is of particular interest to know the enzymatic role of this AT domain and to find out how the organisms that do not possess it can carry out the synthesis.

#### 4.2.1.3. PfaC has two keto synthase and four dehydratase domains.

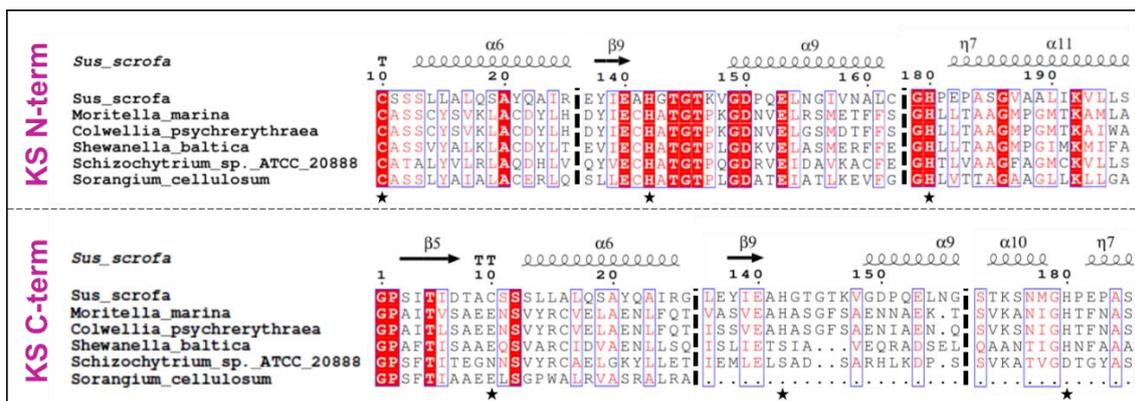
PfaC is the second larger polypeptide of the *pfa* cluster and is also a multi modular protein. To obtain an overview of its organization the same bioinformatic tool, InterPro (Finn et al. 2017), previously used for PfaA, has been tested. In this way we have analyzed the protein sequence and predicted possible functional domains. Figure R-21 shows the prediction of the genetic organization of PfaC where two different folds were found.



**Figure R-21.** Domain prediction for PfaC. InterPro online tool prediction of protein motifs and folds within the PfaC protein sequence. Each predicted domain is represented in a different color.

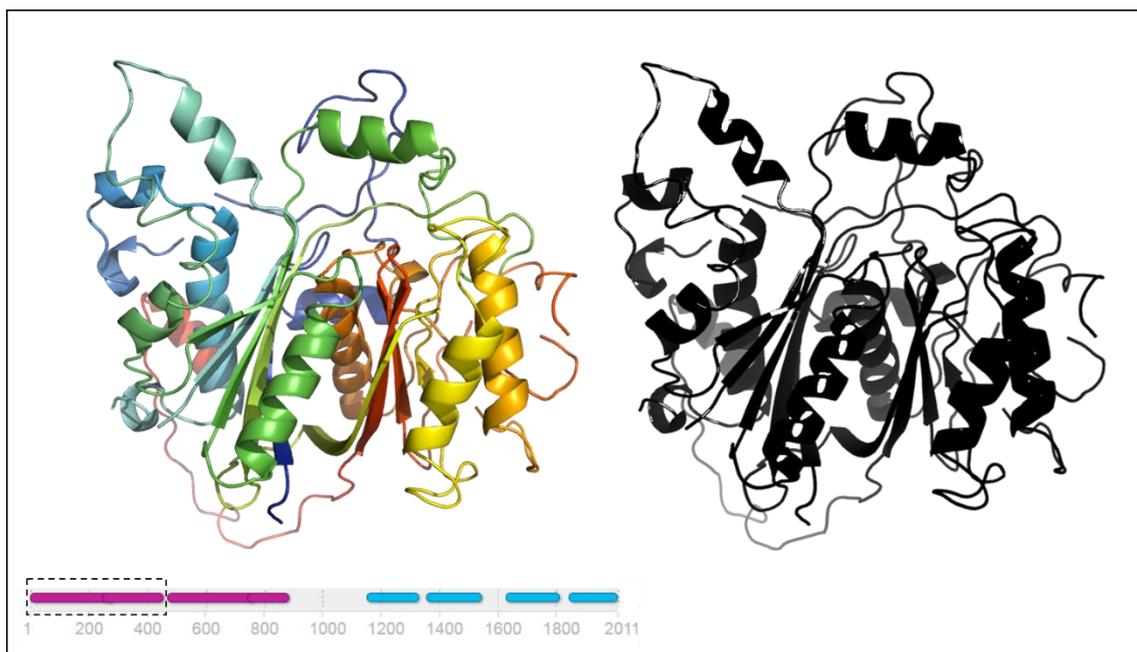
In the N-terminal domain of PfaC two domains that are repetitions of a thiolase-like folding were found. This configuration indicates the presence of two regular KS domains (aa 2-455 and aa 467-883). These domains are quite far from the next functional domain and at least 150 amino acids are predicted as disordered. Just beyond this disordered area, four hot-dog domains predicted to be similar to the ones found in the beta-hydroxydecanoyl thiol ester dehydrases FabA and FabZ were found (Aa 1152-1331; Aa 1358-1546, Aa 1630-1814; Aa 1845-2010). To verify the existence of these domains within PfaC we followed the same procedure as with PfaA. Sequence alignment were made to compare its sequence with similar proteins found in omega-3 producers and structural predictions were made using the Phyre2 online server to verify the protein folding of each domain.

The first two KS domains share homology with the keto synthase of mammalian FAS thus its sequence was also included in the alignment showed in figure R-22 and used for the modeling of the protein structure. The FAS domain comprises only half of the 900 amino acids present in the KS-KS di-domain, so the alignment has been performed twice, one for the complete di-domain and one for the second KS. Only the region of the alignment corresponding to the active sites of both domains is shown but the complete alignment can be found at supplementary material, figure sm-6. Protein cores of both KS domains are different since the residues of the catalytic triad at the C-terminal domain are not fully conserved, thus this one should not be called KS since it probably lacks keto synthase activity. Whereas the N-terminal KS has the typical Cys-His-His triad, the C-terminal domain has a conserved glutamic acid and two histidines, that are not very well conserved. Similar catalytic triad can be found in special keto synthase domains of bacterial PKS usually known as chain length factor or CLF. We have introduced this kind of catalytic entities in the introduction section and there will be a notorious part of this work in the following sections. Since N-terminal part of PfaC is highly related to chain length factor (CLF) is convenient to remember that the KS-CLF domains from PKS has been described to dictate polyketide chain length, to catalyze chain elongation and also to be a factor required for polyketide chain initiation. The CLF was also proposed to be a factor required for polyketide chain initiation (Bao, Wendt-Pienkowski, and Hutchinson 1998; Bisang et al. 1999; Carreras and Khosla 1998; Matharu et al. 1998).



**Figure R-22.** Structure-based multiple sequence alignment of KS-KS di-domain of PfaC. An alignment was made in order to show evolutionary conservation of amino acid residues in active centers of the two KS-like domains of PfaC protein with other representative omega-3 producers. Top and bottom figure show the alignment of the protein cores for the first and the second KS-like domains respectively. Black stars at the bottom of the alignments indicate the conserved active residues, a C-H-H motif for the N-terminal KS domain and E-H-H for the C-terminal one. The dotted lines indicate fragments of the alignment not shown in the figure. Mammalian FAS (accession number: 2VZ8) was used as a model for secondary structure prediction showed above the alignment.

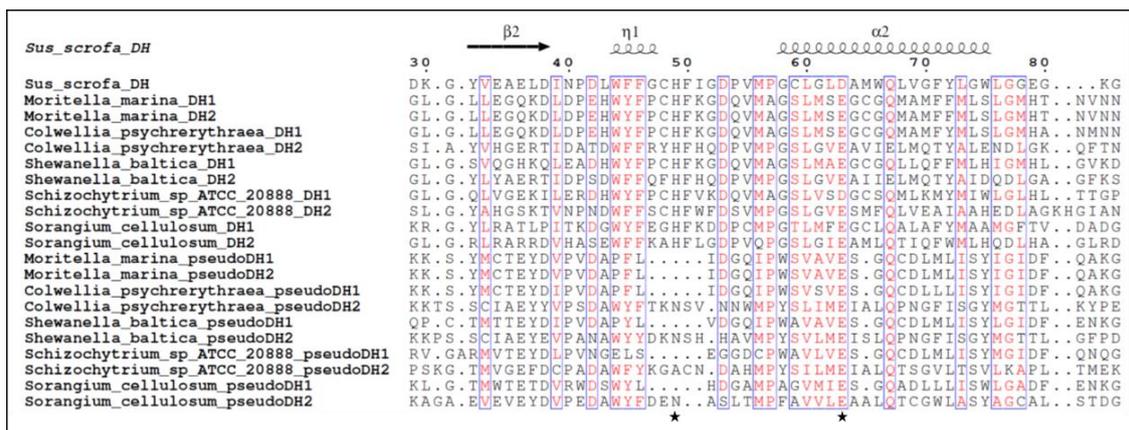
As there is no structure solved for any KS-KS di-domain the modelling to obtain structural predictions for PfaC had to be done using the KS domain from mammalian FAS. To summarize, only one KS domain has been represented, since both predictions are similar. As can be seen in figure R-23, the KS and the KS-like domains, represented in colour and black respectively, share a duplication of the typical thiolase fold, in the same way as the N-terminal fragment of PfaA.



**Figure R-23.** Structural model for the KS-KS di-domain of PfaC. The Pfa-C protein sequence was subjected to structural modelling using the Phyre2 online server. Domain prediction of InterPro online tool is represented at the bottom of the figure with a dotted rectangle indicating the sequence region used for the structural prediction. The KS domain cartoon representation generated with Pymol is shown and coloured in the left side of the figure where in the right the same structure was represented as a shadow indicating the duplication of the protein structural motif. Mammalian FAS (accession number: 2VZ8) was used as a structural template.

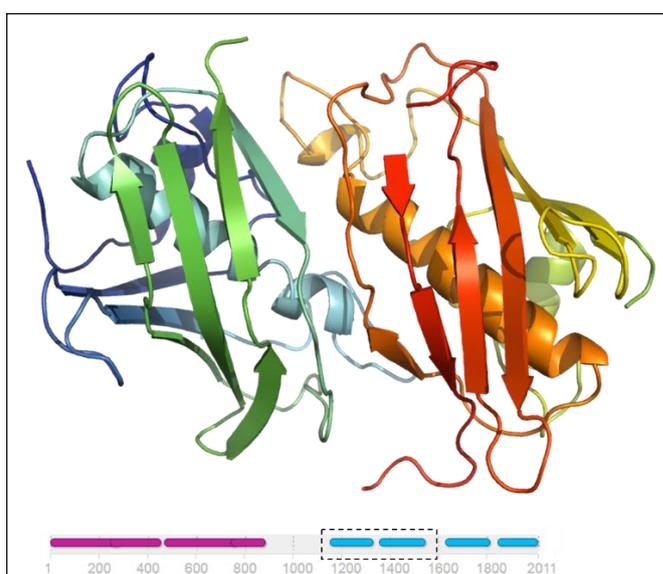
It should be mentioned here that there is a lot of confusion in the annotations found in the literature, since in most of the publications both KS-like domains have been annotated as domains with KS activity, when in fact only the first one has the typical residues of an actual KS domain. This must be taken into account avoid the general confusion and take it as a starting point to study these domains, whose structure and function within PUFA synthases is still unknown.

The C-terminal region of PfaC has been predicted to fold as four hot dog-like domains predicted to be similar to the ones found in some bacterial dehydrases (DH) domains. Homologs of this tandem dehydratases from the organism *Photobacterium profundum* were recently characterized as pairs of dehydratases (DH) and pseudo-dehydratases (DH') domains, being these last ones truncated versions lacking the catalytic histidine responsible of their enzymatic ability (Oyola-Robles et al. 2013b). Having this previous knowledge we have performed an alignment of the four DH/DH' domains to check which of them had conserved catalytic histidines and which did not. This way we were able to define how DH and DH' domains are organized in our system. As is shown in the alignment of figure R-24, half of the dehydratase domains lack the catalytic histidine but retain the glutamic acid that could also be implicated in the dehydratase reaction. The complete sequence alignment can be found at supplementary material, figure sm-7.



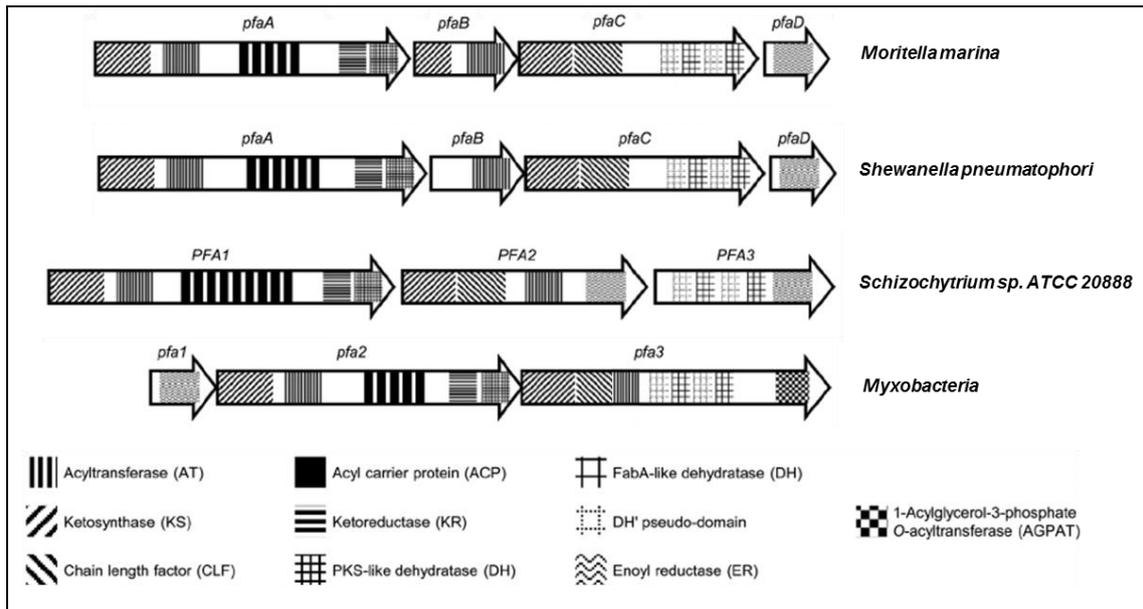
**Figure R-24.** Structure-based multiple sequence alignment of DH'-DH domain of PfaC. An alignment was made in order to show evolutionary conservation of amino acid residues in active centers of the DH domains and pseudo-domains of the PfaC protein with other representative omega-3 producers. Black stars at the bottom of the alignments indicate the conserved active residues: a histidine, that is absent in the pseudodomains and a aspartic/glutamic acid. Mammalian FAS (accession number: 2VZ8) was used as a model for secondary structure prediction showed above the alignment.

Structure prediction for the two firsts hot dog motifs of the C-terminal domain of PfaC, using the mammalian FAS (accession number: 2VZ8) as a template is shown in figure R-25. The DH'-DH model showed a structure very similar to the ones described for PKS or FAS multienzymes. In these cases, the DH motifs, instead of forming homodimers like the one predicted for the DH domain of PfaA, exist as individual hotdog domains followed by an additional C-terminal hotdog “pseudo-domain” that stabilizes double-hotdog structure. DH'-DH structure seems to have the similar organization as the one described for *Photobacterium profundum* where the layout of the domains followed a DH1'-DH1-DH2'-DH2 pattern. Structural prediction for the second pair of DH2'-DH2 is not shown due to their similarities with the first duplication. By examining the catalytic residues conserved within the structural prediction it can be observed that the active histidine of the DH domain is in a position very close to the glutamic acid of DH', so we do not rule out its intervention in the catalytic mechanism.



**Figure R-25.** Structural model for the DH'-DH didomain of PfaC. The C terminal part of PfaC was modeled using the Phyre2 online server. Domain prediction of InterPro online tool is represented at the bottom of the figure with a dotted rectangle indicating the sequence region used for the structural prediction. The DH'-DH-like domain cartoon representation was generated with Pymol software. Mammalian FAS (accession number: 2VZ8) was used as a structural template.

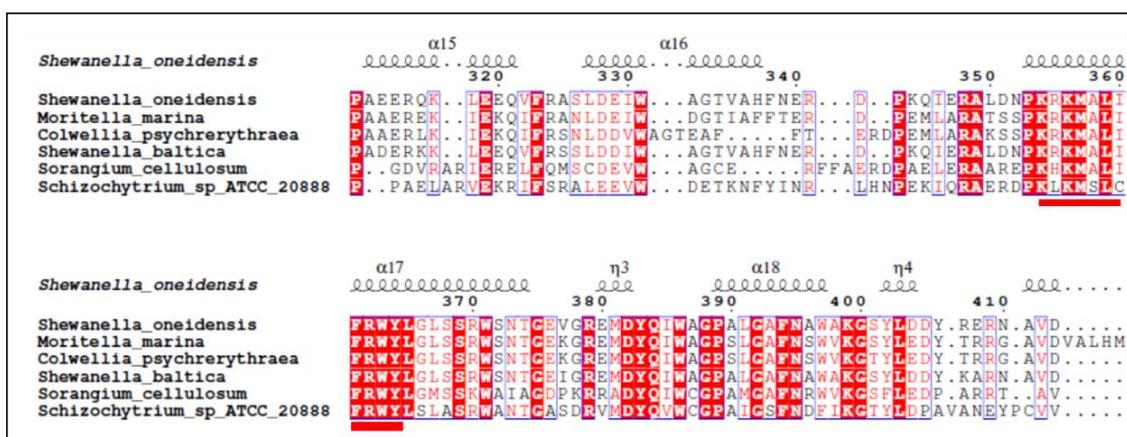
PfaC is a protein whose domains are highly conserved in organisms that produce PUFAs. These domains do not always have the same organization or coexist in the same polypeptide. In myxobacteria such as *S. cellulosum*, the gene that shows homology with PfaC is called Pfa3, in which in addition to the domains described for PfaC, it is found an additional catalytic domain (1-acylglycerol-3-phosphate o-acyltransferase) at its C-terminus. In *Schizochytrium* organisms the PfaC domains are distributed in two different polypeptides, PFA2, in which KS-CLF, an AT and an enoyl reductase (ER) domains are found, and PFA3, where are located two DH'-DH domains and another ER. Figure R-26 shows the genetic organization of the different domains present in the omega-3 synthases of the organisms used as comparative models in this study.



**Figure R-26.** Genetic organization of the *pfa*-like clusters of the omega-3-producer organisms included in previous genetic alignments *pfaA* is conserved in all species included in the analysis but domains of the other genes are located in different positions. Figure adapted from (Gemperlein, Rachid, Garcia, Wenzel, & Müller, 2014)

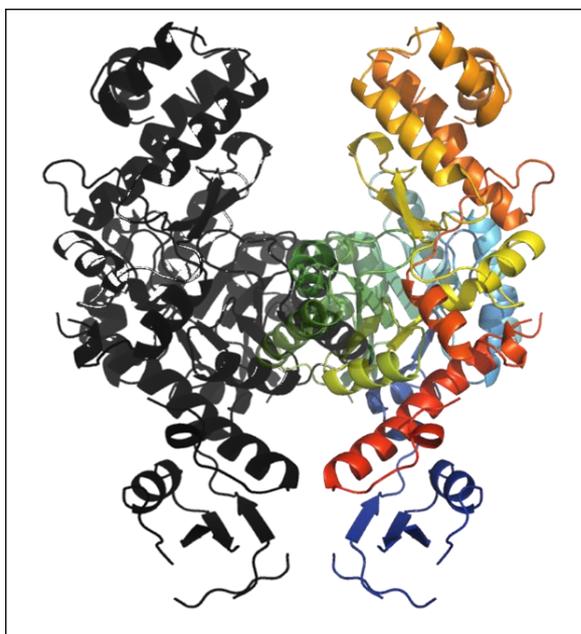
#### 4.2.1.4. PfaD is a enoyl reductase

PfaD is the only polypeptide that only contains a single functional domain of the entire *pfa* genetic cluster. For this reason, its analysis is quite simple. In the protein data bank an homologue whose structure have been solved can be found, the one from the marine bacterium *Shewanella oneidensis* (accession number: 4Z9R). Thus a sequence comparison has been made with using 4Z9R as template. As shown in figure R-27 the sequence conservation in the catalytic center is very high, as well as in the rest of the protein sequence that is not represented here but at supplementary material, figure sm-8. The catalytic motif typical of enoyl reductases is a Y-(X)<sub>6</sub>-K that is also conserved in all sequences analysed, but with an inversion (K-(X)<sub>7</sub>-Y) of the sequence and an extra amino acid inside the motif. The active site is highlighted with a red rectangle.



**Figure R-27.** Structure-based multiple sequence alignment of ER domain of PfaD. An alignment was made in order to show evolutionary conservation of amino acid residues in active centers of the ER protein PfaD with other representative omega-3 producers. A red rectangle bottom of the alignment indicate the conserved active residues the catalytic motif typical of enoyl reductases Y-(X)6-K. PfaD from *Shewanella oneidensis* (accession number: 4Z9R) was used as a model for secondary structure prediction showed above the alignment.

A structural model was made using the structure *Shewanella oneidensis* ER (accession code: 4Z9R) as template. This ER protein from *S. oneidensis* is a dimer whose active center is situated just in the middle of both molecular subunits. Structure prediction for PfaD is represented in figure R-28. Its second monomer was represented nearby the modelled structure as a dark shadow in the position where its homologous has the duplication. With this oligomeric configuration, the active center would be protected from the polar contacts of the solvent.

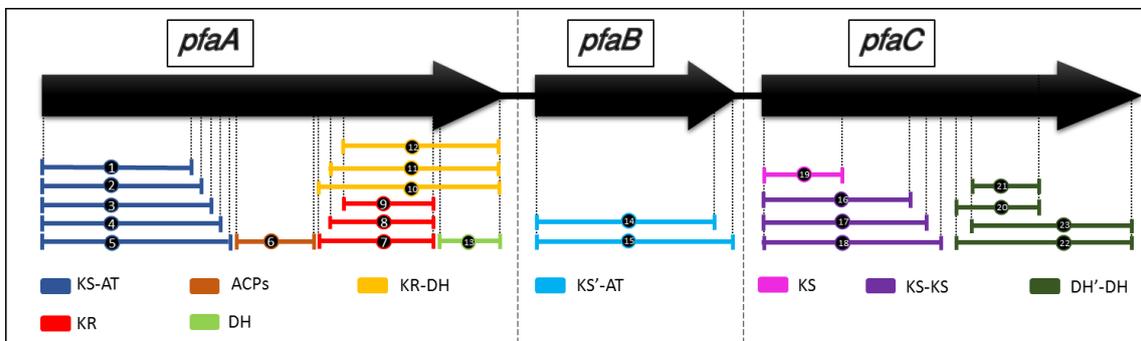


**Figure R-28.** Structural model for the enoyl reductase domain of PfaD. The PfaD protein sequence was modeled using the Phyre2 online server. The ER domain cartoon representation was generated with Pymol. The color model is represented in the right side of the figure where in the left the same structure was represented as a shadow indicating the possible duplication of the protein structural motif showed in other solved structures. PfaD from *Shewanella oneidensis* (accession number: 4Z9R) was used as the structural template.

#### 4.2.2. *pfa* constructions design

The first step to characterize our *pfa* system was to decide how many constructs of what domains were going to be studied to have a good snapshot of the protein complex and perform the crystallization and the biochemical experiments. In the previous section, we have analyzed the *pfa* genetic cluster organization that is naturally divided in 5 different polypeptides (PfaA, PfaB, PfaC, PfaD, PfaE). The first approach was to make constructions of the full-length proteins to try to crystallize them. This is not a simple task since in addition to make the larger constructions, *pfaA* and PfaC, they have to present the stability and solubility necessary to be purified. Aside from the production of full-length proteins, we wanted to clone their different domains as they should be able to act as individual biochemical entities. As explained in the previous section, a sequence comparison with other omega-3 producers and some structural models were used to define the different domains that form the Pfa polypeptides. To decide the borders that define the different constructions, a logical criterion has been followed. To define the beginning and the end of a domain the structural prediction obtained with phyre2 and Jpred have been taken into account. Only the areas whose prediction identify them as unstructured motifs were selected as zones to design the oligonucleotides. Following this methodology we were able to predict the positions of the individual functional domains and their unstructured inter domains.

We thought that in this way, with the functional domains cloned as individual entities, the interpretation of the further biochemical experiments would be easier and the chances of success in crystallization trials would increase. Figure R-29 shows a schematic representation of the different constructions where the wild type polypeptides PfaA, PfaC and PfaD were not included in order to make it simpler.



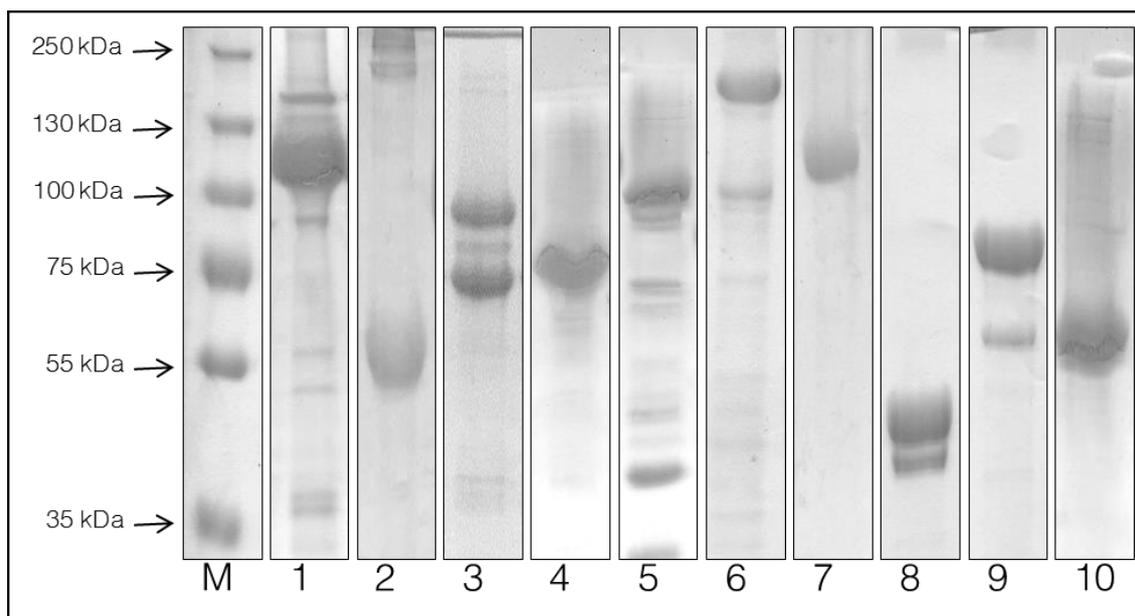
**Figure R-29.** Scheme of the genetic constructions of the *Pfa* genetic cluster from *Moritella marina* made in this work. Full length proteins have not been included in the scheme to facilitate the visualization. The numbering of the constructions is the same as that of the table R-2. A color code and a legend have been used to facilitate the identification of the functional domains included in each construction.

Table R-2 summarizes the constructions designed which are arranged according to the numbering used in the previous figure R-29. Constructions highlighted in green were the ones that yielded soluble proteins. In addition to these constructions of individual domains those which comprise entire wild type proteins were also annotated in the table.

**Table R-2.** List of constructions made in this work. A list has been detailed in which the amino acids included in each of the constructions are specified. The same numbering of the figure R-29 has been used to facilitate the identification of the constructions.

Number	Name	Amino acid range	Number	Name	Amino acid range
1	KSAT1	M1-A934	14	PfaB1	M1-S865
2	KSAT2	M1-G982	<b>15</b>	<b>PfaB2</b>	<b>M1-K883</b>
<b>3</b>	<b>KSAT3</b>	<b>M1-G1150</b>	16	KS-KS1	M1-A876
4	KSAT4	M1-A1197	17	KS-KS2	M1-A925
5	KSAT5	M1-A1235	<b>18</b>	<b>KS-KS3</b>	<b>M1-A1006</b>
<b>6</b>	<b>5ACP</b>	<b>A1258-V1748</b>	19	KS 1M	M1-S418
7	KR1	S1746-Q2291	<b>20</b>	<b>DH1M1</b>	<b>V1030-L1555</b>
<b>8</b>	<b>KR2</b>	<b>A1809-Q2291</b>	21	DH1M2	Y1149-L1555
9	KR3	V1863-Q2291	<b>22</b>	<b>DH2M1</b>	<b>V1030-A2011</b>
10	KRDH1	S1746-S2652	23	DH2M2	Y1149-A2011
<b>11</b>	<b>KRDH2</b>	<b>A1809-S2652</b>	24	PfaA	Wild type
12	KRDH3	V1863-S2652	<b>25</b>	<b>PfaC</b>	<b>Wild type</b>
13	DH ( <i>pfaA</i> )	S2356-S2652	<b>26</b>	<b>PfaD</b>	<b>Wild type</b>

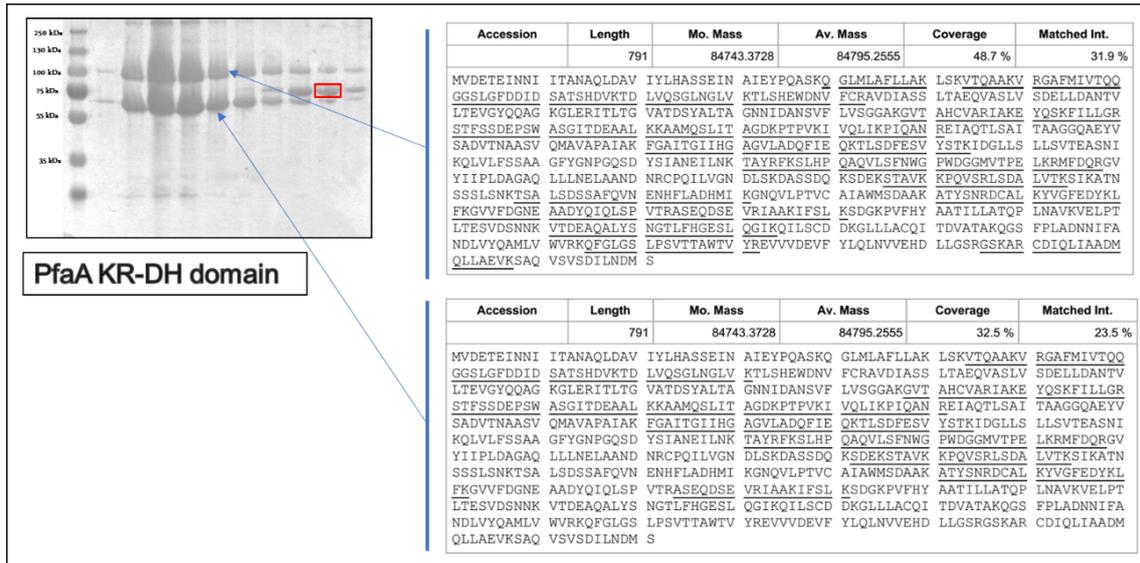
Twenty-six constructs were made, of which ten produced soluble proteins that we were able to purify. All of them have been purified with a histidine tag using nickel affinity columns since they have been cloned using the overexpression vector pet29c. The exceptions is PfaE whose cloning has been performed using the pet23a vector that does not codify for a histidine tag. PfaE overexpression was done in further biochemical experiments with the purpose of *in vivo* activate the ACPs so it is not included in this analysis as we did not purify it. Figure R-30 shows an acrylamide gel with samples of all the constructions made that yielded enough amount of protein to conduct further crystallization and biochemical experiments. As a rule, we used a tris buffer with a pH of 7.5 and a sodium chloride concentration of 150 mM. The temperature of expression has always been 18 °C, with an overexpression time of 20 hours and 1 mM of IPTG, with the exception of the KSAT domain of PfaA and PfaB that were expressed at 15 °C with 0.1mM of IPTG. The detailed purification procedure is described in the Materials and Methods section. We have observed a general tendency to decrease the efficiency of purifications when raising temperature levels. This phenomenon could be a consequence of the evolutionary acclimatization of these proteins to the cold marine environments.



**Figure R-30.** SDS-PAGE gel for testing the quality of the different PUFA synthase domains purified for further crystallization and biochemical experiments. All proteins were purified by nickel affinity chromatography with only one purification step. Line M: Marker; 1: KSAT (clon 3); 2: 5ACP (clon 5); 3: KR-DH (clon 11); 4: KR (clon 8); 5: PfaB (clon 15); 6: PfaC (clon 25); 7: KS-KS' (clon 18); 8: DH1M (clon 20); 9: DH2M (clon 22); 10: PfaD (clon 26).

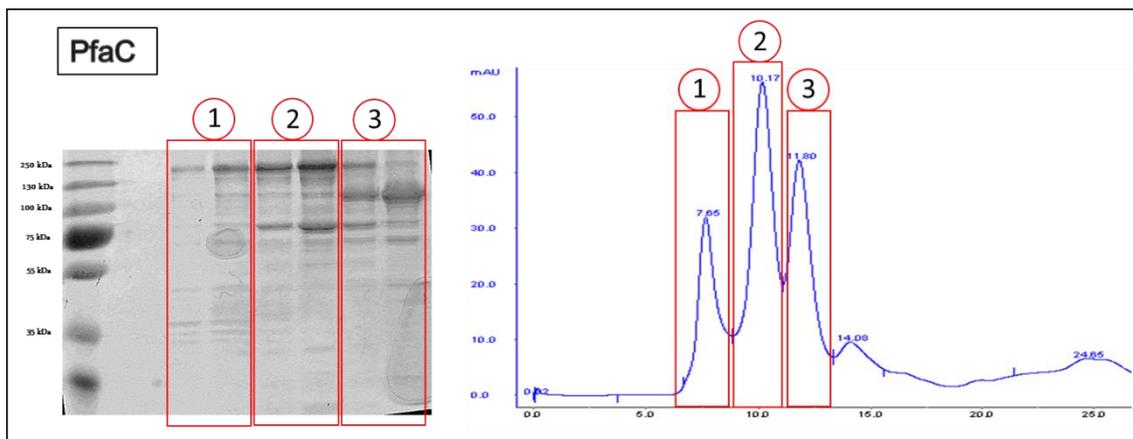
PfaB (construction number 15), PfaC and PfaD were the full-length constructions that yielded soluble protein when were overexpressed. PfaB and PfaD are stable proteins in our working buffer, but PfaC show certain degree of instability and get degraded in less than a week at 4°C. Unfortunately, we were not able to obtain an overexpression of the PfaA full length protein. All protein domains but KR-DH and PfaC were purified as single bands even though some of them have other minor impurities. Two single bands were always obtained when KR-DH module of PfaA was purified. Impure protein was subjected to gel filtration to try to separate both bands with no satisfactory results as is shown in an SDS-PAGE gel in figure R-30. In order to identify both bands obtained when overexpress the KR-DH domain they were subjected to MALDI-TOFF analysis. The spectra have been analyzed with the MASCOT search engine using the sequence of *M. marina* as reference. Right panel of figure R-31 shows the peptides identified from each of the bands of the gel. The upper band was identified with the reference protein sequence (KR-DH; PfaA). A change in the peptides found for the lower band was observed, especially in the C-terminal fragment. With this information we can conclude that this band is a product of degradation of the complete protein.

The red rectangle of the SDS-PAGE gel of the figure R-31 (on the right) is signaling the most common contaminant found in our purifications. We subjected this band to a peptide-mass fingerprinting analysis to try to identify the protein. The protein was identified as ArnA, a very common contaminant from His-tag purifications *in E.coli*, a bifunctional enzyme involved in the modification of lipid A phosphates with aminoarabinose (Andersen, Leksa, and Schwartz 2013). This biochemical reaction in principle does not interfere with the experiments that we will perform in the next sections.



**Figure R-31.** Gel filtration and MALDI-TOFF analysis of the KR-DH protein. KR-DH protein was purified using Nickel affinity column and subjected to gel filtration to try to separate a contaminant band. As is shown in the SDS-PAGE gel the two band elluted together. Both bands were subjected to MALDI-TOFF analysis and the peptide matches are shown in the figure.

In the case of PfaC, two extra bands were obtained that could be degradation fragments of the full-length protein. In this case we only identify with MALDI-TOFF the upper band corresponding to the whole protein since we were able to get rid of the contaminating bands, at least partially. Figure R-32 shows an SDS gel containing the fractions of the gel filtration experiment in which we are partially able to separate the larger band. Gel filtration helped in this case to get rid of the degradation fragments but not in all fractions collected.



**Figure R-32.** Gel filtration of the PfaC protein. PfaC protein was purified using Nickel affinity column and subjected to gel filtration to try to separate the contaminats associated. As is shown in the SDS-PAGE gel (left) and the gel filtration chromatograpy (right) the protein elluted in three different peaks. Fraction 1 contained protein with a better purity tha the original sample.

*pfaE* codes for a PPTase that is needed for the activation of the acyl carrier protein. As was commented in the previous section PfaE was described as a protein that is not actually necessary to produce PUFAs as its catalytic reaction can be performed by other homologous proteins (Sugihara, Orikasa, and Okuyama 2010). Normally all the constructs planned to be purified were done carrying a histidine tail using a pET29c vector that confers Km resistance but in the case of PfaE the vector of choice was pET3a that carry an Amp resistance cassette. The reason for that was that PfaE was only going to be used to activate the ACPs with phosphopantetheinyl group. The Amp cassette let us to co-express the ACPs together with PfaE and be able to purify the acyl carrier proteins previously activated *in vivo* and ready for the further biochemical experiments.

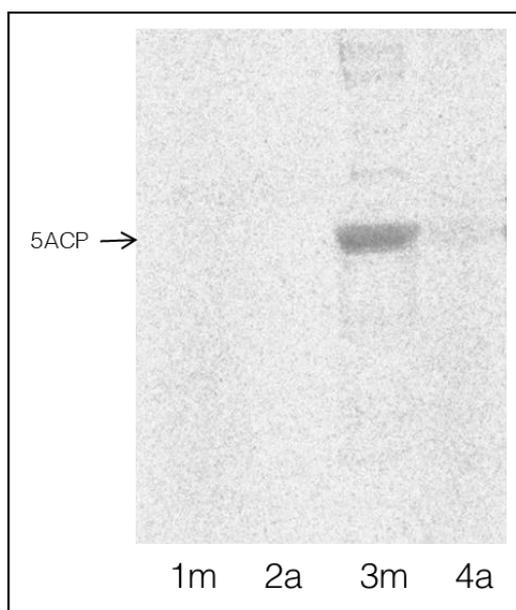
### 4.2.3. Biochemical analysis

The regular synthesis of 16-18 carbon fatty acids in mammals and fungi is performed by FAS type I proteins, biological iterative systems that were studied in depth for many years. The proposed mechanism for fatty acid synthesis was supported by many experiments that mainly prove the individual enzymatic activities of all domains present in the complex. PUFA biosynthesis is presumably also an iterative process through which a series of individual enzymatic activities performed by different protein domains led to the building of an unsaturated fatty acid in a very similar way as the FAS type I systems. When structural comparisons are made with bioinformatic tools it is clear that PUFA and FAS synthases share a common domain architecture and organization. This supports the idea that both systems should carry on very similar biochemical pathways. Unfortunately, not many efforts have been made in the biochemistry field to try to prove the similarities between both systems and explain why the molecular products of both synthesis are quite different in terms of length and unsaturation pattern. Thus, we have performed biochemical experiments that involve the protein domains present in our *pfa* cluster. These experiments will give us information of how the PUFA synthesis is working in the cell. Our main approach was to try to prove the ability of each individual domain or a combination of them to bind and process malonyl-CoA and acetyl-CoA substrates *in vitro*. For that purpose, we designed experiments that use the detection of covalently bound protein intermediates by radio-SDS-PAGE to investigate the enzymatic activity of PUFA synthase components.

#### 4.2.3.1. ACPs are self-acylated with malonylCoA but not acetylCoA

PUFA synthases uses ACPs as covalent attachments for fatty acid synthesis in an iterative reaction where ACPs hold the growing fatty acid chain to transfer it from one module to the next one. The synthesis proceeds with reiterative condensation cycles where the acyl chain is covalently attached to the ACPs, growing by two-carbon units with each round. The fatty acid is bound to the free sulfhydryl group of the 4'-phosphopantetheine (ppt) prosthetic group of the ACP that is covalently linked to a conserved serine residue of their holo- form. This ppt residue is actually enzymatically added to the protein in a posttranslational modification catalyzed by a PPTase (PfaE). The ppt functions as a long flexible arm that carries the growing chain and delivers it to the different domains of the complex responsible for chain extension and modification. In this experiment ACPs were co-purified with PfaE in order to obtain directly the holo-form needed for further experiments. In this way the conversion of the inactive apo- form of the ACPs to the active holo- form is performed *in vivo*.

Normally PUFA synthesis starts when an AT incorporates the malonyl onto the ACPs ppt groups. In that way, when the ACPs are charged, the KS domain can start the claisen condensation cycles. It was described that at least some ACPs from type II PKS are able to perform self-malonylation reactions, being charged without the need of an AT domain (Arthur et al. 2005). Knowing that, we wanted to find out if the tandem ACPs of the *M. marina* Pfa system were able to do the same and also to investigate their substrate preference. We checked the ability of our holo- and apo- ACPs to covalently bind <sup>14</sup>C-radiolabeled malonyl-CoA and acetyl-CoA. The results are shown in figure R-33.

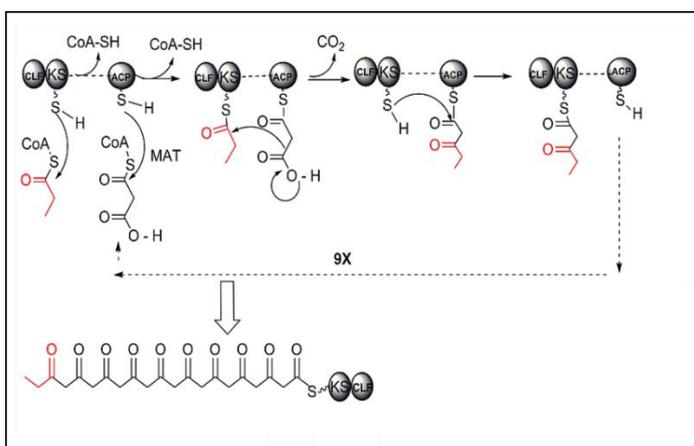


**Figure R-33.** Binding enzyme assays of 5ACP domains. Protein was incubated with radiolabeled substrates, [<sup>14</sup>C]-malonyl-CoA and [<sup>14</sup>C]-acetyl-CoA, and analyzed via radio-SDS-PAGE. Lane 1m: apo-5ACP+Malonyl-CoA; 2a: apo5ACP+Acetyl-CoA; 3m: holo-5ACP+Malonyl-CoA; 4a: holo-5ACP+Acetyl-CoA. Protein mapping positions are indicated on the left of the panel.

In this experiment, the ability ACP holo- and apo- forms to bind acetyl and malonyl-CoA substrates was tested. Apart from including the holo-ACP in the experiment that was expressed in the presence of PfaE to be transformed in the active form, apo-ACP was also purified in order to prove that the PPT arm was strictly necessary for any type of acylation. As is shown in the gel of figure R-33, only holo-ACP is active and able to bind malonyl. It is also remarkable its strong preference for malonyl vs. acetyl.

It is generally accepted that in the first step of the fatty acid synthesis, a molecule of acetyl, normally in the form of acetyl-CoA, is covalently attached to the cysteine of the KS to form acetyl-KS (Staunton and Weissman 2001). The priming with acetyl is involved in the first reaction and its condensed with malonyl-ACP to form acetoacetyl-ACP (Campbell and Cronan 2001) in the first condensation event. To be able to do this condensation the ACPs should be previously charged with extender blocks (malonyl units). In FAS systems this activations of the ACPs is always performed by an AT domain that collaborates with the ACP and the KS. Although it seems that the AT modules are in charge of loading the extender units to the ACPs, it has been shown that the ACP domains can also exhibit intrinsic acyltransferase activity in some PKS systems. This activity is completely necessary in some type II PKS in which the AT module is missing (Arthur et al. 2005; Hitchman et al. 1998). Figure R-34 shows a representation of the typical biochemical collaboration between ACP, KS-CLF and AT domains in PKS systems where the AT domain is not always needed for the acylation of the ACPs.

This experiment points out that in our system, as it was also described for FAS, the acetyl primer cannot be incorporated to the ACP without the help of other protein domain. On the contrary, the 5ACPs in tandem of our Pfa system were able to incorporate malonyl units by themselves, with no other enzymatic help, demonstrating the malonyl transferase ability of the ACPs in a similar manner as was demonstrated for the ACPs of some type II PKS systems. The acetyl needed to start the synthesis has to be present in the KS domains of PfaA in order to condensate the extender units previously charged in the ACPs. As a complementary result to this experiment, we have also tested the auto malonylation activity of the homologous tandem ACP protein from *C. psychrerythraea*. This protein has six tandem repetitions of the ACP motif instead of five but its sequence is similar to the one from *M. marina*. The results obtained were shown in supplementary material figure sm-9.



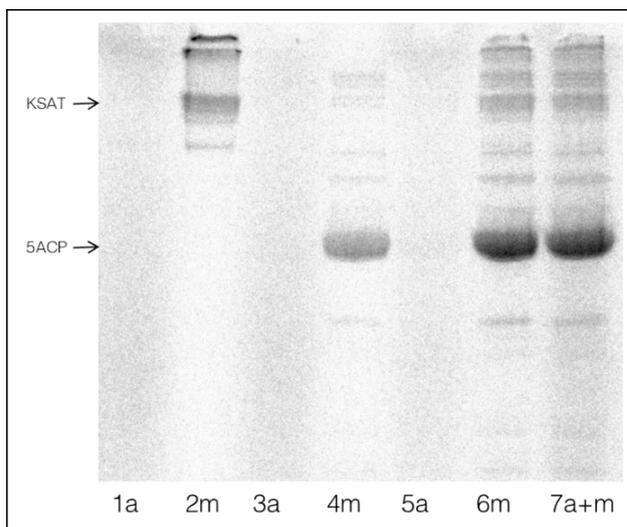
**Figure R-34.** ACP, KS-CLF and AT domain collaboration. Typical mechanism of iterative PKS type II for loading of extender units to the ACP, using an AT helper domain. The condensation of the extender units is carried out by the KS-CLF di-domain.

#### 4.2.3.2. PfaA KS-AT promotes malonyl-CoA binding to ACP

The minimal module required to produce a regular condensation contains three domains that catalyze one cycle of chain extension: A ketosynthase (KS), an acyltransferase (AT) and an acyl carrier protein (ACP) as it was shown in figure R-34. The acyl group usually known as starter unit is attached to a cysteine thiol of the first enzyme, the KS. The chain extender unit, usually malonate, is selected by the AT and bound to a thiol residue of the ACPs ppt arms, ready to contact with the KS module that performs the condensation. In the first module of the *pfa* cluster (PfaA) there are all these elements that are necessary to perform this first condensation. For that reason, we suspect that probably the first steps of the synthesis, must occur within the PfaA module. With the following experiment we wanted to obtain evidence that support this assumption.

We have seen that the ACPs of our system are able to bind malonyl by themselves. But even if the ACP is able to get activated by self-malonylation, as we have demonstrated here, all studies of fatty acid synthesis and polyketide synthesis agree that the help of an acyltransferase is needed to fully activate the ACP with malonyl or other extender units for the proper function of the system. The role of this protein (AT) is to transfer the building blocks from the respective coenzyme A pools onto the active thiol residues of the ACPs, selecting then the specific extender units. We wanted to prove the role of the KSAT domain and its biochemical function with its cognate ACPs. For that reason, we designed an experiment where we mixed the purified KSAT and the ACPs domains of PfaA and mixed them together with radiolabeled substrates to test if the

KSAT domain was able to actively select the extender blocks and transfer them to the ACPs. Results of this experiment are shown in figure R-35.



**Figure R-35.** Binding enzyme assays of KSAT and 5ACP domains. Protein were incubated with radiolabeled substrates, [14C]-malonyl-CoA and [14C]-acetyl-CoA, and analyzed via radio-SDS-PAGE.. Line 1: KSAT+Acetyl-CoA; 2: KSAT+Malonyl-CoA; 3: 5ACP+Acetyl-CoA; 4: 5ACP+Malonyl-CoA; 5: 5ACP+KSAT+Acetyl-CoA; 6: 5ACP+KSAT+Malonyl-CoA; 7: 5ACP+KSAT+Acetyl-CoA+Malonyl-CoA. Protein predicted mapping positions in the gel are indicated on the left of the panel.

This experiment shows that the KSAT domain of PfaA is able to selectively bind malonyl but not acetyl, as a band with the size corresponding to this domain can be detected in the gel. In the same manner, the gel shows how ACPs are able to bind malonyl by themselves, with no enzymatic help from KSAT protein, as we previously shown in the last experiment. When KSAT is mixed with ACPs and malonyl-CoA, a strong signal is detected, demonstrating that the KSAT domain has acyl transferase activity with a strong preference for malonyl. This selectivity is suggesting that, the acyl transferase domain of KSAT is active in the presence of malonyl and is able to charge the ACPs to a level that the ACPs cannot achieve by themselves. Although apparently AT should be the responsible of the transference reaction, more experiments are needed to tell apart the enzymatic activities of the two domains present in the KSAT construction.

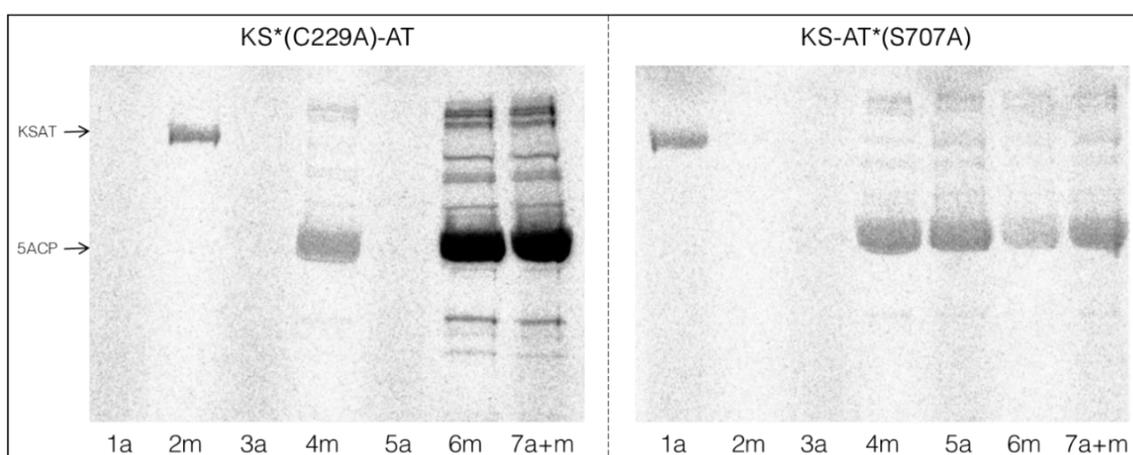
The three domains used here (KS, AT and ACPs) could be able to perform condensation rounds with no help from other proteins. Knowing this, it is important to keep in mind that in this experiment is impossible to identify the size of the radiolabeled molecules that are covalently bound to ACP. It means that we are not able to know if the first condensation rounds were performed or not. This and an important question that will be answered with the next experiment using mutants of the KSAT construction.

The mechanism of loading extender blocks onto the ACPs performed by KSAT demonstrated here highlight that PfaA is able to load the firsts malonyl extender units that the system will use to start the synthesis. Besides having the three domains included in this experiment, PfaA has two extra domains that are predicted to be keto reductases (KR) and dehydratases (DH). The activities of them allow PfaA not only to have its ACPs charged with extender blocks by itself, but to modify them performing reductions that could prepare this extender blocks for the next steps in the synthesis. If this assumption is true, it would have a lot of implications in terms of understanding the overall pathway for DHA synthesis when more biochemical experiments are done with the rest of protein domains.

#### 4.2.3.3. PfaA KS-AT S707A promotes acetyl-CoA binding to ACP

PfaA KS and AT domains were cloned and expressed as a single polypeptides because in all Pfa-like FAS systems these domains are usually expressed as single polypeptides and as a consequence of this they perform their enzymatic reactions in a very coordinated way. After the AT charge the ACPs with the proper extender units, the KS is able to condensate these blocks to make the chain grow. There are also a few KS-AT structures published that show how the KS domains form dimers and the AT domains are spatially organized in very close conformation that they should be coordinated with each other and with the ACPs.

Due to the inability to know which of the two domains (KS or AT) present in the construct was responsible for the effect observed in the previous experiment we decided to construct two new plasmids carrying alanine substitutions in active centers, substituting the cysteine of the KS domain (C229A), and the serine of the AT domain (S707A), respectively. These mutants were used in an experiment with exactly the same conditions as the last one and the results are shown in figure R-36.



**Figure R-36.** Binding enzyme assays of KSAT mutants and 5ACP domain. This assay is a replica of the last radio-SDS-PAGE gel but using mutants for the active centers of both KS and AT domains separately. Right and left panels have the same line organization and show AT (S707A) and KS (C229A) mutants results respectively. Line 1: KSAT+Acetyl-CoA; 2: KSAT+Malonyl-CoA; 3: 5ACP+Acetyl-CoA; 4: 5ACP+Malonyl-CoA; 5: 5ACP+KSAT+Acetyl-CoA; 6: 5ACP+KSAT+Malonyl-CoA; 7: 5ACP+KSAT+Acetyl-CoA+Malonyl-CoA. Protein mapping positions are indicated on the left of the panel.

In this experiment, the KS mutant (C229A) showed a similar pattern as the wild type KSAT protein, being able to bind malonyl and load it to the ACPs as is shown in the left panel of figure R-36. As the KS was completely inactive in this mutant, this indicates that the AT domain was performing this catalysis. When the substrate is incubated only with the KSAT domain, the malonyl is bound to the active serine of the AT domain. Moreover, when is incubated with ACPs, the substrates are efficiently transferred to the thiol group of the ppt arms. The increasing of ACP band intensity observed here when were incubated with KSAT (C229A) domain is consistent with the one shown in figure R-35 with the wild type protein. In this particular case is not possible any chain extension due to the mutation introduced in the cysteine of the KS domain. If KS were doing some condensation rounds we would expect a decreasing in the intensity of the ACPs band

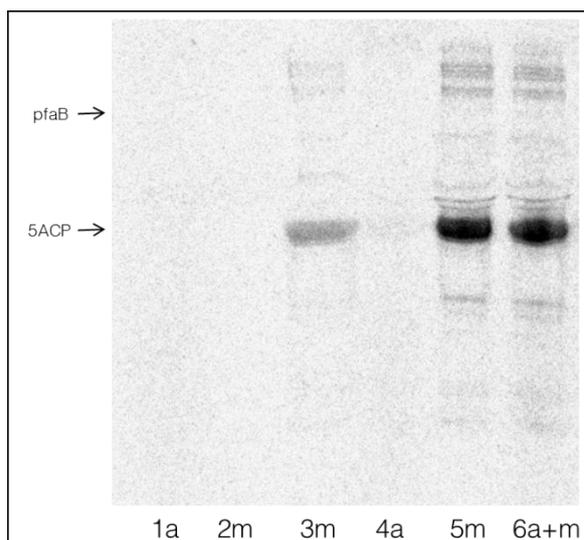
with respect to the previous experiment. This could mean that there is no need of KS enzymatic contribution for replicate exactly the same result, so that the KS is apparently not condensing. More experiments that quantify the amount of substrate bound are needed to demonstrate this hypothesis.

On the other hand, as shown in the right panel of figure R-36, when the KSAT mutant (S707A) is incubated with their substrates, there is a change in the substrate specificity to the use of acetyl-CoA. Moreover, it is observed that AT (S707A) is able to charge the ACPs with acetyl when incubated together. It has to be taken into account that it was not observed binding of acetyl to the ACPs before until this experiment so there is a qualitative change. The protein also lost its ability to charge the ACPs with malonyl, at least with the strength shown before. As the AT mutant has no active serine where the selected substrates should be attached, it is difficult to speculate about where the acetyl could be bound. It is logical to think that the substrate has to be now attached to the active cysteine of the KS domain, that should be completely active, but this particular reaction apparently happens only when the AT domain is inactive. To accept that explanation, it is necessary that the active AT domain has some unknown activity that was preventing the binding of acetyl to the active cysteine of the KS domain. The analysis of both AT and KS domain separately could shed light to this pretty striking effect. Other AT domains have shown before to be able to hydrolyze the acetyl group from an acetyl-enzyme covalent complex (Liew et al. 2012). Is it possible that in the presence of the active AT domain, any acyl group that spontaneously bind to the KS domain is rapidly hydrolyzed making impossible to detect them bound to the KS domains when the AT is active. It was also extensively described the molecular recognition between ACP and its cognate KS (Kapur et al. 2010). This fact, together with the intrinsic acyl transferase activity of the ACP domains previously demonstrated, leads to think that the ACPs, that should be in tight contact with the KS domain (that is now charged with acetyl), will be able to transfer the acetyl prime units from him to the same level as they could take malonyl from the media.

#### **4.2.3.4. PfaB is able to charge ACPs with malonyl-CoA**

PfaB is the second module of the system and it was described to be the key enzyme that determines the final product in EPA/DHA biosynthesis (Orikasa et al. 2009b). Even if it does not have the typical catalytic cysteine in its structural prediction made with Phyre2 server (previously shown in figure R-20) suggested that PfaB N-terminal domain poses a thiolase fold that resembles a KS domain. This KS-like domain is present usually in DHA-producers like *M. marina* but is absent in EPA-producers like *Shewanella baltica*. As this KS-like domain does not have the canonical cysteine in the active center as the rest of known KS in nature its biological function is not clear.

An AT-like domain is always present in the C-terminal position of PfaB proteins and always has the typical serine in the active center as was shown in the alignment of figure R-19. This is suggesting that PfaB should be playing an AT role in the protein complex. To try to understand what could be the function of this protein in the PUFA synthase we subjected the protein to the same assay as we have done with PfaA domains and the results are shown in figure R-37.



**Figure R-37.** Binding enzyme assays of PfaB and 5ACP domains with radiolabeled substrates. Lane 1a: PfaB+Acetyl-CoA; 2m: PfaB+Malonyl-CoA; 3m: 5ACP+Malonyl-CoA; 4a: 5ACP+PfaB +Acetyl-CoA; 5m: 5ACP+PfaB+Malonyl-CoA; 6a+m: 5ACP+KSAT+Acetyl-CoA+Malonyl-CoA. Protein mapping positions are indicated on the left of the panel.

When PfaB was incubated with both substrates, malonyl and acetyl (lines 1a and 2m of figure R-37), as in the previous experiments, no radioactive binding was detected. When the protein was incubated with substrates but in presence of ACPs, it was able to load malonyl onto ACPs but not acetyl. Basically, what we observed here is the same behavior as the AT domain of PfaA, in the sense that is acting like a regular AT. The main difference is that PfaB is only able to bind its substrate when is in contact with ACPs as we did not observe binding when we incubated PfaB alone.

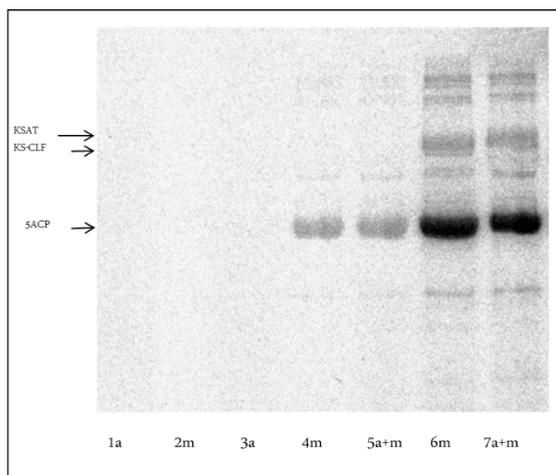
It is difficult to provide an explanation for this results without further experiments. It is known that AT domains in nature can be tolerant to a certain extent to new building blocks (Dunn and Khosla, 2013). Then is not surprising that both PfaA KSAT and PfaB AT domains tested in the experiments behave alike when are exposed to malonyl-CoA. An *in vitro* experiment like this can prove that the protein is active against one particular molecule but more different substrates should be tested to find other possible specificities. We believe that as the AT domain of PfaA is located in the same polypeptide as the ACPs, it is more likely that it is responsible for loading the extender modules to the ACPs. Due to its implications in the determination of the final length of the product (DHA/EPA) we think that PfaB should act in the last steps of the synthesis selecting the last extender units.

#### 4.2.3.5. PfaC KS-CLF domain does not bind malonyl-CoA nor acetyl-CoA

PfaC is the third module, downstream PfaA and PfaB within the *pfa* cluster of *M. marina*. Its protein architecture is composed by two KS domains in the N-terminal domain and four DH domains at the C-terminal domain. We have previously analyze their domain architecture using bioinformatics tools and we have predicted a pseudo-DH-DH repetition very similar to the one found in homologous of *Photobacterium profundum* previously described in the Dr. Abel Baerga's laboratory. We concluded that as in the case of *Photobacterium*, only two of these four domains of PfaC have actual dehydratase activity due to the lack of the active histidine necessary for catalysis of two of them (Oyola-Robles et al. 2013a). The KS repetition of the PfaC N-terminal domain is also present in PKS type II domains previously characterized. While the first KS domain

possess the canonical KS sequence, the second one lacks the cysteine in its active center, being substituted for a glutamic acid. In PKS, this second KS, the one that lacks the active cysteine, is known as chain length factor (CLF) because it has been described to be necessary for dictating polyketide chain length. It was also shown its ability to catalyze chain elongation in iterative PKS together with ACPs and with no help from other protein domains (Bao, Wendt-Pienkowski, and Hutchinson 1998; Carreras and Khosla 1998; Matharu et al. 1998). The CLF was also proposed to be a factor required for polyketide chain initiation in this group of PKS because it has decarboxylase activity towards malonyl-ACP and is able to prime the ACPs with starter units (Bisang et al. 1999).

Because of the apparent importance of KS-CLF domains in the PKS synthesis process, it is crucial to understand their role in PUFA synthases. KS-CLF, as was mentioned before, is able to prime the ACP in some PKS type II through a decarboxylation process that uses malonyl as substrate, making the carbon chain grow through a series of condensation rounds. Some authors also speculate that KS-CLF domains are involved only in the last steps of synthesis in some PKS complexes where the KS-CLF has preference for being primed with bigger units than acetate (Tang, Tsai, and Khosla 2003b). In order to gain some information about the substrate specificity of KS-CLF domains of PUFA synthases, as could be involved in starter units priming and chain extension, we decided to do the same substrate binding experiment as with the KS-AT and ACPs domains and the resulting gel is shown in figure R-38.



**Figure R-38.** Binding enzyme assays of KSAT, CLF and 5ACP domains with radiolabeled substrates. Line 1a: KS-CLF+Acetyl-CoA; 2m: KS-CLF+Malonyl-CoA; 3a: 5ACP+KS-CLF+Acetyl-CoA; 4m: 5ACP+KS-CLF+Malonyl-CoA; 5a+m: 5ACP+KS-CLF+Acetyl-CoA+Malonyl-CoA; 6m: 5ACP+KSAT+Malonyl-CoA; 7a+m: 5ACP+KSAT+KS-CLF+Acetyl-CoA+Malonyl-CoA. Protein mapping positions are indicated on the left of the panel.

PfaC KS-CLF did not show ability to bind malonyl nor acetyl when it was incubated with both substrates, being them attached to coenzyme A or even forming covalent complexes with ACPs or KS-AT domains as was demonstrated before. This result points out that the CLF domains of PUFA synthases probably do not have the same function as some CLF from iterative PKS that are specialized in priming the system with malonyl units and/or performing the first rounds of condensation. The inability to use malonyl-CoA or acetyl-CoA as substrates suggest that perhaps more elongated and/or reduced acyl moieties are the proper KS-CLF substrates. Complementary experiments were performed in which we tried to *in vitro* generate the KS-CLF substrate by addition of the previously purified domains (PfaB, KR-DH, DH2M) and with NADPH but the result was similar to the one showed in figure R-38. Results of these experiments can be consulted in supplementary material, figure sm-10.

#### 4.2.4. Structural analysis of PfaC KS-CLF domain

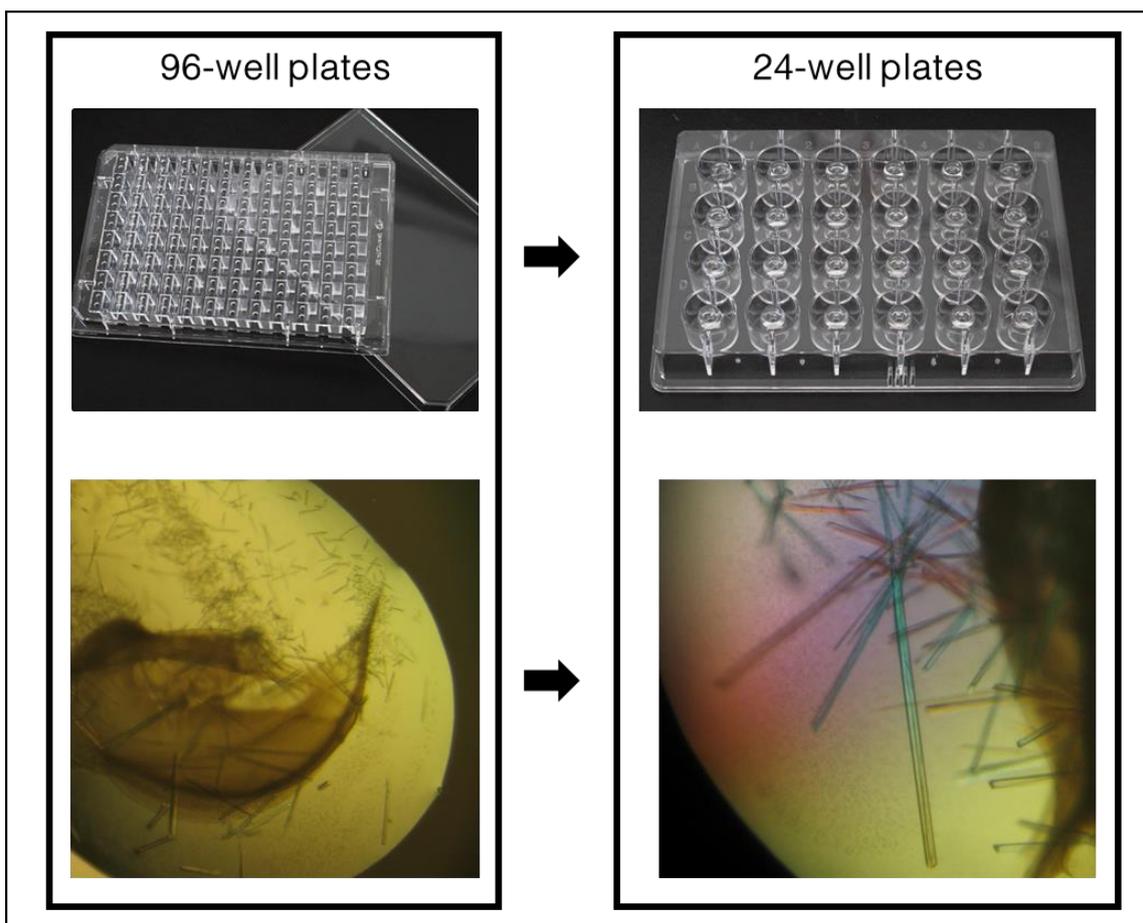
KS-CLF is an interesting domain for us due to the total lack of knowledge we have about their role in the PUFAs production, especially after knowing that it does not react with the most common substrates known for FA synthesis (results shown in last section, figure R-38). A very interesting aspect of CLF domains is that they are the only ones that are not present in regular FAS type I synthases, which indicates that they may play a key role that could explain at least partially the differences in length and saturation pattern that both system products show. Looking in the literature we realized that CLF plays a central role in iterative PKS synthesis as was explained in the “introduction” section of this work. Knowing the homologies that N-terminal part of PfaC show with the KS-CLF domains of PKS, it is natural to think that they could be also a central piece of for PUFA synthase molecular machinery. These PKS domains have been suggested to have decarboxylase activity. They can introduce starter units priming the ACPs or themselves, and make the carbon chain grow by performing condensation reactions, providing a protective environment for the polyketide in its hydrophobic pocket. For that reasons, we thought we must study the molecular architecture of the KS-CLF domain of PfaC from a structural biology perspective using X-ray crystallography.

To date there are no structures published in the literature for KS-CLF domains that belong to PUFA synthases. Homology structure predictions of KS-CLF can be generated using the structure of a fatty acid synthases like FabF from *E.coli* or the mammalian FAS. However, these models do not provide high-resolution information for protein features such as the contact areas between the two KS domains nor the hydrophobic residues of the tunnel that connects both active centers and was proposed to play an important role in the protection of the nascent polyketide and the determination of chain length. There is only one published structure of a KS-CLF intermolecular dimer from the actinorhodin PKS that was determined by X-ray crystallography and explain some aspects of the PKS synthesis (Keatinge-Clay et al. 2004). Although this structure is predicted to be similar to KS-CLF structure of PUFA synthases, they may show some substantial differences that could be important to understand why the synthesis process leads in one case to produce a polyketide and a fatty acid in the other case.

##### 4.2.4.1. Crystallization process

For obtaining a KS-CLF suitable for crystallization, the main strategy was to generate a pool of constructs represented in figure R-29 and table R-2 that gave us a good chance to be able to produce soluble and stable proteins that were good candidates for crystallization trials. After making four constructs of different length, playing with the stop codon position at the C-terminal end we found a construct, KS-KS3 (M1-A1006), that was productive and stable enough at high concentration (up to 18 mg/ml) for crystallization experiments. KS-CLF was screened using two commercially available screening kits (Hampton research crystal screen 1 and 2) in a single 96-well crystallization plates by the classical sitting drop technique using an initial protein concentration of 12 mg/ml and a 1:1 ratio of protein-precipitant in the pedestal (0.7+0.7  $\mu$ l). Plates were incubated at 22°C and 4°C overnight. Twenty hours after setting the plate, small crystals were found in 2 wells of the plate at 22°C, corresponding to the following crystallization conditions: HR-2 condition 38: Hepes 7.5 0,1M; Polyethylene glycol (PEG) 10,000 18% and HR-1 condition 18: 0.2 M Magnesium acetate tetrahydrate;

0.1 M Sodium cacodylate trihydrate pH 6.5 20% w/v PEG 8,000. After finding these promising crystallization conditions, two 24-well plates were set by hand replicating the same crystallization conditions but playing with the pH of the buffer and changing the concentration of the precipitants. We were not able to grow even small crystals in the 18 HR-1 condition. We focused on the 38 HR-2 condition and played with pH (from 6.5 to 8.5) and with PEG 10,000 (from 15% to 25%). Best crystals with an oblique rectangular shape grew in 48 hours at 22 °C at the original pH 7.5 and with a concentration of PEG 10,000 of 20%. A schematic representation of the crystallization process is presented in figure R-39 showing pictures of the crystals used for further x-ray diffraction experiments.



**Figure R-39.** Crystal growing and optimization. Microscopic images of droplets with protein crystals of KS-CLF. Scale-up crystallization is shown from the 96-well plate to the 24-well plate. The initial small crystals shown in the left panel were improved yielding the bigger and stable crystals of the right panel.

#### 4.2.4.2. Data collection and processing

In order to perform the X-ray diffraction experiment, the best crystals were soaked into a buffer with the cryoprotectant ethylene glycol 10%, Hepes (pH 7.5) 0.09 M, PEG 10,000 18%. This was made to prevent ice formation during freezing that could affect to the quality of the diffraction. Three datasets were collected at the Xaloc beamline in ALBA synchrotron (Barcelona, Spain) using 3 different crystals of the same crystallization drop at 100 K. The data set collected using the crystal that diffracted at

best resolution (1.85 Å) was selected to be processed and was of sufficient quality to be able to begin diffraction image analysis. Images were indexed and integrated with the iMOSFLM package (Battye et al. 2011) and further processed using the CCP4 suite of programs (Winn et al. 2011:4). Analysis of the data set revealed that the crystal had a primitive orthorhombic space group where  $a \neq b \neq c$ . Crystals belong to space group  $P2_1P2_1P2_1$ , with unit cell dimensions  $a=78.23$  Å;  $b=90.83$  Å;  $c=132.42$  Å and one KS-CLF molecule per asymmetric unit.

**Table R-3.** Crystal and x-ray diffraction data of native crystals of KSCLF. Statistics for the highest-resolution shell are shown in parentheses.

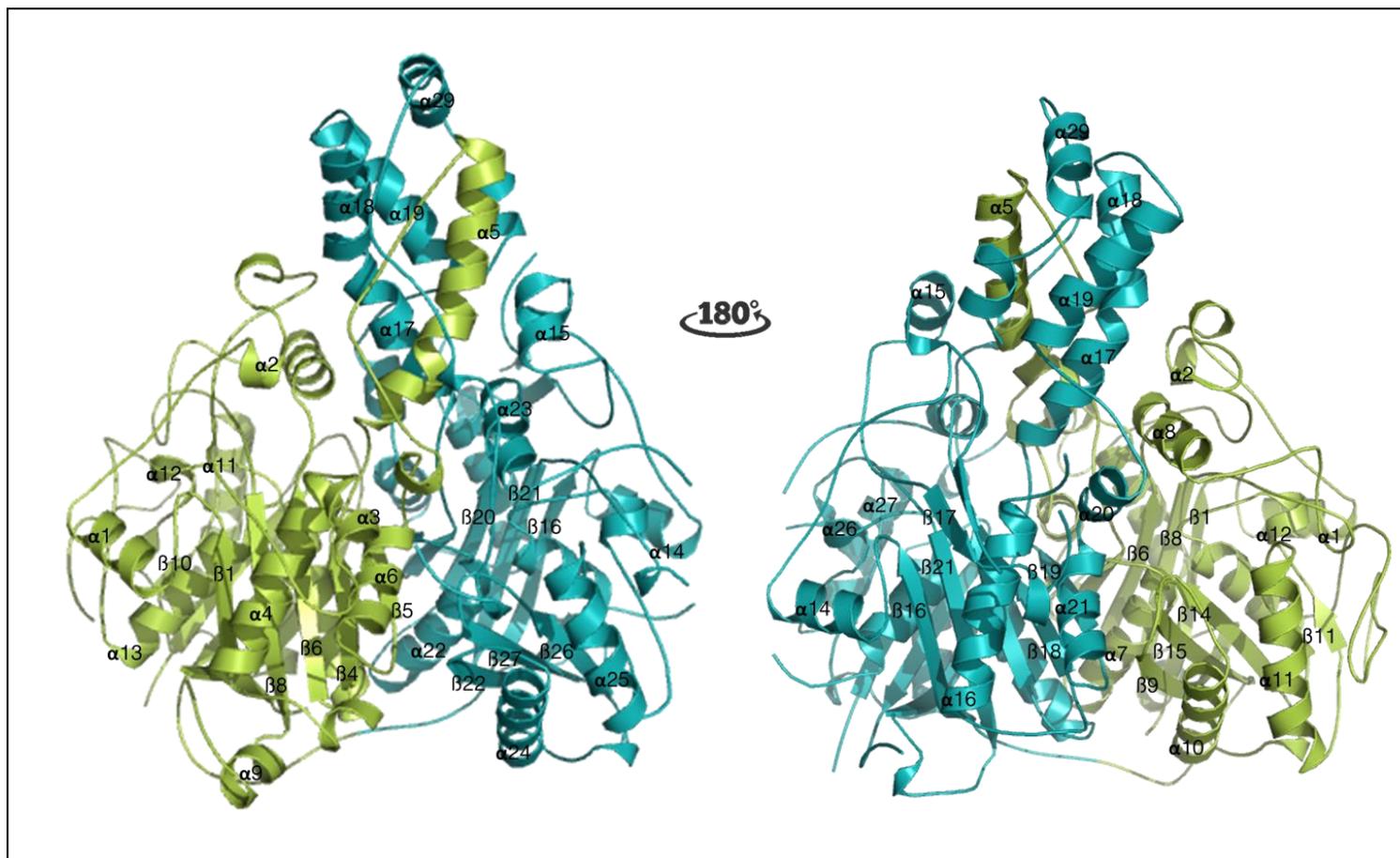
Wavelength	1.127	CC(work)	0.966 (0.891)
Resolution range	27.05 - 1.85 (1.916 - 1.85)	CC(free)	0.951 (0.830)
Space group	P 21 21 21	Number of non-hydrogen atoms	7335
Unit cell	78.2399 90.83 132.42 90 90 90	Macromolecules	6933
Total reflections	151912 (14867)	Protein residues	923
Unique reflections	79983 (7893)	RMS(bonds)	0.007
Multiplicity	1.9 (1.9)	RMS(angles)	1.07
Completeness (%)	0.99 (0.98)	Ramachandran favored (%)	98
Mean I/sigma(I)	12.16 (3.38)	Ramachandran allowed (%)	1.4
Wilson B-factor	20.26	Ramachandran outliers (%)	0.11
R-merge	0.04072 (0.2254)	Rotamer outliers (%)	0.41
R-meas	0.05759 (0.3187)	Clashscore	1.82
CC1/2	0.996 (0.847)	Average B-factor	22.82
CC*	0.999 (0.958)	Macromolecules	22.59
Reflections used in refinement	79978 (7893)	Solvent	26.81
Reflections used for R-free	3939 (402)		
R-work	0.1666 (0.1975)		
R-free	0.2042 (0.2607)		

The structure was determined by molecular replacement using the structural core of curacin polyketide synthase as model (Protein Data Bank accession: 4MZ0). As the KS-CLF domains poses two confronted KS-like domains and the curacin polyketide synthase is composed by a single KS domain the best molecular replacement solution showed two molecules of curacin KS in the asymmetric unit forming a dimer similar to the typical KS-KS dimer. Refinement was performed by Refmac5 and Phenix crystallographic packages. After phase refinement using PHASER, an electron density map was obtained with a resolution of 1.85 Å. Residues 1-966 could be built into the density and refined using Phenix and COOT. A compilation of the Crystal and x-ray diffraction data can be visualized in table R-3.

#### 4.2.4.3. Overall structure of KS-CLF domain of PfaC

KS-CLF domain of PfaC shows an intramolecular dimer architecture formed by two confronted KS domains. These KS domains present the typical thiolase fold, similar to KSs found in fatty acid biosynthesis (Hopwood and Sherman 1990; White et al. 2005). The KS thiolase fold has a core consisting of a duplication of a  $\alpha\beta\alpha\beta\alpha$  motif that can be also identified in the KS and CLF domains in our structure. In the catalytic center of a typical KS domain there is usually a cysteine coordinated with two histidines and for that reason they are known as the CHH group (Robbins et al. 2016). These two histidines could be also found in both active centers of our KS-CLF structure, but only the N-terminal domain carry the catalytic cysteine that is present in all KS domains. In the active center of the C-terminal domain a glutamic acid can be found replacing the cysteine. This

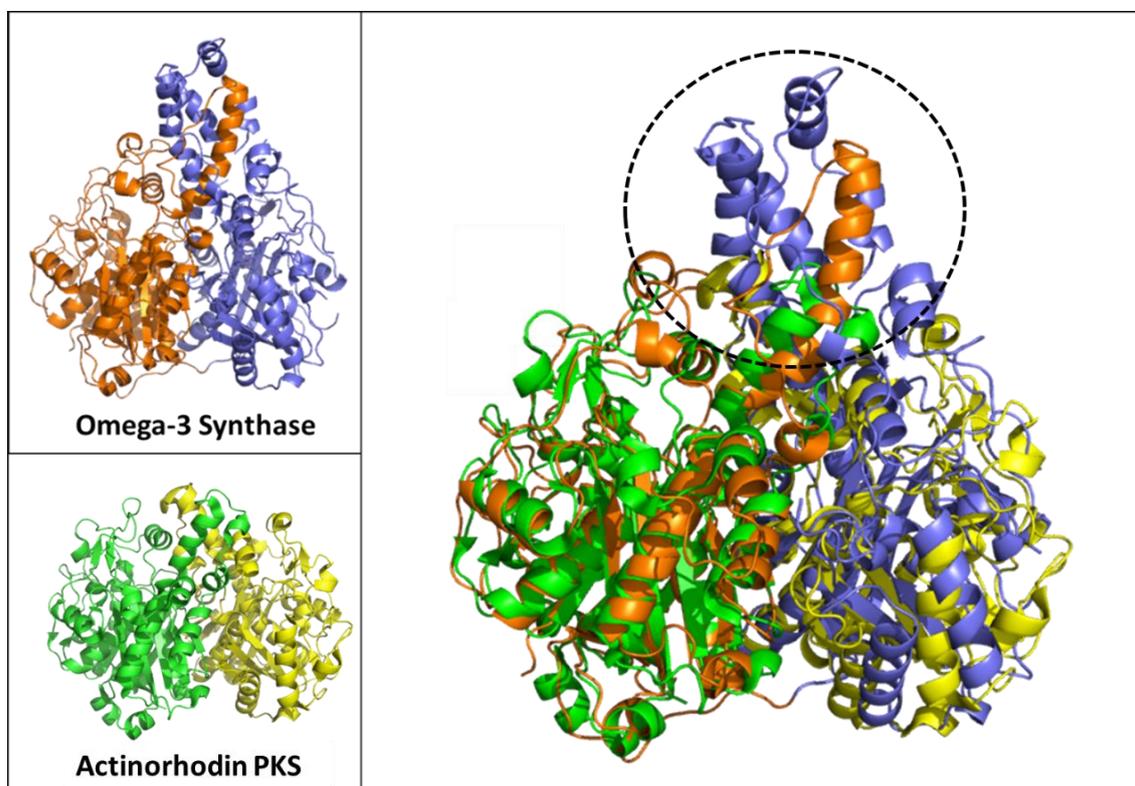
modification is also present in other CLF from PKS complexes and it was described to have decarboxylation activity (Bisang et al. 1999). A helical bundle connects KS and CLF domain and is formed by  $\alpha 5$  from the C-terminal KS domain and  $\alpha 17$ ,  $\alpha 18$   $\alpha 19$  and  $\alpha 29$  from the C-terminal CLF domain. This structure that tightly packs both domains is not present in any KS-CLF structure solved to date and has strong implications for the overall complex stoichiometry architecture that are discussed below. Figure R-40 shows the overall structure of the KS-CLF domain.



**Figure R-40.** Overall structure of KS-CLF domain from PfaC of *Moritella marina*. The protein chain is shown using cartoon representation generated with Pymol software in order to visualize the secondary elements of the structure. KS-CLF is a heterodimeric protein of 107 kDa composed of two domains represented here in green (N-terminal, keto synthase domain) and blue (C-terminal, Chain length factor domain). Alpha helices and beta sheets are numbered in the structure. The two KS domains present a regular thiolase fold and are connected with a helical bundle structure.

#### 4.2.4.4. Helical bundle fold could have structural implications

The only KS-CLF structure solved to date is from the actinorhodin PKS of *Streptomyces coelicolor* (accession number: 1TQY). This structure is also a heterodimer formed by two independent amino acid chains (an intermolecular dimer). On the contrary, the KS-CLF structure solved in this work show an intramolecular dimer as the polypeptide dimerize with itself with both KS and CLF domains faced together. If we compare our KS-CLF with the actinorhodin PKS structure, as is shown in figure R-41, both share the same overall structure organization, with two KS domains confronted showing the same canonical thiolase fold. The main difference that can be appreciated is the presence of a helical bundle within the top of the vertical contact plane of the KS-CLF interface. This structure is formed by alpha helices 5, 17, 18, 19 and 29 and strongly stabilizes the intramolecular dimer. Figure R-41 shows the superimposed protein structures of PfaC KS-CLF and actinorhodin KS-CLF that facilitates the observation of structural differences.



**Figure R-41.** Superimposed protein structure of PfaC KS-CLF and actinorhodin KS-CLF. Left-top panel shows PfaC KS-CLF structure with /KS) N and (CLF) C terminal domains shown in orange and blue respectively and left-bottom panel shows the actinorhodin PKS structure with similar structural layout in green and yellow. Right panel shows superimposed structures of both proteins. The helical bundle of PfaC is highlighted with a dashed circle.

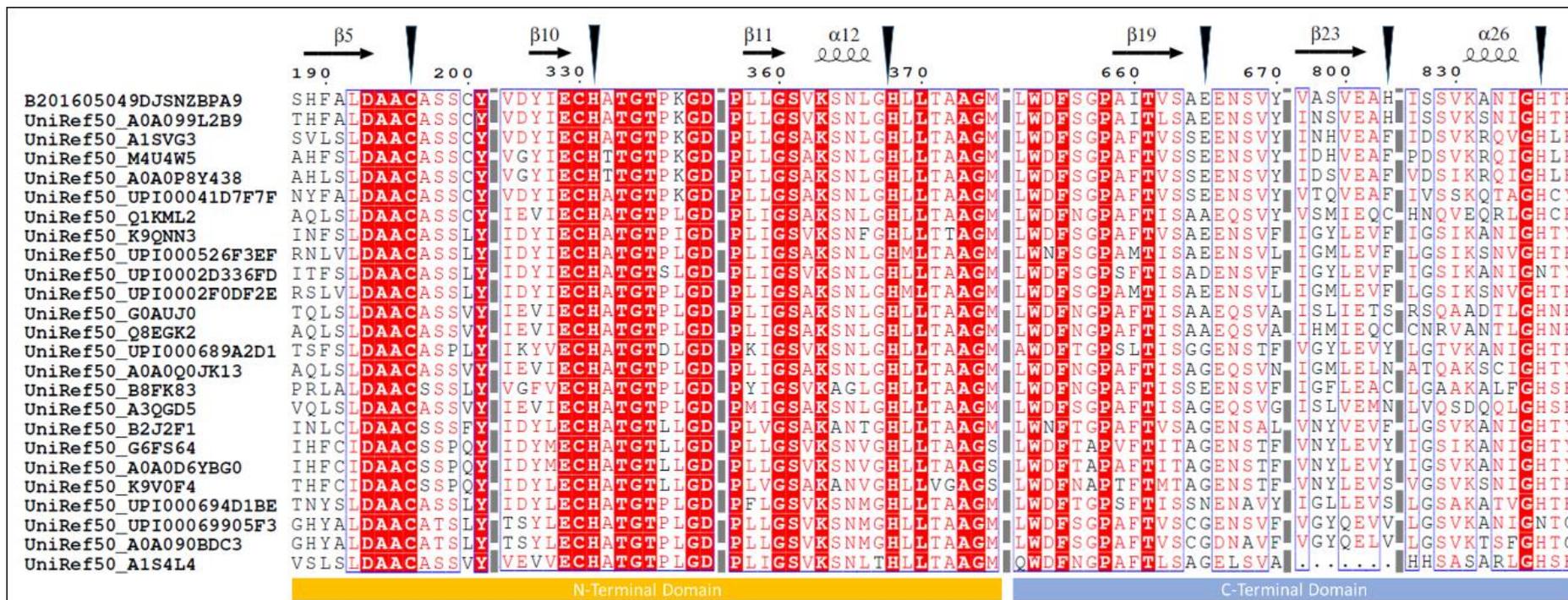
The presence of a helical bundle motif as well as numerous amino acid interactions within the KS-CLF interface have strong implications in the overall structure of PfaC. With the information that this structure give us we can state that PfaC will present a monomer organization within the omega-3 synthase protein complex. This observation is very important to predict the whole complex structural organization and

understand the enzymatic process especially when other structures of key components of the Pfa complex are available.

#### **4.2.4.5. KS-CLF C-terminal domain might have decarboxylation activity**

It was widely reported in the literature that the CLF domain can perform decarboxylation reactions and prime the PKS systems with acetyl or even make the polyketide chain grow condensing different intermediary molecules (Beltran-Alvarez et al. 2007; Bisang et al. 1999; Carreras and Khosla 1998; Kong et al. 2008; Liew et al. 2012; Sun et al. 2009). It was demonstrated that this phenomenon is a consequence of having a glutamic acid replacing the cysteine in the active center of the C-terminal domain (the CLF domain). The understanding of the CLF enzymatic mechanism is definitely relevant to understand PKS synthesis and could also be a key part of the omega-3 synthase machinery. For that reasons, we are interested in demonstrate decarboxylation activity in our system and be able to find the intermediary molecules product of some rounds of condensation of malonyl and some reduction events. It is remarkable to mention here that it was previously shown the inability of KS-CLF to bind malonyl nor acetyl (figure R-38) in vitro as it was shown in the biochemical experiments with radiolabeled substrates.

The first approach was to explore the conservation of the KS-CLF active centers in Omega-3 synthases in order to check if it was possible to find the same amino acid substitution in the CLF C-terminal domain in others omega-3-producer gamma proteobacteria. A structure-based multiple sequence alignment (figure R-42) was made using protein sequences from the Uniprot database and limiting results using the uniref50 filter in order to visualize the conservation through evolution for the residues involved in the catalysis of both active centers.



**Figure R-42.** Structure-based multiple sequence alignment. At the bottom of the figure is indicated what triad correspond to the N-terminal domain (KS, in yellow) and to the C-terminal domain (CLF, in blue). Black arrows on the top of the figure indicate each amino acid that form both catalytic centers. N-terminal domain triad (C196, H331, H368) is highly conserved whereas the first histidine (803) in C-terminal domain triad (E666, H803, H836) is poorly conserved. The alignment was run using protein sequences from the Uniprot database and limiting results using the uniref50 filter.

A perfect conservation of the N-terminal domain (KS) catalytic triad is observed in the structural alignment. This active center is formed by a cysteine and two histidines as in the canonical active site of KS domains. It is also observed that the glutamic acid substitution of the C-terminal domain (CLF) is ubiquitous among the distant relative homologous of KS-CLF from *M. marina* but not perfectly conserved. As previously mentioned, within CLF domains of PKS, this glutamic acid was reported to be responsible for the priming and decarboxylation events occurring during condensation of intermediate polyketide growing chains.

In order to prove the ability of our Omega 3 synthase KS-CLF to bind and even elongate substrates, we decided to repeat the binding assay with radiolabeled acetyl and malonyl-CoA that was previously implemented for the KSAT and ACP domains but this time analyse the results using thin layer chromatography instead of SDS-PAGE gel. TLC allows to check the length of the products synthesized with a particular combination of substrates and proteins. Preliminary results showed that when we incubated our substrate acetyl with the KS-AT didomain we always obtained an unspecific spot, even when we used the truncated versions of KS-AT (S707A) and (C229A). When we incubated the CLF domain with both substrates separately nor mixing CLF with KSAT, ACP or PfaB modules we were not able to observe any changes in the spot pattern. This result indicates that the substrate intermediate that CLF should be able to polymerize is not being generated in this reaction conditions, or the chain transfer mechanism between modules is not working properly due to the isolation of protein domains from their native full-length form. We have also performed the experiments mixing all the previous protein and substrates possible combinations (KS-CLF, KSAT, ACP and PfaB modules) with the modifier modules (KR, DH and ER) also adding NADPH as cofactor, in case any reduction of the extender units was needed to prepare them to be accommodated inside the hydrophobic tunnel of CLF, but none of them generated any conclusive result. This could be indicating that the substrate of our CLF domain is a chain of carbons of a larger size or/and with certain steric and redox state characteristics that would allow a better coupling within the active sites of the protein. Preliminary results of these experiment can be consulted at figure sm-11 of supplementary material.

This negative result indicates that it is possible that the substrate of the KS-CLF domain in omega-3 synthases is none of the typical precursors of fatty acid synthases (acetyl and malonyl). PfaC KSCLF could be working in a similar way to that of type II PKS, in which CLF are often required in later stages of synthesis to condense matured bigger extender units. To prove this, it is necessary to demonstrate the ability of KS-CLF to decarboxylate and elongate bigger substrates and try to crystallize the protein with them into the active sites. This would be probably one of the next challenges for our research group.

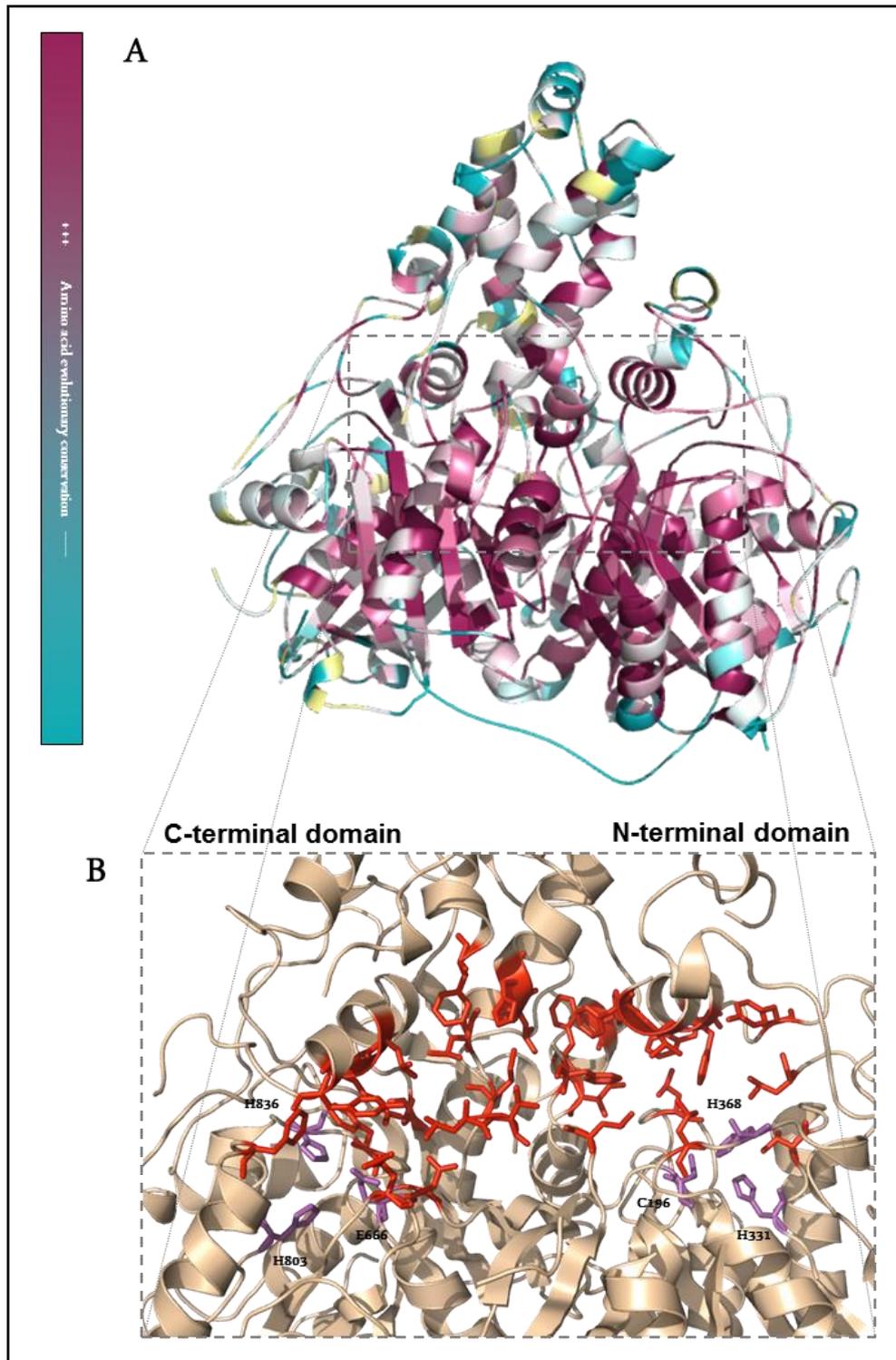
#### **4.2.4.6. KS-CLF has a hydrophobic tunnel that connects both active centers**

Phylogenetic analysis of characterized KS-CLF proteins from PKS type II systems show that these proteins are grouped by chain length, indicating that these domains are important for the determination of the number of carbons that a polyketide can reach in a particular PKS system (Hillenmeyer et al. 2015). KS-CLF coding genes were introduced in type II PKS in coordination with the introduction of cyclases like TcmN and OxyN. Some authors suggest that chain elongation of polyketides that need cyclization events that normally produce polyketides longer than sixteen carbons is

always guided by KS-CLF domains that holds the growing chain in a special hydrophobic tunnel that is found connecting both active centers (Hillenmeyer et al. 2015). The protection of this hydrophobic cavity against the external chemical environment is indispensable for the correct formation of the polyketide. The length of this tunnel and some key amino acids inside determine the final chain length, thus playing a central role in the final polyketide/fatty acid identity. In addition to this and supporting this idea it is known that bigger KS-CLF hydrophobic cavities are correlated with longer polyketides (Hillenmeyer et al. 2015).

All this information found in the literature suggest that at certain stage of the polyketide assembly, the growing chain is transferred to the KS-CLF domain where the final condensations occur and the polyketide can grow protected against undesirable chemical reactions that could affect the integrity of the final molecule. We think that the presence of this KS-CLF domains in all omega-3 synthases known, together with the structural and biochemical information we have, is pointing out a possible functional homology between these two highly related protein complexes. For that reason, we wanted to find in the KS-CLF structure solved here similarities in the amino acid residues that could be disposed in a hydrophobic tunnel similar to the one found in PKS.

In figure R-43 (A) KS-CLF structure is drawn with a color code that represents the conservation of the amino acid sequence across the evolution of the omega-3 synthases, and is based on a sequence alignment. Blue regions are poorly conserved where red ones are conserved among the sequences analyzed. We have found that there is a clear trend towards the conservation of the central core residues and a divergence in the evolution of the peripheral residues. It can be also observed a poor conservation of the helical bundle structure whose importance has been discussed in previous sections. Both active centers are strongly conserved and particularly the region comprising the hydrophobic tunnel that connects them remained highly stable through evolution. This degree of conservation shown for the hydrophobic tunnels suggests that they may be playing a fundamental role in the synthesis of PUFAS in gamma proteobacteria. In order to have a precise view of the amino acid arrangement that forms this hydrophobic cavity, a representation of the residues within the enzyme was made in figure R-43 (B). It can be appreciated that the residues in the central region of the protein are oriented in a way that a hydrophobic cavity is formed in which a chain of polyketide or a fatty acid could supposedly grow.



**Figure R-43.** Amino acid evolutionary conservation within the Pf $\alpha$ C KSCLF central hydrophobic tunnel. Figure shows conservation of residues that cover the internal hydrophobic cavity surface that connects both active centers. (A). structural representation of KS-CLF domain showing the conserved amino acids highlighted in purple where variable residues are colored in green. (B) zoom of the lipophilic tunnel connecting both catalytic triads with the hydrophobic residues colored in red. Presumably, fatty acid chain grows covalently bound to the Cys196 in the N-terminal domain. Both active center triads (Purple) are connected via a hydrophobic tunnel (Red) that holds the growing fatty acid chain.

#### 4.2.5. General Discussion

In this second part of the thesis, we have carried out a detailed biochemical, structural and bioinformatics analysis of the genetic cluster *pfa* of *M. marina* analyzing each of the sequences of the four proteins present in the system that are involved in the production of omega-3 FA. In this way, we have defined their domain organization and predicted their individual biochemical functions as well as their foldings and structures. These predictions indicated the presence of a KS-AT-ACP(x5)-KR'-KR-DH organization in PfaA, a KS'-AT in PfaB, a KS-KS'-DH'-DH-DH'-DH in PfaC and an individual ER motif in PfaD (Figure R-8, R-18, R-21 and R-27 respectively). It has been quite surprising to discover the existence of a KR' pseudo-domain within PfaA sequence, similar to the ones found in bacteria and mammals since it had not been previously described for PUFA producers. PfaC also contains a pseudo DH'-DH repetition similar to the one previously described for *Photobacterium profundum*. The pseudo-KS present in PfaB is also an interesting domain since it shows a predicted thiolase fold but does not pose the C-H-H triad of a regular KS. This motif is not present in most of the EPA producers like *Shewanella baltica* thus it may play a role in the final product determination ability of PfaB. With the information provided by the *in silico* analysis, we have designed twenty six constructions whose sequences cover all the domains coded in the genetic cluster. Ten of these constructs produced soluble proteins that we could purify to perform further experiments. Unfortunately, we have not gotten any soluble protein for the individual C-terminal DH domain of PfaA so future efforts are needed to properly define this module. In the same way, the expression of the KR-DH domain of PfaA has always generated a degradation fragment that could not be separated from the full-length domain. For future experiments, it would be convenient to design a construction that maintains the complete KR-DH domain stable.

Some biochemical experiments were performed to check the substrate specificity of the KS-AT and ACP domains present in the *pfa* cluster of *M. marina*. The presence of multiple domains and their similarity to FAS and PKS systems suggests that omega-3 synthases may act, at least partially, in an iterative and complex way. Although the enzymatic functions of some domains that show homology with the ones present in the system under study are known, biochemical experiments are needed to correctly characterize them. Within PfaA, the first module of our omega-3 synthase, we find a KS, AT, and ACP domains that are, in principle, necessary to select the extender units and begin the synthesis in the classical FAS model. In addition to this, the synthesis has to be carried out with the growing acyl chain covalently bound to the ACPs at least in the initial stages of the process. For these reasons, we decided to focus our efforts in characterizing and defining how these domains could be initiating the carbon assembly in the PfaA module. In addition, we also sought to define substrate affinities for the AT domain of PfaB and the KS-CLF domains of PfaC.

The results of the experiments using radio-labeled substrates with PfaA domains have demonstrated the ACP ability to perform self-malonylation as it was previously reported for other ACP from PKS type II systems. This self-malonylation provides an additional evidence of the close evolutionary relationship between these two systems. The ability of auto acylation implies a certain independence of the ACPs. PfaA ACPs do not need the activity of the AT modules (at least partially), adding a step of complexity and versatility to our Pfa system. In addition to that, the substrate specificity of the AT module to select extender units of malonyl and the ability to strongly charge the ACPs to an extent that the ACPs could not achieve by themselves was demonstrated. This means that the independence of the ACPs is relative and when the AT domains are active they

will be the ones that modulate the acylation state of the ACPs. In the same way as the self-malonylation activity of the ACPs reminds us the PKS type II systems, the affinity demonstrated by the AT domains to select malonyl resembles the one described for FAS systems, evidencing the evolutionary that these three systems share. Surprisingly, and as a remarkable difference with FAS systems, we were not able to demonstrate acetyl binding with any of the wild type domains analyzed. Despite having defined the ability of these systems to use malonyl, a more comprehensive characterization of the AT modules will be needed in a near future in order to test their selectivity against other substrates to properly define the mechanism of extender unit selection.

Experiments with the C229A and S707A mutants of the KS-AT didomain have confirmed that the activity of the AT module is independent of the KS activity. Strikingly, the KS-AT S707A mutant revealed an unknown mechanism by which the KS module is able to actively bind acetyl only when the AT domain is inactive. This may be indicating an enzymatic pressure exerted by the AT module to the KS by which KS has its ability to bind acetyl dramatically reduced while AT is selecting malonyl molecules. More experiments are needed to demonstrate whether this is a novel enzymatic co-operative mechanism or perhaps an artifact product of the non-natural mutants generated.

We have also demonstrated the ability of PfaB to select malonyl units in the same way as the AT domain of PfaA does. We do not know if this activity is a product of the general promiscuity of the AT domains, but we suspect it, since PfaB has been previously described as a final product determinant (DHA or EPA), so we understand that it should act in the final stages of the synthesis. In this way PfaB AT domain should select the extender units in a critical step of the synthesis as it was described to be able to change the final length of the product when is heterologously expressed (Orikasa et al. 2009b). A simple explanation of this would be that in some way PfaB is responsible for transferring the extender units to PfaC but unfortunately, we did not succeed in finding protein-substrate complexes when the KS-CLF and PfaB domains were tested in experiments with radiolabeled substrates. Complementary experiments are needed to demonstrate this hypothesis.

On the other hand, we have presented the first solved structure of a KS-CLF dimeric protein that correspond to the N-terminal domain of PfaC. This is the first KS-CLF solved structure of an omega-3 synthase complex. Protein structures that show homology with the KS-CLF domains of PKS systems are also formed by two KS domains, being the second one an atypical KS domain that lacks the cysteine in its active center but has a glutamic acid instead. Despite sharing certain characteristics regarding the general architecture of the protein, in our PfaC structure we can observe the presence of a helical bundle motif that keeps both domains permanently attached. This feature points out that the full-size protein PfaC has to remain in a monomeric state form due to this intramolecular dimerization discovered here. This may have many structural implications in the general architecture and organization of the protein complex.

In addition, structural evidences such as the presence of a conserved hydrophobic tunnel, typical of PKS KS-CLF domains, or the sequence homologies found for certain protein regions are indicating a possible functional homology in relation to the KS-CLF of PKS systems. In these systems, the hydrophobic tunnel plays a fundamental role protecting the growing chain from the outside chemical environment and preventing cyclization events that may affect to the final molecule structure. As PUFAS are long molecules that need to be reduced before their release, they may need the protection of the KS-CLF interface in the same way as the long polyketides use it in PKS systems. It is possible that KS-CLF domains in omega-3 synthases are playing similar roles as in

the case of PKS where they condense longer and reduced extender units in the last steps of synthesis. More experiments are needed to demonstrate the specificity that our KS-CLF may show for other unconventional extender units and in this way, clarify the mechanism by which the DHA is synthesized.

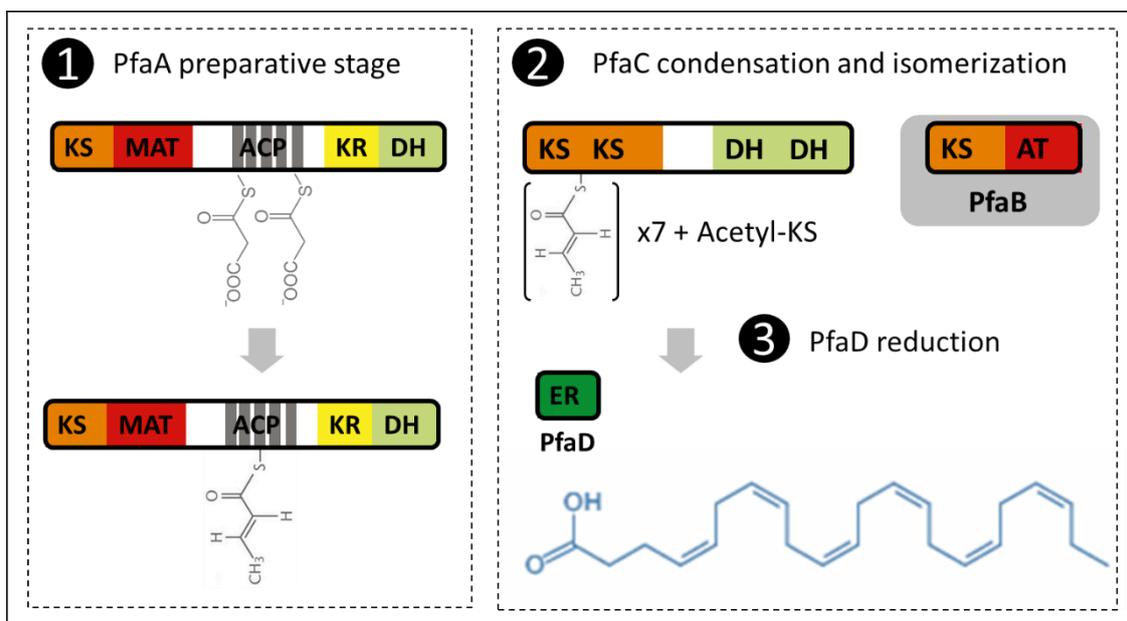
Although we have focused our study of the Pfa condensation domains (KS, AT and ACPs), we believe that for a full characterization of the Pfa biochemical system is necessary to study the individual enzymatic activities of the modifying domains (KR, DH and ER). We are very interested in elucidate how the characteristic double bond pattern of the omega-3 FA is generated. Thus, we need to understand how the stereochemistry of the keto reduction works. KR domains catalyze the stereospecific reduction of the C-3-ketone groups that arise from the chain extension reaction, to give both possible stereoisomers of the resulting hydroxy groups. It was proposed in the literature that the double bond isomer (cis or trans) synthesized in the dehydration process depends on the chirality of the  $\beta$ -OH product of the KR reaction and not on the specificity of the DH domain (Akey et al. 2010; Bonnett et al. 2013; Caffrey 2003; Reid et al. 2003).

A classification of the KR domains was proposed based on the isomer of the  $\beta$ -OH intermediate molecules they produce. The change of the KR stereospecificity is due to the conservation or the absence of a Asp residue located nearby the active site of the protein (Akey et al. 2010; Reid et al. 2003; Weissman 2017). As a rule, the KR domains that poses this Asp residue (that are included in the “B” group) tends to produce D-isomers where the ones that lacks it produce isomers with an L-configuration. When D- and L-isomers are subjected to dehydration (Reid et al. 2003) they produce trans and cis double bonds respectively. Analyzing the sequence of the KR domain present in the PfaA protein of *M. marina* we were able to find the Asp residue (D2177). This amino acid is conserved in all the omega-3 producers analyzed in this work, thus it may produce D-isomers that are reduced to trans double bonds with the adjacent DH domain of PfaA. This biosynthesis scheme is typical of the saturated FA synthesis since the ER domains act only on trans double bonds, producing fully-saturated methylene groups (Weissman 2017). DH domains of PfaC should isomerize the double bonds that are going to be preserved in the final structure

With all the information that we have now, extracted from the experimental part of this work and also with the information available in the literature, we believe that we can try to propose a general model that explain the synthesis of omega-3 FA. We have demonstrated the ability of PfaA to activate its ACP with extender units, either using the ACPs ability of auto-malonylation or the AT domain transferase activity. For that reason, we believe that this first polypeptide can be proposed as the module where the DHA synthesis should start. Apart from loading the malonyl extender units to the ACPs, PfaA might perform the first rounds of condensation extending the carbon chain of these first molecular building blocks. As the KR and DH domains may be in tight contact with the ACPs due to its structural proximity, the first keto reduction and the successive coupled dehydration are likely to generate trans double bonds. This first stage of the synthesis could be considered an iterative process, in which PfaA would act as a “preparative” protein generating matured extender units.

The second part of the synthesis would be carried out by the PfaC protein. As was discussed before, the KS-CLF domains of PKSs are usually associated with the condensation of larger extender units and bringing molecular protection against undesirable chemical reactions during the lasts steps of the synthesis process. The matured extender units generated by PfaA would be condensed within the KS-CLF interface and the associated DH domains would isomerize specific double bonds to the

cis- configuration in order to maintain them in the final FA structure. The ER will be now able to reduce only the double bonds that remain in a trans- configuration. The role of PfaB, at least in *Shewanella pneumatophori* and *M. marina* is to determine whether two carbons are incorporated or not into the FA molecule. Since PfaB showed AT activity it may interact with the growing molecule and with one of the domains with potential to condense (KS or CLF) at some point of the synthesis and influence the incorporation of this “extra” unit to the FA. The synthesis model proposed here is illustrated in figure R-44.



**Figure R-44.** Proposed vectorial model for DHA synthesis. First part of the synthesis is the preparation of the extender units performed by PfaA and represented in panel one. In the second step of the synthesis the condensation of the mature extender units and the isomerization of the double bonds is carried out by PfaC. Double bonds that are now in a cis configuration will be protected from the action of PfaD in the third step. PfaB is shown in a grey box as its mechanism of determination of the omega-3 final length remains unclear.

In this preliminary model we have proposed a synthesis process that has two iterative steps (1 and 2) which are embedded in a general vectorial process with marked directionality. From a general point of view, this biochemical model is similar to the one found in modular PKS as the system proposed has a strong compartmentalization of the biochemical steps. PfaA was proposed to be the one that prepare the extender units to be used in the next step of condensation performed by PfaC. The DH domains of PfaC will isomerize the double bonds protecting them from the action of PfaD. On the other hand, PfaB has not a clear role in the model since its function is not known yet, but we think that it is probably involved in the last steps of the synthesis deciding whether or not an extra extender unit is incorporated to the final molecule before it is released. Thus, the results presented in this thesis allow us to propose the first mechanistic model of DHA synthesis by Pfa synthases. This paves the way for the modification of the system and the production of DHA derivatives with an ample commercial interest.



**Resumen en Castellano**

---



## 5. Resumen en Castellano

---

### 5.1. Introducción

Los lípidos son moléculas que se encuentran en todas las células vivas y están formados por cadenas de hidrocarburos. Éstos forman estructuras importantes como las membranas celulares y también son utilizados como reservorio de energía o incluso actúan como moléculas de señalización (Berg, Tymoczko, and Stryer 2002a; Pucci, Chiovato, and Pinchera 2000). Este trabajo de tesis está enfocado en el estudio de dos familias de proteínas productoras de lípidos de interés comercial: los triglicéridos y los ácidos poliinsaturados omega-3.

Los ácidos grasos son ácidos carboxílicos con una cadena alifática y con un número par de átomos de carbono. De acuerdo con el número de carbonos que poseen éstos se clasifican generalmente en tres grupos: ácidos grasos de cadena corta (menos de 6 carbonos), de cadena media (6-10 carbonos) y cadena larga (más de 12 carbonos). Los ácidos grasos pueden ser saturados (sin doble enlace), insaturados (con un doble enlace) o poliinsaturados (PUFA) (con más de 1 doble enlace) (Chow 1999). El enlace doble puede ocurrir en una configuración *cis*-, cuando los átomos de hidrógeno adyacentes se sitúan al mismo lado de la cadena, o *trans*-, (cuando se encuentran en posiciones opuestas) (Keweloh and Heipieper 1996). Los ácidos grasos se pueden encontrar en las células en forma de triglicéridos (TAG) como forma principal de almacenamiento de energía. Al mantener los ácidos grasos almacenados en moléculas TAG, la célula experimenta cambios de presión osmótica y evita toxicidades, normalmente asociadas a la existencia de ácidos grasos libres. Tanto los TAG como los ácidos grasos poliinsaturados presentan muchas propiedades interesantes tanto para la industria como para la salud y bienestar humano.

Los TAG son moléculas que han sido históricamente utilizadas para muchos propósitos como la producción de aceites utilizados en cocina o la fabricación de todo tipo de grasas y cosméticos. Al ser moléculas muy energéticas con unas propiedades muy características pueden ser fácilmente utilizados para la producción de biodiesel (Balat 2009; Fox and Stachowiak 2007). Históricamente los TAG para fabricar biodieseles se obtienen a partir de biomasa de plantas extraídas de cultivos masivos, principalmente aceite de palma, soja y colza (Metzger and Meier 2011). Las áreas utilizadas para estos cultivos compiten con el uso del suelo con los cultivos de vegetales para el consumo humano, generando de este modo un problema de recursos evidente, más aun si se tienen en cuenta los procesos de industrialización que aumentan la demanda de combustibles fósiles (Hajjari et al. 2017).

Los PUFAs, especialmente los ácidos grasos omega 3 también son moléculas con un gran interés, especialmente desde el punto de vista médico. El término omega-3 hace referencia a la posición de su primer doble que se encuentra alejado tres carbonos del extremo metil del ácido graso. Éstos ácidos grasos son largos (con más de 12 carbonos) poseen al menos 2 insaturaciones que se encuentran siempre en configuración *cis*-. Los PUFAS más relevantes en la naturaleza se conocen como ácido araquidónico (ARA), ácido eicosapentaenoico (EPA) y ácido docosahexaenoico (DHA). Éstos ácidos grasos son esenciales y se ha visto que en mamíferos su consumo está directamente relacionados con una disminución en la aparición de ciertos tipos de cáncer, con la reducción de procesos inflamatorios, alergias o con la protección del

sistema vascular y el nervioso (Cole et al. 2005; Rose and Connolly 1999; Ruxton et al. 2004). Se ha estudiado que en mamíferos existe una deficiencia en la producción de este tipo de ácidos grasos y por ello es necesaria una ingesta en la dieta. La forma de obtención en la actualidad de estos ácidos grasos omega-3 es a partir del aceite de pescado, ya que la fuente principal de producción de los mismos se encuentra en ciertas comunidades bacterianas en los ecosistemas marinos. De este modo se genera de nuevo un problema de recursos ya que las limitaciones naturales en la extracción de pescado imposibilitan satisfacer la demanda mundial de éste tipo de complementos para la alimentación.

Ya que ambas moléculas, TAGs y PUFAs, son interesantes desde el punto de vista biológico e industrial nos planteamos estudiar las proteínas que originan su síntesis en bacterias y profundizar en los conocimientos que tenemos de ellas desde un punto de vista de la ingeniería de proteínas, la biología estructural y la bioquímica. De este modo podremos utilizar sus capacidades de síntesis y modificarlas a nuestro antojo para ser utilizadas en organismos heterólogos fáciles de manipular en el laboratorio, como la bacteria *Escherichia coli*.

Los triglicéridos son sintetizados por células eucariotas de una forma generalizada, pero su síntesis en procariontes está restringida únicamente a algunas bacterias aerobias heterotróficas y algunas cianobacterias. Concretamente dentro del género *Actinomycetes* han sido descritas varias bacterias con capacidades para sintetizar TAG y ceras como *Mycobacterium* sp., *Nocardia* sp., *Rhodococcus* sp., *Micromonospora* sp., *Dietzia* sp., *Gordonia* sp. y *Streptomyces* (Manilla-Pérez et al. 2011; Santala et al. 2014). Estos organismos pueden utilizar muchas fuentes de carbono diferentes para producir TAG y por ello existe un interés en la industria para utilizarlos de modo que puedan transformar sustancias de deshecho en productos útiles como biocombustibles (Nigam and Singh 2011). Las rutas bioquímicas utilizadas por estos microorganismos son diferentes de las utilizadas por organismos eucariotas, especialmente el último paso de la síntesis en el que se condensan un ácido graso con un diglicérido. Para realizar este paso enzimático estas bacterias poseen una enzima denominada como WS/DGAT del término en inglés “wax ester synthase/acyl-CoA:diacylglycerol acyltransferase” (Kalscheuer and Steinbüchel 2003). Éste término hace referencia a que la enzima tiene una capacidad bifuncional, siendo capaz de producir ceras o triglicéridos dependiendo del sustrato utilizado. En nuestro laboratorio hemos descubierto recientemente una enzima de ésta familia perteneciente al organismo *Thermomonospora curvata* (la cual nombramos como tDGAT) y que ha demostrado capacidad para producir rápidamente TAG cuando es expresada en *E.coli* (Lázaro et al., 2017, *in press*). Puesto que este sistema de producción no es óptimo debido a las diferencias intrínsecas entre el organismo original y el hospedador en el que se realiza la expresión heteróloga creemos que el sistema puede ser mejorado. Para ello nos hemos planteado utilizar un mecanismo de evolución dirigida para modificar la proteína y hacer que el sistema de acumulación funcione mejor, ya sea por mejoras en la catálisis o por el acondicionamiento de las interacciones entre la proteína y el sistema de membranas de *E.coli*.

Los ácidos grasos saturados o monoinsaturados se sintetizan de un modo generalizado mediante unas proteínas llamadas ácido graso sintetasas o FAS. Estos sistemas de proteínas se presentan en la naturaleza organizados de dos formas, formando grandes polipéptidos en los que se pueden diferenciar dominios catalíticos individuales, en cuyo caso son clasificados como FAS tipo I, o en proteínas individuales con dominios claramente separados en polipéptidos diferentes, en cuyo caso se trata

de los sistemas FAS tipo II. El primer sistema es típico de mamíferos y hongos y el segundo es típico de bacterias (Bloch and Vance 1977; Campbell and Cronan 2001). En ambos sistemas encontramos los mismos dominios catalíticos característicos que producen la serie de reacciones iterativas necesarias para producir ácidos grasos: ACP o “acyl carrier protein” cuya función es la de portar los intermediarios unidos covalentemente y transferirlos al resto de dominios; la AT o acil transferasa que selecciona los bloques moleculares de malonil-CoA con los que se va a construir la nueva molécula, la KS o keto sintasa que realiza las condensaciones con las que aumenta el tamaño del ácido graso en dos carbonos; y los dominios modificadores KR o keto reductasa, DH o dehidratasa y ER o enoil reductasa, que reducen los grupos ceto generados en las condensaciones y generan el patrón de dobles enlaces característico (Beld, Lee, and Burkart 2014).

Los ácidos grasos poliinsaturados son sintetizados mediante dos rutas diferentes en la naturaleza: la ruta aerobia y la anaerobia. La más común de ellas es la ruta aerobia además de ser la mejor estudiada por estar ampliamente representada en plantas y animales. Esta ruta se caracteriza por la presencia de desaturasas y elongasas que incorporan unidades de malonil nuevas e insaturaciones nuevas a los ácidos grasos previamente existentes generando de esta forma ácidos grasos poliinsaturados (Lee et al. 2016). Estas rutas son extremadamente complicadas desde el punto de vista de la ingeniería genética, por lo que siempre ha existido una ambición por la búsqueda de rutas más simples que puedan ser manipuladas en el laboratorio. Se conoce una ruta alternativa para la producción de ácidos grasos omega 3 conocida como la ruta anaerobia. Esta ruta es llevada a cabo por gamma-proteobacterias marinas como *Shewanella pealeana* o *Moritella marina* las cuales son capaces de producir grandes cantidades de PUFAs (Nichols et al. 1999) utilizando moléculas básicas como el malonil y el acetil. Para realizar esta ruta de síntesis poseen clusters genéticos compuestos por 4-5 pautas abiertas de lectura denominados pfa sintasas (Okuyama et al. 2007; Orikasa et al. 2004). Éstos poseen alta homología con los dominios enzimáticos involucrados en la síntesis de ácidos grasos convencionales (Yazawa 1996b) y tienen una organización modular que recuerda a las FAS tipo I y a las poliketido sintasas bacterianas (PKs), por lo que se presupone que su síntesis debería ser similar. Desgraciadamente con la información existente hasta la fecha no se ha podido proponer un modelo coherente y sólido que explique la totalidad del mecanismo de síntesis de PUFAs por esta vía de las pfa sintasas. Por este motivo se ha planteado el objetivo de estudiar la biología estructural y la bioquímica de los dominios implicados en esta ruta para profundizar en la medida de lo posible en el desarrollo de un modelo definitivo.

## 5.2. Resultados y Discusión

### 5.2.1. Evolución dirigida de tDGAT

Se ha sometido a la proteína tDGAT a un proceso de evolución dirigida *in vitro* con el objetivo de mejorar la productividad de la misma cuando es expresada en *E.coli* y conseguir de esta forma una acumulación de triglicéridos y ceras más eficiente. Además de este objetivo se pretende adquirir ciertos conocimientos estructurales de la proteína. Esta información resultará útil para comprender el mecanismo enzimático de las proteínas WS/DGAT ya que no existe ninguna estructura disponible hasta la fecha.

Para mejorar la proteína se ha diseñado un sistema de evolución dirigida in vitro. El primer paso ha sido la obtención de variabilidad genética utilizando para ello la técnica de mutagénesis aleatoria por PCR (K and M 2002), introduciendo así mutaciones al azar en el fragmento a mutar de la proteína. Después se ha generado una librería de mutantes mediante el proceso de megawhop (Miyazaki 2011) que se ha electroporado a bacterias *E.coli* competentes. Para seleccionar los mutantes que presentaron una mejora en la acumulación de TAG y ceras se puso a punto un sistema de selección basado en un colorante que aporta fluorescencia en presencia de lípidos neutros, el rojo nilo (Greenspan, Mayer, and Fowler 1985). Con la ayuda de este colorante discriminamos entre las diferentes variantes de la proteína, siendo capaces de analizar una librería de más de 10.000 mutantes (puntuales en su mayoría). Se seleccionaron 6 mutantes puntuales y un mutante triple con una producción mejorada de TAG y 4 mutantes cuya producción de TAG había sido completamente bloqueada pero que conservaban producción de ceras. Estos mutantes fueron comprobados y analizados por cromatografía en capa fina.

La mayoría de las mutaciones beneficiosas para la producción de TAG mapeaban en el dominio N-terminal de la proteína, y más concretamente en la superficie de la misma. Además de esto, solamente se seleccionaron mutaciones que disminuían el índice de hidrofobicidad superficial, por lo que entendemos que la mejora en la producción está relacionada de alguna forma con cambios en las interacciones proteína-membrana, ya que éstas son presumiblemente fundamentales en la formación de las gotas lipídicas (Stöveken et al. 2005). En relación a las mutaciones encontradas que bloqueaban selectivamente la producción de TAG pero no la de ceras, no tenemos una explicación completa que explique este fenómeno, pero es probable que las interacciones superficiales de la proteína con los lípidos de membrana o con otras proteínas sean de alguna forma responsables de estos fenotipos.

Para analizar cuantitativamente la mejora en la producción de triglicéridos obtenida se utilizó un sistema de densitometría para medir resultados de cromatografía en capa fina. De esa forma se compararon las producciones de TAG y ceras de los 4 mutantes obtenidos más productivos, obteniéndose valores de hasta un 350% más producción para el mutante P35L.

También analizamos la sobreexpresión de la proteína tDGAT y sus variantes seleccionadas mediante geles de SDS-PAGE, no observándose cambios en la cantidad de proteína soluble producida. De este modo pudimos descartar que cambios en la solubilidad pudiesen estar influenciando la producción de lípidos neutros.

Finalmente, mediante experimentos de ultracentrifugación pudimos comprobar como la proteína tDGAT tiende a unirse a las membranas celulares de *E.coli* y a la fracción lipídica de TAGs, manteniéndose unida a éstos lípidos y formando cuerpos de inclusión. De esta forma aportamos un dato experimental que apoya la hipótesis de que las interacciones proteína-lípidos probablemente juegan un papel fundamental en este tipo de síntesis. En este caso, la insolubilidad de la proteína tDGAT aporta un ambiente hidrofóbico que resulta clave para la preservación de los TAG en forma de cuerpos lipoprotéicos.

### **5.2.2. Estudio bioquímico y estructural de sintasas de ácidos grasos omega-3**

Se ha realizado un estudio bioquímico y estructural de los dominios presentes en las sintasas de ácidos grasos poliinsaturados de bacterias marinas. Para ello se ha

escogido el organismo *Moritella marina* como modelo para estudiar el clúster pfa para la producción de DHA. Debido al carácter modular de las proteínas presentes en este sistema, se realizaron diferentes construcciones representativas para obtener al menos una proteína soluble por cada uno de los dominios con las que una vez purificadas realizar los experimentos de cristalización y bioquímica según el caso. Para la correcta definición de los dominios individuales presentes en nuestro sistema pfa se realizó un estudio bioinformático previo que ayudó a definir correctamente cada una de las unidades funcionales del sistema. Mediante clonación con la técnica de isothermal assembly se consiguió producir menos una construcción soluble para cada uno de los dominios de este complejo protéico.

Una vez purificadas las proteínas, se realizaron ensayos de cristalización probando hasta 96 condiciones diferentes para cada una de ellas y diferentes concentraciones. Se obtuvieron cristales repetibles una construcción del módulo N-terminal de pfaC, que codifica para un dominio keto sintasa- chain length factor (KS-CLF), previamente descrito como elemento central en la síntesis de algunos policétidos de cadena larga. Estos cristales pertenecientes al grupo espacial  $P2_1P2_1P2_1$  difractaron a una resolución 1.85 Å midiendo su celdilla unidad  $a=78.23$  Å;  $b=90.83$  Å;  $c=132.42$  Å. La estructura fue finalmente obtenida por reemplazo molecular utilizando como modelo el núcleo conservado de la proteína curacin policétido sintasa (4MZ0).

La arquitectura general de la proteína presenta dos motivos confrontados con un plegamiento similar al visto anteriormente para otros dominios KS. Cada uno de éstos tiene un plegamiento característico de las proteínas tiolasas, similar a las proteínas encontradas en las FAS (Hopwood and Sherman 1990; White et al. 2005). La característica del dominio C-terminal o CLF es que no presenta la triada catalítica de C-H-H, sino que la cisteína está reemplazada por un ácido glutámico, de una forma similar a los dominios homólogos encontrados en PKS (Keatinge-Clay et al. 2004). Encontramos una diferencia fundamental en nuestra estructura: la presencia de un motivo en la parte superior de la proteína formado por cuatro hélices alfa y que mantiene el dímero intramolecular fuertemente pegado. La presencia de este motivo resulta fundamental a la hora de emitir hipótesis acerca del estado oligomérico de la proteína completa pfaC. Ésta debería encontrarse formando monómeros con una probabilidad alta debido a la imposibilidad de que el dominio KS se disocie para formar dímeros intermoleculares.

Así como otros dominios KS-CLF en PKSs tienen actividad decarboxilasa contra malonil (Beltran-Alvarez et al. 2007; Bisang et al. 1999; Carreras and Khosla 1998; Kong et al. 2008; Liew et al. 2012; Sun et al. 2009), en nuestro caso, a pesar de haber intentado cocristalizar la proteína con éste sustrato, no hemos conseguido ver diferencias en el mapa de densidad electrónica que indiquen la incorporación de unidades de carbono a ninguno de los centros activos. Además de esto, experimentos bioquímicos con malonil y acetil CoA marcados con radiactividad realizados en nuestro laboratorio sugieren de nuevo que nuestra proteína KS-CLF tiene que funcionar preferentemente con otro tipo de sustratos más complejos. La presencia de un túnel hidrofóbico conectando ambos centros activos en nuestra estructura también ha sido descrita anteriormente para otras KS-CLF (Keatinge-Clay et al. 2004). Éste podría tener un papel protector a la hora de polimerizar la cadena de carbonos, impidiendo reacciones de oxidación anómalas antes de finalizar la síntesis de DHA. Se necesitan más estudios que identifiquen el sustrato preferente para proteínas KS-CLF y de ese modo desarrollar un modelo que incluya a éstos en la síntesis de omega-3.

Además de los estudios estructurales previamente descritos se han realizado ensayos de unión proteína-sustrato con acetil y malonil CoA marcados con carbono 14. En estos experimentos se han utilizado los dominios acyl carrier protein (ACP), keto sintasa (KS) y acil transferasa (AT) del dominio pfaA. De este modo se pretendió caracterizar la especificidad de los mismos y desarrollar un modelo preliminar que explique algunos de los pasos de la síntesis de DHA. Tras incubar estos dominios con ambos sustratos radioactivos se pudo comprobar que las ACPs en su forma “holo” son capaces de unir malonyl selectivamente sin la ayuda de ningún dominio AT. Esta habilidad para auto malonilarse ya había sido descrita con anterioridad para otras ACP de PKSs pero nunca para las ACP de complejos pfa (Arthur et al. 2005). Las ATs, tanto de pfaA como de pfaB fueron capaces de unir selectivamente malonil y transferírsele a las ACP, por lo que poseen la habilidad de seleccionar unidades extensoras. Cuando el dominio KS-AT fue mutado en su residuo catalítico responsable de la transferencia de malonil (S707A) sorprendentemente éste adquirió la habilidad de unir acetil y de cargarlo a las ACPs. Quizás este resultado indique que el dominio AT impide la unión de acetil al centro activo de la KS cuando su serina se encuentra en un estado funcional.

Con estos experimentos hemos demostrado que pfaA posee todos los dominios necesarios para secuestrar las unidades de malonil en sus ACP mediante la acción selectora de su dominio AT. Una vez cargadas estas unidades de malonil, al poseer un dominio KS, la proteína pfaA podría realizar las primeras rondas de condensación y posteriormente de reducción, debido a la presencia de dominios modificadores KR y DH, preparando de esta forma las unidades extensoras del sistema. Se ha descubierto la presencia de un residuo conservado en el dominio KR de pfaA (D2177) que indica que el producto de este dominio será un isómero de tipo D-, que son procesados normalmente por los módulos DH adyacentes generando moléculas con dobles enlaces trans- (Akey et al. 2010; Reid et al. 2003; Weissman 2017). Éstos a su vez son normalmente reducidos por el dominio ER, a no ser que sufran una isomerización (Weissman 2017). Nuestro modelo de síntesis propone que los dobles enlaces que saldrían de las unidades extensoras maduras generadas en pfaA serán de tipo trans-. pfaC será el encargado de condensar estas unidades extensoras debido a su homología estructural con los dominios KS-CLF de sistemas PKS y de sus módulos adyacentes DH, que isomerizarán solamente los dobles enlaces que vayan a permanecer en la estructura final de DHA. De este modo proponemos un modelo que en esencia se trata de un proceso unidireccional muy parecido al presente en PKS tipo I y no iterativo como en el caso de los sistemas FAS.

## **References**

---



## 6. References

---

Aasen, Inga Marie, Helga Ertesvåg, Tonje Marita Bjerkan Heggeset, et al, (2016). Thraustochytrids as Production Organisms for Docosahexaenoic Acid (DHA), Squalene, and Carotenoids. *Applied Microbiology and Biotechnology* 100(10): 4309–4321.

Ac, Price, Zhang Ym, Rock Co, and White Sw, 2001. Structure of Beta-Ketoacyl-[Acyl Carrier Protein] Reductase from *Escherichia Coli*: Negative Cooperativity and Its Structural Basis. *Biochemistry* 40(43): 12772–12781.

Adams, P. D., R. W. Grosse-Kunstleve, L.-W. Hung, et al., (2002)PHENIX: Building New Software for Automated Crystallographic Structure Determination. *Acta Crystallographica Section D: Biological Crystallography* 58(11): 1948–1954.

Adarme-Vega, T. Catalina, David K. Y. Lim, Matthew Timmins, et al, (2012). Microalgal Biofactories: A Promising Approach towards Sustainable Omega-3 Fatty Acid Production. *Microbial Cell Factories* 11: 96.

Adarme-Vega, T Catalina, Skye R Thomas-Hall, and Peer M Schenk, (2014). Towards Sustainable Sources for Omega-3 Fatty Acids Production. *Current Opinion in Biotechnology* 26. *Food Biotechnology* □ *Plant Biotechnology*: 14–18.

Agarwal, Avinash Kumar, (2007). Biofuels (Alcohols and Biodiesel) Applications as Fuels for Internal Combustion Engines. *Progress in Energy and Combustion Science* 33(3): 233–271.

Aguinaldo, Anna Marie, and Frances H. Arnold, (2003). Staggered Extension Process (StEP) In Vitro Recombination. *In Directed Evolution Library Creation*. Frances H. Arnold and George Georgiou, eds. Pp. 105–110. *Methods in Molecular Biology*<sup>TM</sup>, 231. Humana Press. <http://link.springer.com/protocol/10.1385/1-59259-395-X%3A105>, accessed January 13, 2015.

Akey, David L., Jamie R. Razelun, Jason Tehranisa, et al., (2010). Crystal Structures of Dehydratase Domains from the Curacin Polyketide Biosynthetic Pathway. *Structure* (London, England : 1993) 18(1): 94–105.

Alberts, A W, and P R Vagelos, (1968). Acetyl CoA Carboxylase. I. Requirement for Two Protein Fractions. *Proceedings of the National Academy of Sciences of the United States of America* 59(2): 561–568.

Alberts, Alfred W., Philip W. Majerus, Barbara Talamo, and P. Roy Vagelos, (1964). Acyl-Carrier Protein. II. Intermediary Reactions of Fatty Acid Synthesis. *Biochemistry* 3(10): 1563–1571.

## References

Alvarez, H. M., R. Kalscheuer, and A. Steinbüchel, (2000). Accumulation and Mobilization of Storage Lipids by *Rhodococcus Opacus* PD630 and *Rhodococcus Ruber* NCIMB 40126. *Applied Microbiology and Biotechnology* 54(2): 218–223.

Alvarez, H. M., and A. Steinbüchel (2002). Triacylglycerols in Prokaryotic Microorganisms. *Applied Microbiology and Biotechnology* 60(4): 367–376.

Andersen, Kasper R., Nina C. Leksa, and Thomas U. Schwartz, (2013). Optimized *E. Coli* Expression Strain LOBSTR Eliminates Common Contaminants from His-Tag Purification. *Proteins* 81(11): 1857–1861.

Arthur, Christopher J., Anna Szafranska, Simon E. Evans, et al., (2005). Self-Malonylation Is an Intrinsic Property of a Chemically Synthesized Type II Polyketide Synthase Acyl Carrier Protein. *Biochemistry* 44(46): 15414–15421.

Bagautdinov, Bagautdin, Yoko Ukita, Masashi Miyano, and Naoki Kunishima. (2008). Structure of 3-Oxoacyl-(Acyl-Carrier Protein) Synthase II from *Thermus Thermophilus* HB8. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 64(Pt 5): 358–366.

Balat, M., (2009). Biodiesel Fuel from Triglycerides via Transesterification—A Review. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 31(14): 1300–1314.

Baldock, Clair, John B. Rafferty, Svetlana E. Sedelnikova, et al.,(1996). A Mechanism of Drug Action Revealed by Structural Studies of Enoyl Reductase. *Science* 274(5295): 2107–2110.

Baldock, Clair, John B. Rafferty, Antoine R. Stuitje, Antoni R. Slabas, and David W. Rice (1998). The X-Ray Structure of *Escherichia Coli* Enoyl Reductase with Bound NAD<sup>+</sup> at 2.1 Å Resolution 1. *Journal of Molecular Biology* 284(5): 1529–1546.

Bao, Wuli, Evelyn Wendt-Pienkowski, and C. Richard Hutchinson, (1998). Reconstitution of the Iterative Type II Polyketide Synthase for Tetracenomycin F2 Biosynthesis. *Biochemistry* 37(22): 8132–8138.

Battye, T. Geoff G., Luke Kontogiannis, Owen Johnson, Harold R. Powell, and Andrew G. W. Leslie, (2011). iMOSFLM: A New Graphical Interface for Diffraction-Image Processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography* 67(Pt 4): 271–281.

Beld, Joris, D. John Lee, and Michael D. Burkart, (2014). Fatty Acid Biosynthesis Revisited: Structure Elucidation and Metabolic Engineering 11(1): 38–59.

Beltran-Alvarez, Pedro, Russell J. Cox, John Crosby, and Thomas J. Simpson, (2007). Dissecting the Component Reactions Catalyzed by the Actinorhodin Minimal Polyketide Synthase. *Biochemistry* 46(50): 14672–14681.

- Berg, Jeremy M., John L. Tymoczko, and Lubert Stryer, (2002). Triacylglycerols Are Highly Concentrated Energy Stores. <https://www.ncbi.nlm.nih.gov/books/NBK22369/>, accessed January 30, 2017.
- Bisang, C., P. F. Long, J. Cortés, et al., (1999). A Chain Initiation Factor Common to Both Modular and Aromatic Polyketide Synthases. *Nature* 401(6752): 502–505.
- Bloch, K., and D. Vance, (1977). Control Mechanisms in the Synthesis of Saturated Fatty Acids. *Annual Review of Biochemistry* 46: 263–298.
- Bloom, Jesse D., Claus O. Wilke, Frances H. Arnold, and Christoph Adami (2004). Stability and the Evolvability of Function in a Model Protein. *Biophysical Journal* 86(5): 2758–2764.
- Böhm, I., I. E. Holzbaur, U. Hanefeld, et al., (1998). Engineering of a Minimal Modular Polyketide Synthase, and Targeted Alteration of the Stereospecificity of Polyketide Chain Extension. *Chemistry & Biology* 5(8): 407–412.
- Bonnett, Shilah A., Jonathan R. Whicher, Kancharla Papireddy, et al., (2013). Structural and Stereochemical Analysis of a Modular Polyketide Synthase Ketoreductase Domain Required for the Generation of a Cis-Alkene. *Chemistry & Biology* 20(6). <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3875705/>, accessed April 9, 2017.
- Broadhurst, R. William, Daniel Nietlispach, Michael P Wheatcroft, Peter F Leadlay, and Kira J Weissman, (2003). The Structure of Docking Domains in Modular Polyketide Synthases. *Chemistry & Biology* 10(8): 723–731.
- Brown, Deborah A, (2001). Lipid Droplets: Proteins Floating on a Pool of Fat. *Current Biology* 11(11): R446–R449.
- Bruegger, Joel, Robert W. Haushalter, Bob Haushalter, et al., (2013). Probing the Selectivity and Protein-protein Interactions of a Nonreducing Fungal Polyketide Synthase Using Mechanism-Based Crosslinkers. *Chemistry & Biology* 20(9): 1135–1146.
- Bugnicourt, Elodie, Patrizia Cinelli, Andrea Lazzeri, and Vera Alejandra Alvarez, (2014). Polyhydroxyalkanoate (PHA): Review of Synthesis, Characteristics, Processing and Potential Applications in Packaging. <http://ri.conicet.gov.ar/handle/11336/4563>, accessed January 27, 2017.
- Caffrey, Patrick, (2003). Conserved Amino Acid Residues Correlating With Ketoreductase Stereospecificity in Modular Polyketide Synthases. *ChemBioChem* 4(7): 654–657.

## References

- Caffrey, Patrick, Debra J. Bevitt, James Staunton, and Peter F. Leadlay, (1992). Identification of DEBS 1, DEBS 2 and DEBS 3, the Multienzyme Polypeptides of the Erythromycin-Producing Polyketide Synthase from *Saccharopolyspora Erythraea*. *FEBS Letters* 304(2): 225–228.
- Campbell, J. W., and J. E. Cronan, (2001). Bacterial Fatty Acid Biosynthesis: Targets for Antibacterial Drug Discovery. *Annual Review of Microbiology* 55: 305–332.
- Carreras, C. W., and C. Khosla, (1998). Purification and in Vitro Reconstitution of the Essential Protein Components of an Aromatic Polyketide Synthase. *Biochemistry* 37(8): 2084–2088.
- Chakravarty, Bornali, Ziwei Gu, Subrahmanyam S. Chirala, Salih J. Wakil, and Florante A. Quioco, (2004). Human Fatty Acid Synthase: Structure and Substrate Selectivity of the Thioesterase Domain. *Proceedings of the National Academy of Sciences of the United States of America* 101(44): 15567–15572.
- Chan, David I., and Hans J. Vogel, (2010). Current Understanding of Fatty Acid Biosynthesis and the Acyl Carrier Protein. *The Biochemical Journal* 430(1): 1–19.
- Chertkov, Olga, Johannes Sikorski, Matt Nolan, et al., (2011). Complete Genome Sequence of *Thermomonospora Curvata* Type Strain (B9T). *Standards in Genomic Sciences* 4(1): 13–22.
- Chow, Ching Kuang, (1999). *Fatty Acids in Foods and Their Health Implications*, Second Edition,. CRC Press.
- Clardy, Jon, Michael A. Fischbach, and Christopher T. Walsh, (2006). New Antibiotics from Bacterial Natural Products. *Nature Biotechnology* 24(12): 1541–1550.
- Cohen-Gonsaud, Martin, Stéphanie Ducasse, Francois Hoh, et al., (2002). Crystal Structure of MabA from *Mycobacterium Tuberculosis*, a Reductase Involved in Long-Chain Fatty Acid Biosynthesis. *Journal of Molecular Biology* 320(2): 249–261.
- Cole, Greg M., Giselle P. Lim, Fusheng Yang, et al., (2005). Prevention of Alzheimer's Disease: Omega-3 Fatty Acid and Phenolic Anti-Oxidant Interventions. *Neurobiology of Aging* 26(1, Supplement). Supplement: Aging, Diabetes, Obesity, Mood and Cognition, "SPARK" Workshop: 133–136.
- Crump, M. P., J. Crosby, C. E. Dempsey, et al., (1997). Solution Structure of the Actinorhodin Polyketide Synthase Acyl Carrier Protein from *Streptomyces Coelicolor* A3(2). *Biochemistry* 36(20): 6000–6008.
- Daniel, Jaiyanth, Chirajyoti Deb, Vinod S. Dubey, et al., (2004). Induction of a Novel Class of Diacylglycerol Acyltransferases and Triacylglycerol Accumulation in *Mycobacterium Tuberculosis* as It Goes into a Dormancy-Like State in Culture. *Journal of Bacteriology* 186(15): 5017–5030.

- Davies, C., R. J. Heath, S. W. White, and C. O. Rock, (2000). The 1.8 Å Crystal Structure and Active-Site Architecture of Beta-Ketoacyl-Acyl Carrier Protein Synthase III (FabH) from *Escherichia Coli*. *Structure* (London, England: 1993) 8(2): 185–195.
- Del Vecchio, Francesca, Hrvoje Petkovic, Steven G. Kendrew, et al., (2003). Active-Site Residue, Domain and Module Swaps in Modular Polyketide Synthases. *Journal of Industrial Microbiology & Biotechnology* 30(8): 489–494.
- Delano, WL, (2002). The PyMOL Molecular Graphics System. <http://www.pymol.org>, accessed November 1, 2016.
- Dillon, Shane C., and Alex Bateman, (2004). The Hotdog Fold: Wrapping up a Superfamily of Thioesterases and Dehydratases. *BMC Bioinformatics* 5: 109.
- Ding, Yunfeng, Li Yang, Shuyan Zhang, et al., (2012). Identification of the Major Functional Proteins of Prokaryotic Lipid Droplets. *Journal of Lipid Research* 53(3): 399–411.
- Dreier, J., and C. Khosla, (2000). Mechanistic Analysis of a Type II Polyketide Synthase. Role of Conserved Residues in the Beta-Ketoacyl Synthase-Chain Length Factor Heterodimer. *Biochemistry* 39(8): 2088–2095.
- Dutta, Somnath, Jonathan R. Whicher, Douglas A. Hansen, et al., (2014). Structure of a Modular Polyketide Synthase. *Nature* 510(7506): 512–517.
- Elder, Robert L.,(1984). The Cosmetic Ingredient Review—a Safety Evaluation Program. *Journal of the American Academy of Dermatology* 11(6): 1168–1174.
- Emsley, P., and K. Cowtan, (2004). Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallographica Section D: Biological Crystallography* 60(12): 2126–2132.
- Escobar, José C., Electo S. Lora, Osvaldo J. Venturini, et al., (2009). Biofuels: Environment, Technology and Food Security. *Renewable and Sustainable Energy Reviews* 13(6–7): 1275–1287.
- Evans, Simon E., Christopher Williams, Christopher J. Arthur, et al., (2008). An ACP Structural Switch: Conformational Differences between the Apo and Holo Forms of the Actinorhodin Polyketide Synthase Acyl Carrier Protein. *ChemBioChem* 9(15): 2424–2432.
- Finn, Robert D., Teresa K. Attwood, Patricia C. Babbitt, et al., (2017). InterPro in 2017—beyond Protein Family and Domain Annotations. *Nucleic Acids Research* 45(Database issue): D190–D199.

## References

Fisher, M., J. T. Kroon, W. Martindale, et al., (2000). The X-Ray Structure of Brassica Napus Beta-Keto Acyl Carrier Protein Reductase and Its Implications for Substrate Binding and Catalysis. *Structure* (London, England: 1993) 8(4): 339–347.

Flugel, Roger S., Yon Hwangbo, Ralph H. Lambalot, John E. Cronan, and Christopher T. Walsh, (2000). Holo-(Acyl Carrier Protein) Synthase and Phosphopantetheinyl Transfer in Escherichia Coli. *Journal of Biological Chemistry* 275(2): 959–968.

Fox, N. J., and G. W. Stachowiak, (2007). Vegetable Oil-Based lubricants—A Review of Oxidation. *Tribology International* 40(7): 1035–1046.

Fukuda, Hideki, Akihiko Kondo, and Hideo Noda, (2001). Biodiesel Fuel Production by Transesterification of Oils. *Journal of Bioscience and Bioengineering* 92(5): 405–416.

Gally, H. U., A. K. Spencer, I. M. Armitage, J. H. Prestegard, and J. E. Cronan, (1978). Acyl Carrier Protein from Escherichia Coli: Characterization by Proton and Fluorine-19 Nuclear Magnetic Resonance and Evidence for Restricted Mobility of the Fatty Acid Chain in Tetradecanoyl-Acyl-Carrier Protein. *Biochemistry* 17(25): 5377–5382.

Ghayourmanesh, Soraya, and Soma Kumar, (1981). Synthesis of Acetoacetyl-CoA by Bovine Mammary Fatty Acid Synthase. *FEBS Letters* 132(2): 231–234.

Gibellini, Federica, and Terry K. Smith, (2010). The Kennedy pathway—De Novo Synthesis of Phosphatidylethanolamine and Phosphatidylcholine. *IUBMB Life* 62(6): 414–428.

Gibson, Daniel G., Lei Young, Ray-Yuan Chuang, et al., (2009). Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nature Methods* 6(5): 343–345.

Greenspan, P., E. P. Mayer, and S. D. Fowler, (1985). Nile Red: A Selective Fluorescent Stain for Intracellular Lipid Droplets. *The Journal of Cell Biology* 100(3): 965–973.

Hajjari, Masoumeh, Meisam Tabatabaei, Mortaza Aghbashlo, and Hossein Ghanavati, (2017). A Review on the Prospects of Sustainable Biodiesel Production: A Global Scenario with an Emphasis on Waste-Oil Biodiesel Utilization. *Renewable and Sustainable Energy Reviews* 72: 445–464.

Hayashi, Shohei, Yasuharu Satoh, Tetsuro Ujihara, Yusuke Takata, and Tohru Dairi, (2016). Enhanced Production of Polyunsaturated Fatty Acids by Enzyme Engineering of Tandem Acyl Carrier Proteins. *Scientific Reports* 6: 35441.

Heath, Richard J., and Charles O. Rock, (1995). Enoyl-Acyl Carrier Protein Reductase (fabI) Plays a Determinant Role in Completing Cycles of Fatty Acid Elongation in Escherichia Coli. *Journal of Biological Chemistry* 270(44): 26538–26542.

- Herbst, Dominik A., Roman P. Jakob, Franziska Zähringer, and Timm Maier, (2016). Mycocerosic Acid Synthase Exemplifies the Architecture of Reducing Polyketide Synthases. *Nature* 531(7595): 533–537.
- Hernández, Martín A., William W. Mohn, Eliana Martínez, et al., (2008). Biosynthesis of Storage Compounds by *Rhodococcus Jostii* RHA1 and Global Identification of Genes Involved in Their Metabolism. *BMC Genomics* 9: 600.
- Hillenmeyer, Maureen E., Gergana A. Vandova, Erin E. Berlew, and Louise K. Charkoudian, (2015). Evolution of Chemical Diversity by Coordinated Gene Swaps in Type II Polyketide Gene Clusters. *Proceedings of the National Academy of Sciences* 112(45): 13952–13957.
- Hitchman, Timothy S., John Crosby, Kate J. Byrom, Russell J. Cox, and Thomas J. Simpson, (1998). Catalytic Self-Acylation of Type II Polyketide Synthase Acyl Carrier Proteins. *Chemistry & Biology* 5(1): 35–47.
- Hopwood, D. A., and D. H. Sherman, (1990). Molecular Genetics of Polyketides and Its Comparison to Fatty Acid Biosynthesis. *Annual Review of Genetics* 24: 37–66.
- ircsofia, (2015). Avoiding Bioenergy Competition for Food Crops and Land. Information Resource Center. <https://ircsofia.wordpress.com/2015/02/03/avoiding-bioenergy-competition-for-food-crops-and-land/>, accessed December 10, 2016.
- Ishige, Takeru, Akio Tani, Yasuyoshi Sakai, and Nobuo Kato, (2003). Wax Ester Production by Bacteria. *Current Opinion in Microbiology* 6(3): 244–250.
- Janßen, Helge Jans, and Alexander Steinbüchel, (2014). Production of Triacylglycerols in *Escherichia Coli* by Deletion of the Diacylglycerol Kinase Gene and Heterologous Overexpression of *atfA* from *Acinetobacter Baylyi* ADP1. *Applied Microbiology and Biotechnology* 98(4): 1913–1924.
- Jenke-Kodama, Holger, Axel Sandmann, Rolf Müller, and Elke Dittmann, (2005). Evolutionary Implications of Bacterial Polyketide Synthases. *Molecular Biology and Evolution* 22(10): 2027–2039.
- Jenni, Simon, Marc Leibundgut, Daniel Boehringer, et al., (2007). Structure of Fungal Fatty Acid Synthase and Implications for Iterative Substrate Shuttling. *Science (New York, N.Y.)* 316(5822): 254–261.
- Jiang, Hui, Scott R. Rajski, and Ben Shen, (2009). Tandem Acyl Carrier Protein Domains in Polyunsaturated Fatty Acid Synthases. *Methods in Enzymology* 459: 79–96.
- Jiang, Hui, Ross Zirkle, James G. Metz, et al., (2008). The Role of Tandem Acyl Carrier Protein Domains in Polyunsaturated Fatty Acid Biosynthesis. *Journal of the American Chemical Society* 130(20): 6336–6337.

## References

- John E. Cronan, Jr, and Charles O. Rock, (2008). Biosynthesis of Membrane Lipids. *EcoSalPlus* <http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosalplus.36.4>, accessed January 5, 2017.
- Johnson, Margaret A., Wolfgang Peti, Torsten Herrmann, Ian A. Wilson, and Kurt Wüthrich, (2006). Solution Structure of Asl1650, an Acyl Carrier Protein from *Anabaena* Sp. PCC 7120 with a Variant Phosphopantetheinylation-Site Sequence. *Protein Science: A Publication of the Protein Society* 15(5): 1030–1041.
- Jörnvall, H., B. Persson, M. Krook, et al., (1995). Short-Chain Dehydrogenases/Reductases (SDR). *Biochemistry* 34(18): 6003–6013.
- Joshi, Anil K., Andrzej Witkowski, and Stuart Smith, (1997). Mapping of Functional Interactions between Domains of the Animal Fatty Acid Synthase by Mutant Complementation in Vitro. *Biochemistry* 36(8): 2316–2322.
- K, Miyazaki, and Takenouchi M, (2002). Creating Random Mutagenesis Libraries Using Megaprimer PCR of Whole Plasmid. *BioTechniques* 33(5): 1033–4, 1036–8.
- Kalscheuer, Rainer, and Alexander Steinbüchel, (2003). A Novel Bifunctional Wax Ester Synthase/Acyl-CoA:Diacylglycerol Acyltransferase Mediates Wax Ester and Triacylglycerol Biosynthesis in *Acinetobacter Calcoaceticus* ADP1. *Journal of Biological Chemistry* 278(10): 8075–8082.
- Kalscheuer, Rainer, Torsten Stölting, and Alexander Steinbüchel, (2006). Microdiesel: *Escherichia Coli* Engineered for Fuel Production. *Microbiology (Reading, England)* 152(Pt 9): 2529–2536.
- Kaneda, T., (1991). Iso- and Anteiso-Fatty Acids in Bacteria: Biosynthesis, Function, and Taxonomic Significance. *Microbiological Reviews* 55(2): 288–302.
- Kapur, Shiven, Alice Y. Chen, David E. Cane, and Chaitan Khosla, (2010). Molecular Recognition between Ketosynthase and Acyl Carrier Protein Domains of the 6-Deoxyerythronolide B Synthase. *Proceedings of the National Academy of Sciences of the United States of America* 107(51): 22066–22071.
- Katz, L., and S. Donadio, (1993). Polyketide Synthesis: Prospects for Hybrid Antibiotics. *Annual Review of Microbiology* 47: 875–912.
- Keatinge-Clay, Adrian T., David A. Maltby, Katalin F. Medzihradzky, Chaitan Khosla, and Robert M. Stroud, (2004). An Antibiotic Factory Caught in Action. *Nature Structural & Molecular Biology* 11(9): 888–893.
- Keatinge-Clay, Adrian T., and Robert M. Stroud, (2006). The Structure of a Ketoreductase Determines the Organization of the  $\beta$ -Carbon Processing Enzymes of Modular Polyketide Synthases. *Structure* 14(4): 737–748.

- Kelley, Lawrence A., Stefans Mezulis, Christopher M. Yates, Mark N. Wass, and Michael J. E. Sternberg, (2015). The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nature Protocols* 10(6): 845–858.
- Keweloh, Heribert, and Hermann J. Heipieper, (1996). Trans Unsaturated Fatty Acids in Bacteria. *Lipids* 31(2): 129–137.
- Khosla, Chaitan, Yinyan Tang, Alice Y. Chen, Nathan A. Schnarr, and David E. Cane, (2007). Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. *Annual Review of Biochemistry* 76: 195–221.
- Kim, Y., and J. H. Prestegard, (1990). Refinement of the NMR Structures for Acyl Carrier Protein with Scalar Coupling Data. *Proteins* 8(4): 377–385.
- Kim, Yangmee, Evgenii L. Kovrigin, and Ziad Eletr, (2006). NMR Studies of Escherichia Coli Acyl Carrier Protein: Dynamic and Structural Differences of the Apo- and Holo-Forms. *Biochemical and Biophysical Research Communications* 341(3): 776–783.
- Kimber, Matthew S., Fernando Martin, Yingjie Lu, et al., (2004). The Structure of (3R)-Hydroxyacyl-Acyl Carrier Protein Dehydratase (FabZ) from Pseudomonas Aeruginosa. *The Journal of Biological Chemistry* 279(50): 52593–52602.
- Klaus, Maja, Matthew P. Ostrowski, Jonas Austerjost, et al., (2016). Protein Protein Interactions, Not Substrate Recognition, Dominates the Turnover of Chimeric Assembly Line Polyketide Synthases. *Journal of Biological Chemistry*: jbc.M116.730531.
- Kong, Rong, Lan Pei Goh, Chong Wai Liew, et al., (2008). Characterization of a Carbonyl-Conjugated Polyene Precursor in 10-Membered Eneidyne Biosynthesis. *Journal of the American Chemical Society* 130(26): 8142–8143.
- Lai, Chiou-Yan, and John E. Cronan, (2004). Isolation and Characterization of  $\beta$ -Ketoacyl-Acyl Carrier Protein Reductase (fabG) Mutants of Escherichia Coli and Salmonella Enterica Serovar Typhimurium. *Journal of Bacteriology* 186(6): 1869–1878.
- Lázaro, B.; Villa, J.A.; Santín, O.; Cabezas, M.; Milagre, C.D.F.; de la Cruz, F. and Moncalián, G. (2017). Heterologous Expression of a Thermophilic Diacylglycerol Acyltransferase Triggers Triglyceride Accumulation in Escherichia coli. *Plos One*. In Press.
- Lee, Je Min, Hyungjae Lee, SeokBeom Kang, and Woo Jung Park, (2016). Fatty Acid Desaturases, Polyunsaturated Fatty Acid Regulation, and Biotechnological Advances. *Nutrients* 8(1). <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4728637/>, accessed January 23, 2017.

## References

- Leesong, Minsun, Barry S Henderson, James R Gillig, John M Schwab, and Janet L Smith, (1996). Structure of a Dehydratase–isomerase from the Bacterial Pathway for Biosynthesis of Unsaturated Fatty Acids: Two Catalytic Activities in One Active Site. *Structure* 4(3): 253–264.
- Lewendon, A., I. A. Murray, W. V. Shaw, M. R. Gibbs, and A. G. Leslie, (1994). Replacement of Catalytic Histidine-195 of Chloramphenicol Acetyltransferase: Evidence for a General Base Role for Glutamate. *Biochemistry* 33(7): 1944–1950.
- Li, Yang, Greg J. Dodge, William D. Fiers, et al., (2015). Functional Characterization of a Dehydratase Domain from the Pikromycin Polyketide Synthase. *Journal of the American Chemical Society* 137(22): 7003–7006.
- Liew, Chong Wai, Martina Nilsson, Ming Wei Chen, et al., (2012). Crystal Structure of the Acyltransferase Domain of the Iterative Polyketide Synthase in Eneidyne Biosynthesis. *Journal of Biological Chemistry* 287(27): 23203–23215.
- Lin, Fengming, Yu Chen, Robert Levine, et al., (2013). Improving Fatty Acid Availability for Bio-Hydrocarbon Production in *Escherichia Coli* by Metabolic Engineering. *PLOS ONE* 8(10): e78595.
- Lowry, Brian, Thomas Robbins, Chih-Hisang Weng, et al., (2013). In Vitro Reconstitution and Analysis of the 6-Deoxyerythronolide B Synthase. *Journal of the American Chemical Society* 135(45): 16809–16812.
- Lu, Peng, Christine Vogel, Rong Wang, Xin Yao, and Edward M. Marcotte, (2007). Absolute Protein Expression Profiling Estimates the Relative Contributions of Transcriptional and Translational Regulation. *Nature Biotechnology* 25(1): 117–124.
- Ma, Fangrui, and Milford A Hanna, (1999) Biodiesel Production: A Review. *Bioresource Technology* 70(1): 1–15.
- Maier, Timm, Marc Leibundgut, and Nenad Ban, (2008). The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science* 321(5894): 1315–1322.
- Manilla-Pérez, Efraín, Alvin Brian Lange, Heinrich Luftmann, Horst Robenek, and Alexander Steinbüchel, (2011). Neutral Lipid Production in *Alcanivorax Borkumensis* SK2 and Other Marine Hydrocarbonoclastic Bacteria. *European Journal of Lipid Science and Technology* 113(1): 8–17.
- Marrakchi, H., Y.-M. Zhang, and C. O. Rock, (2002). Mechanistic Diversity and Regulation of Type II Fatty Acid Synthesis. *Biochemical Society Transactions* 30(6): 1050–1055.
- Massengo-Tiassé, R. Prisca, and John E. Cronan, (2009). Diversity in Enoyl-Acyl Carrier Protein Reductases. *Cellular and Molecular Life Sciences : CMLS* 66(9): 1507.

- Matharu, A. L., R. J. Cox, J. Crosby, K. J. Byrom, and T. J. Simpson, (1998). MCAT Is Not Required for in Vitro Polyketide Synthesis in a Minimal Actinorhodin Polyketide Synthase from *Streptomyces Coelicolor*. *Chemistry & Biology* 5(12): 699–711.
- McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, et al., (2007). Phaser Crystallographic Software. *Journal of Applied Crystallography* 40(4): 658–674.
- McDaniel, R., S. Ebert-Khosla, D. A. Hopwood, and C. Khosla, (1993). Engineered Biosynthesis of Novel Polyketides. *Science* 262(5139): 1546–1550.
- McDaniel, Robert, Camilla M. Kao, Sue J. Hwang, and Chaitan Khosla, (1997). Engineered Intermodular and Intramodular Polyketide Synthase Fusions. *Chemistry & Biology* 4(9): 667–674.
- Meng, Xin, Jianming Yang, Xin Xu, et al., (2009). Biodiesel Production from Oleaginous Microorganisms. *Renewable Energy* 34(1): 1–5.
- Metz, James G., Paul Roessler, Daniel Facciotti, et al., (2001). Production of Polyunsaturated Fatty Acids by Polyketide Synthases in Both Prokaryotes and Eukaryotes. *Science* 293(5528): 290–293.
- Metzger, Jürgen O., and Michael A. R. Meier, (2011). Fats and Oils as Renewable Feedstock for the Chemical Industry. *European Journal of Lipid Science and Technology* 113(1): 1–2.
- Meurer, G., and C. R. Hutchinson, (1995). Functional Analysis of Putative Beta-Ketoacyl:acyl Carrier Protein Synthase and Acyltransferase Active Site Motifs in a Type II Polyketide Synthase of *Streptomyces Glaucescens*. *Journal of Bacteriology* 177(2): 477–481.
- Miyazaki, Kentaro, (2011). MEGAWHOP Cloning: A Method of Creating Random Mutagenesis Libraries via Megaprimer PCR of Whole Plasmids. *Methods in Enzymology* 498: 399–406.
- Moche, M., G. Schneider, P. Edwards, K. Dehesh, and Y. Lindqvist, (1999). Structure of the Complex between the Antibiotic Cerulenin and Its Target, Beta-Ketoacyl-Acyl Carrier Protein Synthase. *The Journal of Biological Chemistry* 274(10): 6031–6034.
- Morris, S. A., W. P. Reville, J. Staunton, and P. F. Leadlay, (1993). Purification and Separation of Holo- and Apo-Forms of *Saccharopolyspora Erythraea* Acyl-Carrier Protein Released from Recombinant *Escherichia Coli* by Freezing and Thawing. *Biochemical Journal* 294(2): 521–527.
- Nawabi, Parwez, Stefan Bauer, Nikos Kyripides, and Athanasios Lykidis, (2011). Engineering *Escherichia Coli* for Biodiesel Production Utilizing a Bacterial Fatty Acid Methyltransferase. *Applied and Environmental Microbiology* 77(22): 8052–8061.

## References

Nichols, David, John Bowman, Kevin Sanderson, et al., (1999). Developments with Antarctic Microorganisms: Culture Collections, Bioactivity Screening, Taxonomy, PUFA Production and Cold-Adapted Enzymes. *Current Opinion in Biotechnology* 10(3): 240–246.

Nigam, Poonam Singh, and Anoop Singh, (2011). Production of Liquid Biofuels from Renewable Resources. *Progress in Energy and Combustion Science* 37(1): 52–68.

Nishida, Takanori, Yoshitake Orikasa, Yukiya Ito, et al., (2006). *Escherichia Coli* Engineered to Produce Eicosapentaenoic Acid Becomes Resistant against Oxidative Damages. *FEBS Letters* 580(11): 2731–2735.

Nordling, Erik, Hans Jörnvall, and Bengt Persson, (2002). Medium-Chain Dehydrogenases/Reductases (MDR). Family Characterizations Including Genome Comparisons and Active Site Modeling. *European Journal of Biochemistry* 269(17): 4267–4276.

Ochsenreither, Katrin, Claudia Glück, Timo Stressler, Lutz Fischer, and Christoph Sydtk, (2016). Production Strategies and Applications of Microbial Single Cell Oils. *Frontiers in Microbiology* 7. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5050229/>, accessed April 12, 2017.

Oefner, C., H. Schulz, A. D'Arcy, and G. E. Dale, (2006). Mapping the Active Site of *Escherichia Coli* Malonyl-CoA–acyl Carrier Protein Transacylase (FabD) by Protein Crystallography. *Acta Crystallographica Section D: Biological Crystallography* 62(6): 613–618.

Okuyama, Hidetoshi, Yoshitake Orikasa, Takanori Nishida, Kazuo Watanabe, and Naoki Morita, (2007). Bacterial Genes Responsible for the Biosynthesis of Eicosapentaenoic and Docosahexaenoic Acids and Their Heterologous Expression. *Applied and Environmental Microbiology* 73(3): 665–670.

Olsen, J. G., A. Kadziola, P. von Wettstein-Knowles, M. Siggaard-Andersen, and S. Larsen, (2001). Structures of Beta-Ketoacyl-Acyl Carrier Protein Synthase I Complexed with Fatty Acids Elucidate Its Catalytic Machinery. *Structure* (London, England: 1993) 9(3): 233–243.

Olsen, Johan Gotthardt, Anders Kadziola, Penny von Wettstein-Knowles, Mads Siggaard-Andersen, and Sine Larsen, (2001). Structures of  $\beta$ -Ketoacyl-Acyl Carrier Protein Synthase I Complexed with Fatty Acids Elucidate Its Catalytic Machinery. *Structure* 9(3): 233–243.

Orikasa, Y., A. Yamada, R. Yu, et al., (2004). Characterization of the Eicosapentaenoic Acid Biosynthesis Gene Cluster from *Shewanella* Sp. Strain SCRC-2738. *Cellular and Molecular Biology* (Noisy-Le-Grand, France) 50(5): 625–630.

- Orikasa, Yoshitake, Takanori Nishida, Akiko Yamada, et al., (2006). Recombinant Production of Docosahexaenoic Acid in a Polyketide Biosynthesis Mode in *Escherichia Coli*. *Biotechnology Letters* 28(22): 1841–1847.
- Orikasa, Yoshitake, Mika Tanaka, Shinji Sugihara, et al., (2009). *pfaB* Products Determine the Molecular Species Produced in Bacterial Polyunsaturated Fatty Acid Biosynthesis. *FEMS Microbiology Letters* 295(2): 170–176.
- Oyola-Robles, Delise, Darren C Gay, Uldaeliz Trujillo, et al., (2013). Identification of Novel Protein Domains Required for the Expression of an Active Dehydratase Fragment from a Polyunsaturated Fatty Acid Synthase. *Protein Science: A Publication of the Protein Society* 22(7): 954–963.
- Pereira, Suzette L, Amanda E Leonard, and Pradip Mukerji, (2003). Recent Advances in the Study of Fatty Acid Desaturases from Animals and Lower Eukaryotes. *Prostaglandins, Leukotrienes and Essential Fatty Acids* 68(2): 97–106.
- Plourde, Mélanie, and Stephen C. Cunnane, (2007). Extremely Limited Synthesis of Long Chain Polyunsaturates in Adults: Implications for Their Dietary Essentiality and Use as Supplements. *Applied Physiology, Nutrition, and Metabolism = Physiologie Appliquee, Nutrition Et Metabolisme* 32(4): 619–634.
- Price, A. C., K. H. Choi, R. J. Heath, et al., (2001). Inhibition of Beta-Ketoacyl-Acyl Carrier Protein Synthases by Thiolactomycin and Cerulenin. Structure and Mechanism. *The Journal of Biological Chemistry* 276(9): 6551–6559.
- Pucci, E., L. Chiovato, and A. Pinchera, (2000). Thyroid and Lipid Metabolism. *International Journal of Obesity and Related Metabolic Disorders: Journal of the International Association for the Study of Obesity* 24 Suppl 2: S109-112.
- Qiu, X., C. A. Janson, W. W. Smith, et al., (2001). Refined Structures of Beta-Ketoacyl-Acyl Carrier Protein Synthase III. *Journal of Molecular Biology* 307(1): 341–356.
- Qiu, Xiayang, Anthony E. Choudhry, Cheryl A. Janson, et al., (2005). Crystal Structure and Substrate Specificity of the  $\beta$ -Ketoacyl-Acyl Carrier Protein Synthase III (FabH) from *Staphylococcus Aureus*. *Protein Science: A Publication of the Protein Society* 14(8): 2087–2094.
- Ranganathan, A., M. Timoney, M. Bycroft, et al., (1999). Knowledge-Based Design of Bimodular and Trimodular Polyketide Synthases Based on Domain and Module Swaps: A Route to Simple Statin Analogues. *Chemistry & Biology* 6(10): 731–741.
- Reeves, Christopher D., Sumati Murli, Gary W. Ashley, et al., (2001). Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations. *Biochemistry* 40(51): 15464–15470.

## References

- Reid, Ralph, Misty Piagentini, Eduardo Rodriguez, et al., (2003). A Model of Structure and Catalysis for Ketoreductase Domains in Modular Polyketide Synthases. *Biochemistry* 42(1): 72–79.
- Ridley, Christian P., Ho Young Lee, and Chaitan Khosla, (2008). Evolution of Polyketide Synthases in Bacteria. *Proceedings of the National Academy of Sciences* 105(12): 4595–4600.
- Robbins, Thomas, Joshua Kapilivsky, David E. Cane, and Chaitan Khosla, (2016). Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. *Biochemistry* 55(32): 4476–4484.
- Rock, Charles O., and John E. Cronan, (1996). *Escherichia Coli* as a Model for the Regulation of Dissociable (Type II) Fatty Acid Biosynthesis. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 1302(1): 1–16.
- Rose, David P., and Jeanne M. Connolly, (1999). Omega-3 Fatty Acids as Cancer Chemopreventive Agents. *Pharmacology & Therapeutics* 83(3): 217–244.
- Röttig, Annika, and Alexander Steinbüchel, (2013). Acyltransferases in Bacteria. *Microbiology and Molecular Biology Reviews* 77(2): 277–321.
- Rubio-Rodríguez, Nuria, Sagrario Beltrán, Isabel Jaime, et al., (2010). Production of Omega-3 Polyunsaturated Fatty Acid Concentrates: A Review. *Innovative Food Science & Emerging Technologies* 11(1): 1–12.
- Rucker, Joanna, Julie Paul, Blaine A. Pfeifer, and Kyongbum Lee, (2013). Engineering *E. Coli* for Triglyceride Accumulation through Native and Heterologous Metabolic Reactions. *Applied Microbiology and Biotechnology* 97(6): 2753–2759.
- Ruxton, C. H. S., S. C. Reed, M. J. A. Simpson, and K. J. Millington, (2004). The Health Benefits of Omega-3 Polyunsaturated Fatty Acids: A Review of the Evidence. *Journal of Human Nutrition and Dietetics* 17(5): 449–459.
- Samel, Stefan A., Georg Schoenafinger, Thomas A. Knappe, Mohamed A. Marahiel, and Lars-Oliver Essen, (2007). Structural and Functional Insights into a Peptide Bond-Forming Bidomain from a Nonribosomal Peptide Synthetase. *Structure* 15(7): 781–792.
- Santala, Suvi, Elena Efimova, Virpi Kivinen, et al., (2011). Improved Triacylglycerol Production in *Acinetobacter Baylyi* ADP1 by Metabolic Engineering. *Microbial Cell Factories* 10: 36.
- Santala, Suvi, Elena Efimova, Perttu Koskinen, Matti Tapani Karp, and Ville Santala, (2014). Rewiring the Wax Ester Production Pathway of *Acinetobacter Baylyi* ADP1. *ACS Synthetic Biology* 3(3): 145–151.

Schindelin, Johannes, Ignacio Arganda-Carreras, Erwin Frise, et al., (2012). Fiji: An Open-Source Platform for Biological-Image Analysis. *Nature Methods* 9(7): 676–682.

Serre, L., E. C. Verbree, Z. Dauter, A. R. Stuitje, and Z. S. Derewenda, (1995). The Escherichia Coli Malonyl-CoA:acyl Carrier Protein Transacylase at 1.5-Å Resolution. Crystal Structure of a Fatty Acid Synthase Component. *The Journal of Biological Chemistry* 270(22): 12961–12964.

Shen, Y., P. Yoon, T. W. Yu, et al., (1999). Ectopic Expression of the Minimal whiE Polyketide Synthase Generates a Library of Aromatic Polyketides of Diverse Sizes and Shapes. *Proceedings of the National Academy of Sciences of the United States of America* 96(7): 3622–3627.

Sherkhanov, Saken, Tyler P. Korman, Steven G. Clarke, and James U. Bowie, (2016). Production of FAME Biodiesel in E. Coli by Direct Methylation with an Insect Enzyme. *Scientific Reports* 6: 24239.

Shulse, Christine N., and Eric E. Allen, (2011). Widespread Occurrence of Secondary Lipid Biosynthesis Potential in Microbial Lineages. *PLoS One* 6(5): e20146.

Smith, Stuart, and Shiou-Chuan Tsai, (2007). The Type I Fatty Acid and Polyketide Synthases: A Tale of Two Megasyntases. *Natural Product Reports* 24(5): 1041–1072.

Smith, Stuart, Andrzej Witkowski, and Anil K. Joshi, (2003). Structural and Functional Organization of the Animal Fatty Acid Synthase. *Progress in Lipid Research* 42(4): 289–317.

Staunton, James, and Kira J. Weissman, (2001). Polyketide Biosynthesis: A Millennium Review 18(4): 380–416.

Stöveken, Tim, Rainer Kalscheuer, Ursula Malkus, Rudolf Reichelt, and Alexander Steinbüchel, (2005). The Wax Ester Synthase/Acyl Coenzyme A:Diacylglycerol Acyltransferase from *Acinetobacter* Sp. Strain ADP1: Characterization of a Novel Type of Acyltransferase. *Journal of Bacteriology* 187(4): 1369–1376.

Stöveken, Tim, Rainer Kalscheuer, and Alexander Steinbüchel, (2009). Both Histidine Residues of the Conserved HHXXXDG Motif Are Essential for Wax Ester Synthase/Acyl-CoA:diacylglycerol Acyltransferase Catalysis. *European Journal of Lipid Science and Technology* 111(2): 112–119.

Stöveken, Tim, and Alexander Steinbüchel, (2008). Bacterial Acyltransferases as an Alternative for Lipase-Catalyzed Acylation for the Production of Oleochemicals and Fuels. *Angewandte Chemie (International Ed. in English)* 47(20): 3688–3694.

## References

- Subramanian, Sandhya, Jan Abendroth, Isabelle Q. H. Phan, et al., (2011). Structure of 3-Ketoacyl-(Acyl-Carrier-Protein) Reductase from *Rickettsia prowazekii* at 2.25 Å Resolution. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 67(Pt 9): 1118–1122.
- Sugihara, Shinji, Yoshitake Orikasa, and Hidetoshi Okuyama, (2010). The *Escherichia coli* Highly Expressed *entD* Gene Complements the *pfaE* Deficiency in a *Pfa* Gene Clone Responsible for the Biosynthesis of Long-Chain N-3 Polyunsaturated Fatty Acids. *FEMS Microbiology Letters* 307(2): 207–211.
- Sun, Huihua, Rong Kong, Di Zhu, et al., (2009). Products of the Iterative Polyketide Synthases in 9- and 10-Membered Ene-diyne Biosynthesis. *Chemical Communications (Cambridge, England)*(47): 7399–7401.
- Sundermann, Uschi, Kenny Bravo-Rodriguez, Stephan Klopries, et al., (2013). Enzyme-Directed Mutasynthesis: A Combined Experimental and Theoretical Approach to Substrate Recognition of a Polyketide Synthase. *ACS Chemical Biology* 8(2): 443–450.
- Szu, Ping-Hui, Sridhar Govindarajan, Michael J. Meehan, et al., (2011). Analysis of the Ketosynthase-Chain Length Factor Heterodimer from the Fredericamycin Polyketide Synthase. *Chemistry & Biology* 18(8): 1021–1031.
- Tang, Yi, Andrew T. Koppisch, and Chaitan Khosla, (2004) The Acyltransferase Homologue from the Initiation Module of the R1128 Polyketide Synthase Is an Acyl-ACP Thioesterase That Edits Acetyl Primer Units. *Biochemistry* 43(29): 9546–9555.
- Tang, Yi, Shiou-Chuan Tsai, and Chaitan Khosla, (2003). Polyketide Chain Length Control by Chain Length Factor. *Journal of the American Chemical Society* 125(42): 12708–12709.
- Tang, Yinyan, Alice Y. Chen, Chu-Young Kim, David E. Cane, and Chaitan Khosla, (2007). Structural and Mechanistic Analysis of Protein Interactions in Module 3 of the 6-Deoxyerythronolide B Synthase. *Chemistry & Biology* 14(8): 931–943.
- Tang, Yinyan, Chu-Young Kim, Irimpan I. Mathews, David E. Cane, and Chaitan Khosla, (2006). The 2.7-Å Crystal Structure of a 194-kDa Homodimeric Fragment of the 6-Deoxyerythronolide B Synthase. *Proceedings of the National Academy of Sciences of the United States of America* 103(30): 11124–11129.
- Tapiero, H, G Nguyen Ba, P Couvreur, and K. D Tew, (2002). Polyunsaturated Fatty Acids (PUFA) and Eicosanoids in Human Health and Pathologies. *Biomedicine & Pharmacotherapy* 56(5): 215–222.
- Toomey, R. E., and S. J. Wakil, (1966). Studies on the Mechanism of Fatty Acid Synthesis. XV. Preparation and General Properties of Beta-Ketoacyl Acyl Carrier Protein Reductase from *Escherichia coli*. *Biochimica Et Biophysica Acta* 116(2): 189–197.

- Trujillo, Uldaeliz, Edwin Vázquez-Rosa, Delise Oyola-Robles, et al., (2013). Solution Structure of the Tandem Acyl Carrier Protein Domains from a Polyunsaturated Fatty Acid Synthase Reveals Beads-on-a-String Configuration. *PLOS ONE* 8(2): e57859.
- Vignot, S., S. Faivre, D. Aguirre, and E. Raymond, (2005). mTOR-Targeted Therapy of Cancer with Rapamycin Derivatives. *Annals of Oncology* 16(4): 525–537.
- Villa, Juan A, Matilde Cabezas, Fernando de la Cruz, and Gabriel Moncalián, (2014). Use of Limited Proteolysis and Mutagenesis to Identify Folding Domains and Sequence Motifs Critical for Wax Ester Synthase/Acyl Coenzyme A:diacylglycerol Acyltransferase Activity. *Applied and Environmental Microbiology* 80(3): 1132–1141.
- Wakil, S. J., J. K. Stoops, and V. C. Joshi, (1983). Fatty Acid Synthesis and Its Regulation. *Annual Review of Biochemistry* 52: 537–579.
- Wall, Rebecca, R. Paul Ross, Gerald F. Fitzgerald, and Catherine Stanton, (2010). Fatty Acids from Fish: The Anti-Inflammatory Potential of Long-Chain Omega-3 Fatty Acids. *Nutrition Reviews* 68(5): 280–289.
- Wältermann, M., H. Luftmann, D. Baumeister, R. Kalscheuer, and A. Steinbüchel, (2000). *Rhodococcus Opacus* Strain PD630 as a New Source of High-Value Single-Cell Oil? Isolation and Characterization of Triacylglycerols and Other Storage Lipids. *Microbiology (Reading, England)* 146 ( Pt 5): 1143–1149.
- Wältermann, Marc, Andreas Hinz, Horst Robenek, et al., (2005). Mechanism of Lipid-Body Formation in Prokaryotes: How Bacteria Fatten up. *Molecular Microbiology* 55(3): 750–763.
- Wältermann, Marc, and Alexander Steinbüchel, (2005). Neutral Lipid Bodies in Prokaryotes: Recent Insights into Structure, Formation, and Relationship to Eukaryotic Lipid Depots. *Journal of Bacteriology* 187(11): 3607–3619.
- Walther, Tobias C., and Robert V. Farese Jr., (2009). The Life of Lipid Droplets. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1791(6). Lipid Droplets as Dynamic Organelles Connecting Influx, Efflux and Storage of Lipids: 459–466.
- Weeks, G., and S. J. Wakil, (1968). Studies on the Mechanism of Fatty Acid Synthesis. 18. Preparation and General Properties of the Enoyl Acyl Carrier Protein Reductases from *Escherichia Coli*. *The Journal of Biological Chemistry* 243(6): 1180–1189.
- Weissman, Kira J., (2006). The Structural Basis for Docking in Modular Polyketide Biosynthesis. *ChemBiochem: A European Journal of Chemical Biology* 7(3): 485–494.

## References

- Weissman, Kira J, (2017). Polyketide Stereocontrol: A Study in Chemical Biology. *Beilstein Journal of Organic Chemistry* 13: 348–371.
- Wertheimer, E., and B. Shapiro, (1948). The Physiology of Adipose Tissue. *Physiological Reviews* 28(4): 451–464.
- White, Stephen W., Jie Zheng, Yong-Mei Zhang, and Charles O. Rock, (2005). The Structural Biology of Type II Fatty Acid Biosynthesis. *Annual Review of Biochemistry* 74(1): 791–831.
- Winn, Martyn D., Charles C. Ballard, Kevin D. Cowtan, et al., (2011). Overview of the CCP4 Suite and Current Developments. *Acta Crystallographica Section D: Biological Crystallography* 67(Pt 4): 235–242.
- Witkowski, Andrzej, Anil K. Joshi, and Stuart Smith, (1997). Characterization of the Interthiol Acyltransferase Reaction Catalyzed by the  $\beta$ -Ketoacyl Synthase Domain of the Animal Fatty Acid Synthase. *Biochemistry* 36(51): 16338–16344.
- Wong, Hing C., Gaohua Liu, Yong-Mei Zhang, Charles O. Rock, and Jie Zheng, (2002) The Solution Structure of Acyl Carrier Protein from *Mycobacterium Tuberculosis*. *The Journal of Biological Chemistry* 277(18): 15874–15880.
- Yang, Li, Yunfeng Ding, Yong Chen, et al., (2012). The Proteomics of Lipid Droplets: Structure, Dynamics, and Functions of the Organelle Conserved from Bacteria to Humans. *Journal of Lipid Research* 53(7): 1245–1253.
- Yazawa, K., (1996). Production of Eicosapentaenoic Acid from Marine Bacteria. *Lipids* 31 Suppl: S297-300.
- Yoshida, Kiyohito, Mikako Hashimoto, Ryuji Hori, et al., (2016). Bacterial Long-Chain Polyunsaturated Fatty Acids: Their Biosynthetic Genes, Functions, and Practical Use. *Marine Drugs* 14(5).
- Zhang, Y., and J. E. Cronan, (1998). Transcriptional Analysis of Essential Genes of the *Escherichia Coli* Fatty Acid Biosynthesis Gene Cluster by Functional Replacement with the Analogous *Salmonella Typhimurium* Gene Cluster. *Journal of Bacteriology* 180(13): 3295–3303.
- Zhang, Y. M., M. S. Rao, R. J. Heath, et al., (2001). Identification and Analysis of the Acyl Carrier Protein (ACP) Docking Site on Beta-Ketoacyl-ACP Synthase III. *The Journal of Biological Chemistry* 276(11): 8231–8238.
- Zhang, Yong-Mei, Bainan Wu, Jie Zheng, and Charles O. Rock, (2003). Key Residues Responsible for Acyl Carrier Protein and Beta-Ketoacyl-Acyl Carrier Protein Reductase (FabG) Interaction. *The Journal of Biological Chemistry* 278(52): 52935–52943.
- Zhao, Huimin, and Wenjuan Zha, (2006). In Vitro “Sexual” Evolution through the PCR-Based Staggered Extension Process (StEP). *Nature Protocols* 1(4): 1865–1871.

Zheng, Jianting, Clint A. Taylor, Shawn K. Piasecki, and Adrian T. Keatinge-Clay, (2010). Structural and Functional Analysis of A-Type Ketoreductases from the Amphotericin Modular Polyketide Synthase. *Structure* 18(8): 913–922.

PHYRE2 Protein Fold Recognition Server.

PyMOL | [www.pymol.org](http://www.pymol.org)



## **Publications and Conferences**

---



## 7. Publications and Conferences

---

### 7.1 Scientific papers

Lázaro, B.; Villa, J.A.; Santín, O.; Cabezas, M.; Milagre, C.D.F.; de la Cruz, F. and Moncalián, G. (2017). Heterologous Expression of a Thermophilic Diacylglycerol Acyltransferase Triggers Triglyceride Accumulation in *Escherichia coli*. Plos One. (In Press)

Santín, O., Galié S., Moncalián, G. (2017) WS/DGAT acyltransferase from *Thermomonospora curvata*: Directed evolution and structural insights. (In preparation)

Santín, O, Moncalián, G. (2017). Structural and biochemical characterization of a KS-CLF di-domain from the omega-3 producer *Moritella marina* (In preparation)

Santín, O, Moncalián, G. (2017). Biochemical characterization of PfaA, a protein involved in the initiation of the DHA synthesis. (In preparation)

### 7.2. Scientific posters

Santín, O, Moncalián G., “Directed evolution of acyltransferases for the improvement of triglyceride production in *E. coli*” XXXVII Congreso de la SEBBM, Granada, (2014)

Santín, O, Moncalián G., “Structure of a ketosynthase-chain length factor of a polyunsaturated fatty acid synthase” XVI International conference on the Crystallization of biological macromolecules, Prague, (2016)\*

\*This work was awarded with the prize for the best scientific poster of the meeting.

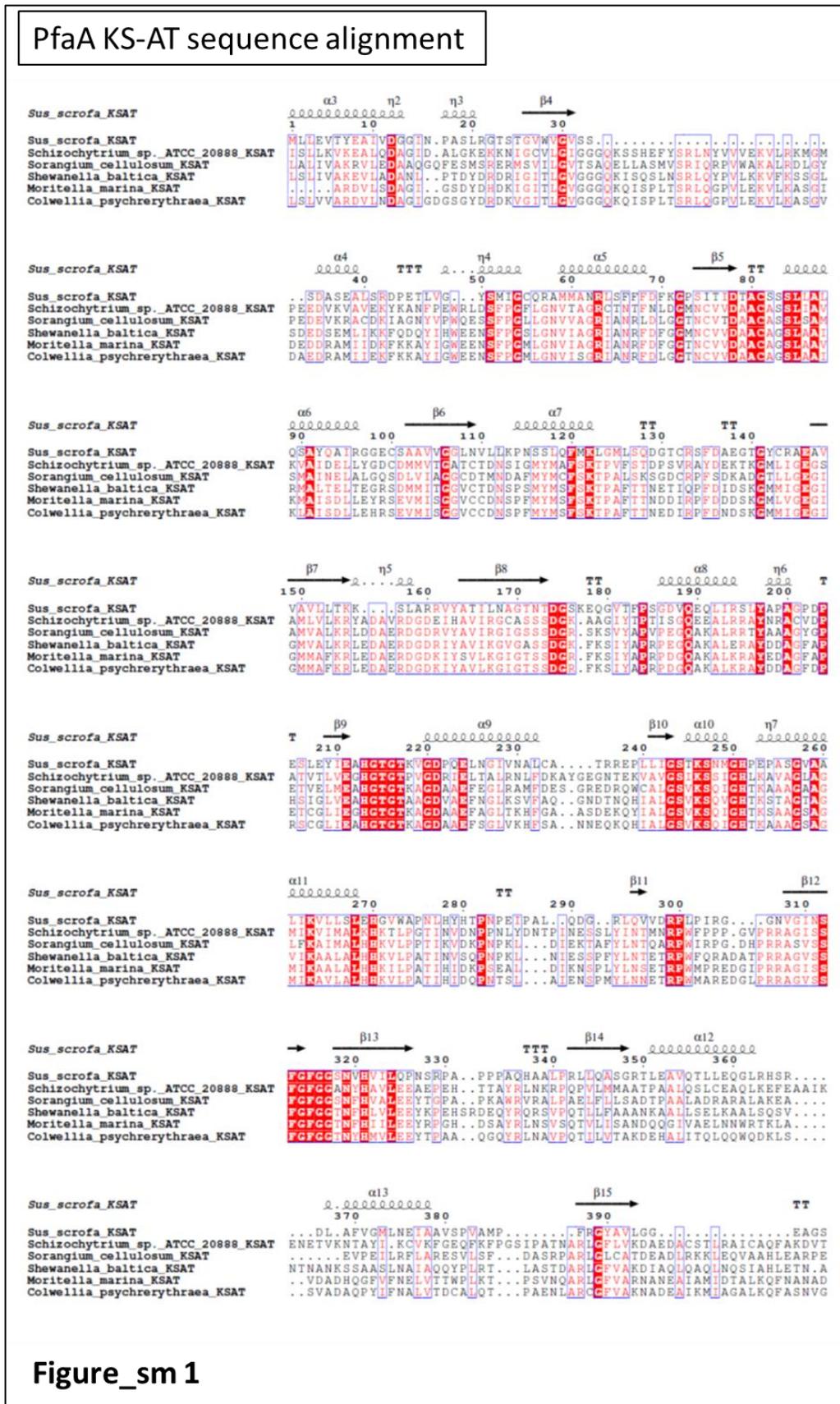


## **Supplementary Material**

---

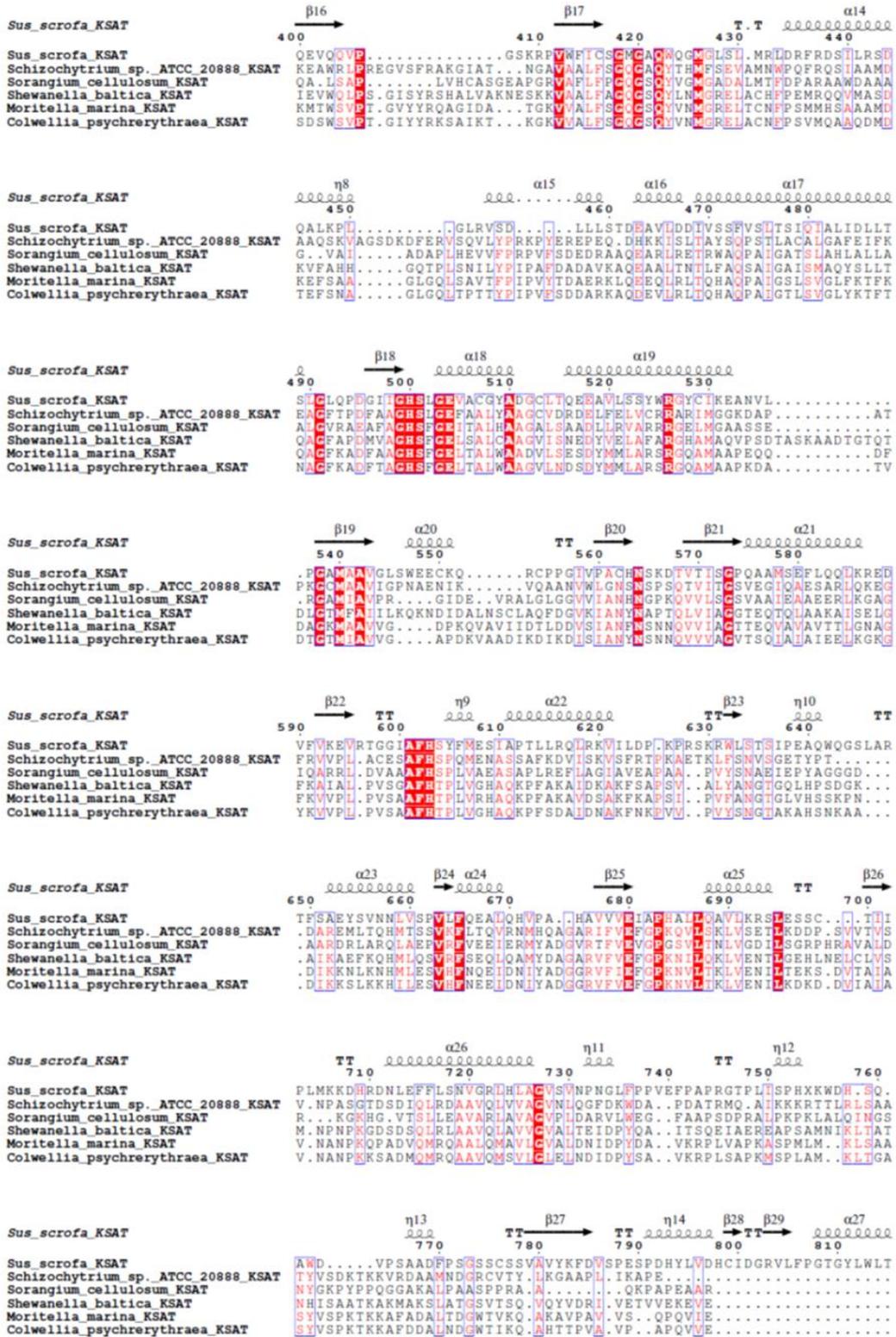


# 8. Supplementary Material



Figure\_sm 1

# PfaA KS-AT sequence alignment



Figure\_sm 1

## PfaA ACP sequence alignment

```

Mycobacterium_tuberculosis      α1      TTT TT      α2
                                10 20 30 40 50
Mycobacterium_tuberculosis    MPVTQEEI IAGIAETIEEVTGIEPSEITPEKSFVDDLDLDSLSMVEIAVQTEDEKY.GV.K
Moritella_marina               AT....IN KVMLEVVADRTGYPTDML ELSMDMEADLGIDSIKRVEILGAVQELIPDLPE
Moritella_marina_1             AH....IQ NVMLEVVADRTGYPTDML ELSMDMEADLGIDSIKRVEILGAVQEIITDLPE
Colwellia_sp._MT41             .....IQT VMLDVVADRTGYPT EML ELM DMEADLGIDSIKRVEILGAVQELIPDLPE
Colwellia_sp._MT41_1           .....IQ SVMTVVAERTGYPT EML ELAMDMEADLGIDSIKRVEILGAVQETIPDLPE
Sorangium_cellulosum           .....IER LVLDVVADRTGYPV SML GLQMELESDLGIDSIKRVEILSSVRDRTPGLSE
Sorangium_cellulosum_1         AE....VER VVMAVVAERTGYPA EML GLQMELESDLGIDSIKRVEILSAVRDRTPGLSE
Schizochytrium_sp._ATCC_20888 .....V VMEVLAARTGYETDMI ESDMELETELGIDSIKRVEILSEVQAML.NVEA
Schizochytrium_sp._ATCC_20888_1 .....V VMEVLAARTGYETDMI ESDMELETELGIDSIKRVEILSEVQAML.NVEA
  
```

```

Mycobacterium_tuberculosis      α3      α4
                                60 70
Mycobacterium_tuberculosis    TPD E D L A G L R T V C D V V A Y T Q
Moritella_marina              L N P E D L A E L R T I G E I V D Y M N
Moritella_marina_1            L N P E D L A E L R T I G E I V S Y M Q
Colwellia_sp._MT41            L N P E D L A E L R T I G E I V D Y M K
Colwellia_sp._MT41_1          L N P E D L A E L R T I G E I V S Y M Q
Sorangium_cellulosum           V D A S A L A Q L R T I G Q V V D H L R
Sorangium_cellulosum_1        V D A S A L A Q L R T I G Q V V D H L R
Schizochytrium_sp._ATCC_20888 K D V D A L S R T R T V G E V V D A M K
Schizochytrium_sp._ATCC_20888_1 K D V D A L S R T R T V G E V V D A M K
  
```

Figure\_sm 2



# PfaA DH sequence alignment

*Lyngbya\_majuscula* β1 β2

1 TTT → 10 20

*Lyngbya\_majuscula* AHP L L G E K I N LAGIEDQHRFQ .....

*Moritella\_marina* PQIL L V G N D L S KDASSDQKSDS...KSTAVKKPQVSRSLSDALVTKSI.KATNSSSLNKT

*Colwellia\_psychrerythraea* AQIL L V G N D L S KDAS...VKKPQVSRLLTTS.VSKTL.LAT.....

*Shewanella\_baltica* AQL L V G T D M Q GSAPHDDTPNEVQETEGSNLKKPEADLTTDASGPHALLNALNPSAVN..I

*Schizochytrium\_sp\_ATCC\_20888* .....

*Lyngbya\_majuscula* η1 β3 β4 α1 β5

30 40 50 60 70

*Lyngbya\_majuscula* .....SYIGAESPGY L N H H Q V F G K V L F P S T G Y L E I A A S A G K S L F T S Q .....EQVVVSDV

*Moritella\_marina* SALSDSSAFQVNNH F L A D H M I K G N Q V L P T V C A I A W M S D A A K A T Y S N R D C A L K Y V G F E D Y

*Colwellia\_psychrerythraea* .....NNAFLAD H T I G D D V L P T V C A I A W M S E A A T A A Y T G F .....HYQGLDNY

*Shewanella\_baltica* SAVKLRQRTLDPKAM I F I E D H C I N G N P V L P T V C A I Q W M R E A A F D V L K .....QPVKVQSY

*Schizochytrium\_sp\_ATCC\_20888* .....N P F L E D H V I Q G R R V L P M T L A I G S L A E T C L G L F P G Y .....S L W A I D D A

*Lyngbya\_majuscula* β6 β7 β8

80 90 100 110

*Lyngbya\_majuscula* D T L Q S L V I P E T E T K I V Q T .....V V S F A E .....N . N S Y K F E T F S . P . S E C E

*Moritella\_marina* K L F K G V V F D G N E A A D Y Q I .....Q L S P V T . R A S E Q D S E V R I A A K I F S L K . S G C K

*Colwellia\_psychrerythraea* K L F K G I I F D G N E A S D Y Q I .....D M N A E V . F . D D G T R L V V E T K I F S T N . A D G K

*Shewanella\_baltica* K L L K G I I F D A M T L E N G A L E N G A P I T L E L E L A P M A L T D K A A K D T D E Y L S G Q F S A L I S F C R

*Schizochytrium\_sp\_ATCC\_20888* Q L F K G V T V D G D V N C . E V .....T L T P S T . A P . . S G R V N V Q A T L K T . F . S S C K

*Lyngbya\_majuscula* β9 α2 β10 α3 β11

120 130 140 150 160

*Lyngbya\_majuscula* ..NQQT P Q W V L H A Q G K I Y T E P T R N S Q A K I . . D L E K Y Q A E C S Q A I E I . E E H Y R E Y R S K G I D

*Moritella\_marina* P V F H Y A A T I L L A . . . . . T O P L N A V K V E L P T L T E S V D S . . . . . N N K V I D E A Q A T Y S N G T L F

*Colwellia\_psychrerythraea* L I F H Y G A Q L T L V . . . . . N K L K A S T E I N I . . . . . V V P E L D E . . . . . K S G E C A A A T Y T N G T L F

*Shewanella\_baltica* . . P Q Y Q A I L V I D . . . . . D A P S D N L A T N S . . . . . K A T A F N A H S L A G L S A I T T A S S Y S D G T L F

*Schizochytrium\_sp\_ATCC\_20888* L V P A Y R A V I V I S . . . . . N Q G A P P A N A T M . . . . . Q P P S L D A . . . . . D P . A L Q G S Y Y D G K T L F

*Lyngbya\_majuscula* η2 β12 β13 η3 α4

170 180 190 200

*Lyngbya\_majuscula* Y C S S F Q C I K O L W K G Q G K A T G . E M A F P E E L T A Q L A D Y Q L . . . . . H P A L L D

*Moritella\_marina* H C E S L Q C I K O I L S C D D K G T L L A C Q I T D V A T A K Q G S F P L . . . . . A . D N N I F A N D

*Colwellia\_psychrerythraea* H C P S L H C I S E I L E C N E Q G T L L S C Q V P T V A S A K Q G D F T V P A L G N I R G N I I G K N Q S N I F A N D

*Shewanella\_baltica* H C P R L Q C I E S V V K F D G A S T V A R K V S I P H V A I A D C G S F V P N L . . . . . A P K G S Q A F A N D

*Schizochytrium\_sp\_ATCC\_20888* H C P A F R C I D D V L S C T K S C I V A R C S A V P G S D A R G E F A T . . . . . D T D A H D P F V N D

*Lyngbya\_majuscula* η4 β14 β15 TT

210 220 230 240 250

*Lyngbya\_majuscula* A A F Q I V S Y A I P H T E T D K I Y L P V G V E K F K L Y R Q T I S Q V . . . . . W A I A E I R Q T N I T A

*Moritella\_marina* L V Y Q A M L V W R K K O . F G L G S L P S V T T A W T V Y R E V V V D E V F Y L Q L N V V E H D L L G S R G S K A R C

*Colwellia\_psychrerythraea* I A Y Q A M L V W R K K O . L G L G S L P S S T Q S W T V Y R E V A V G R E R F Y L Q L N V V K S S G K G R G S L V A

*Shewanella\_baltica* L L L Q A M L V W A R L K . Y G A A S L P S S I G E F I S H A P P A F G D T G Y L V L E V V K H S G . . . . . R A L E A

*Schizochytrium\_sp\_ATCC\_20888* L A F Q A M L V W V R R T . L G Q A A L P N S I Q R I V Q H R P V P Q D K P F Y I T L R . . . . .

*Lyngbya\_majuscula* β16 β17

260 270 280

*Lyngbya\_majuscula* N I F L V D N Q G T V L V E I E G L R V K V T E P . . . . .

*Moritella\_marina* D I Q L I A A D M Q L L A E V K S A Q V S V S D I L N D M S

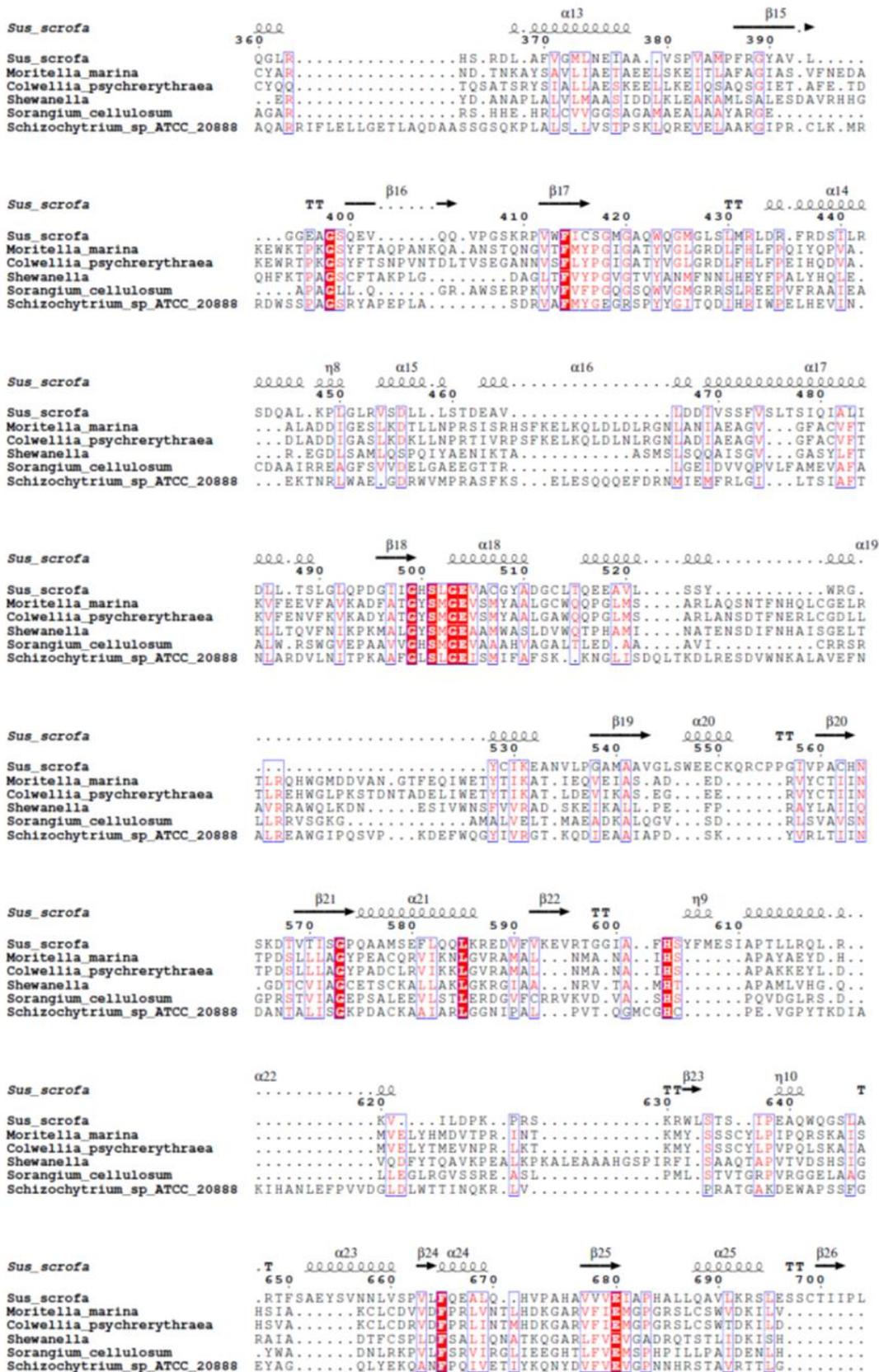
*Colwellia\_psychrerythraea* D I Q M I D E N M Q L L S E I K S A K V T A S A N L N D L .

*Shewanella\_baltica* N I A L Y H Q D G R L S C E M N N A K V T I S K N L N . . . . .

*Schizochytrium\_sp\_ATCC\_20888* . . . . .

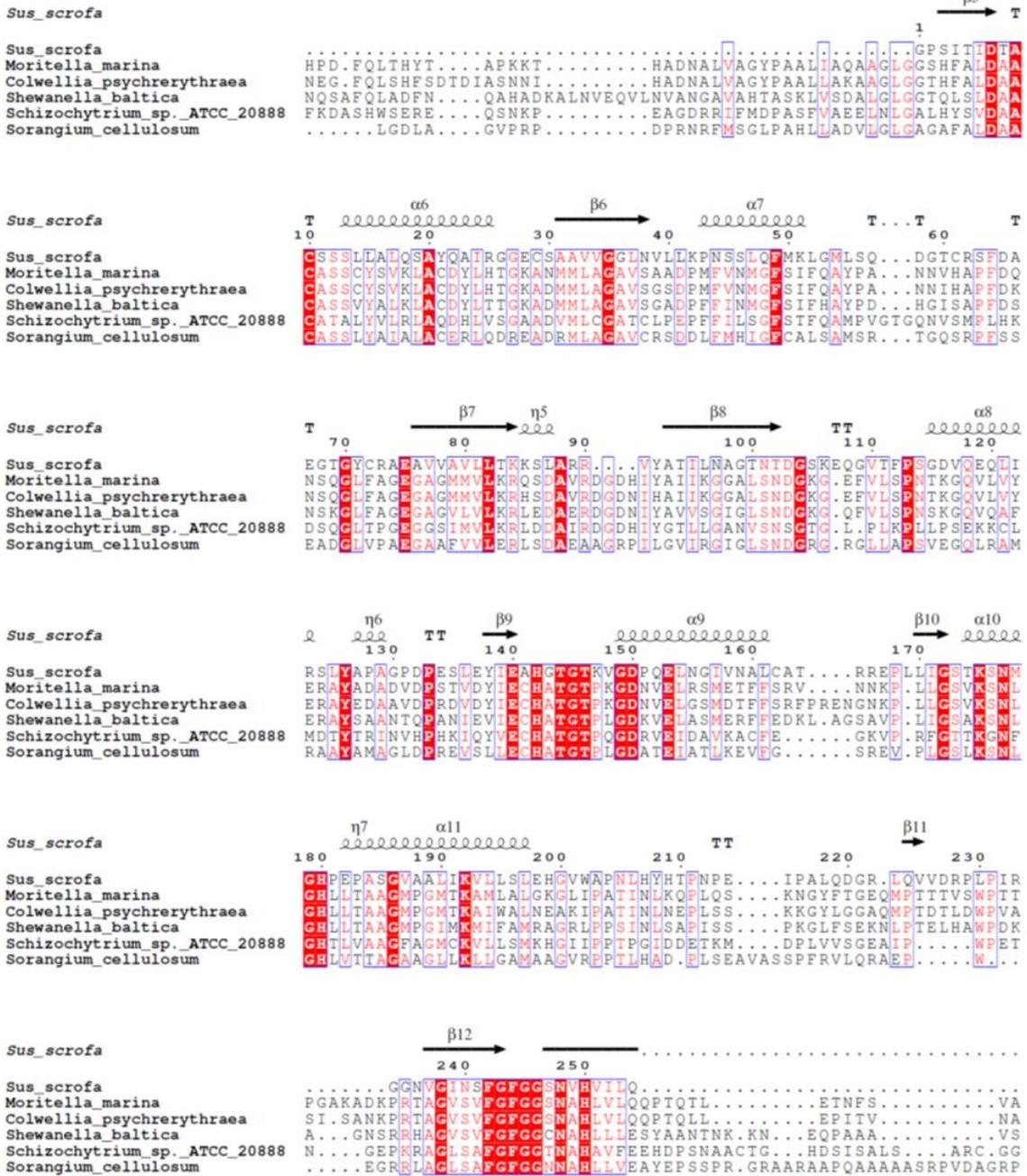
Figure\_sm 4

# PfaB sequence alignment



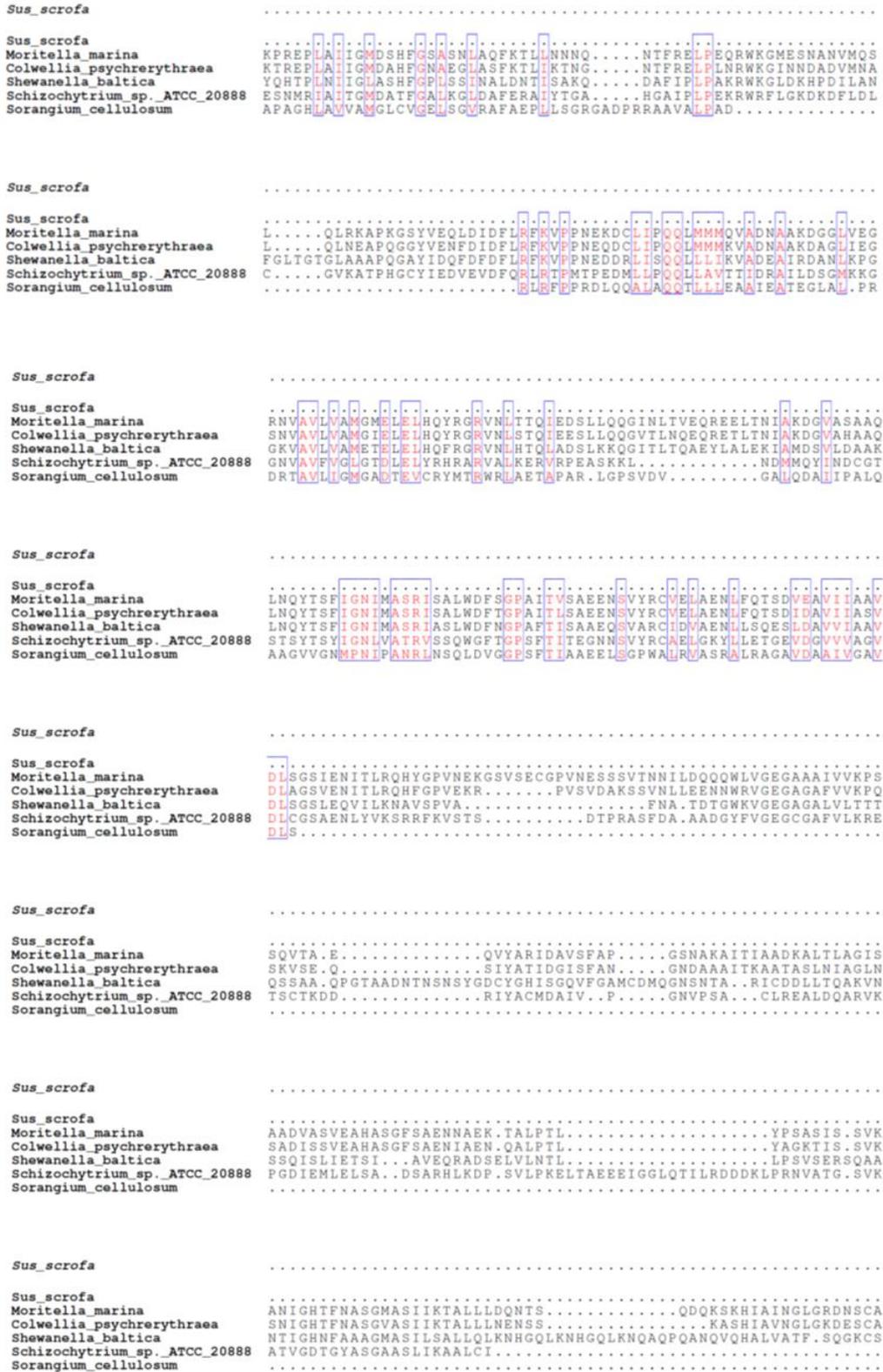
Figure\_sm 5

# PfaC KS-KS sequence alignment



Figure\_sm 6

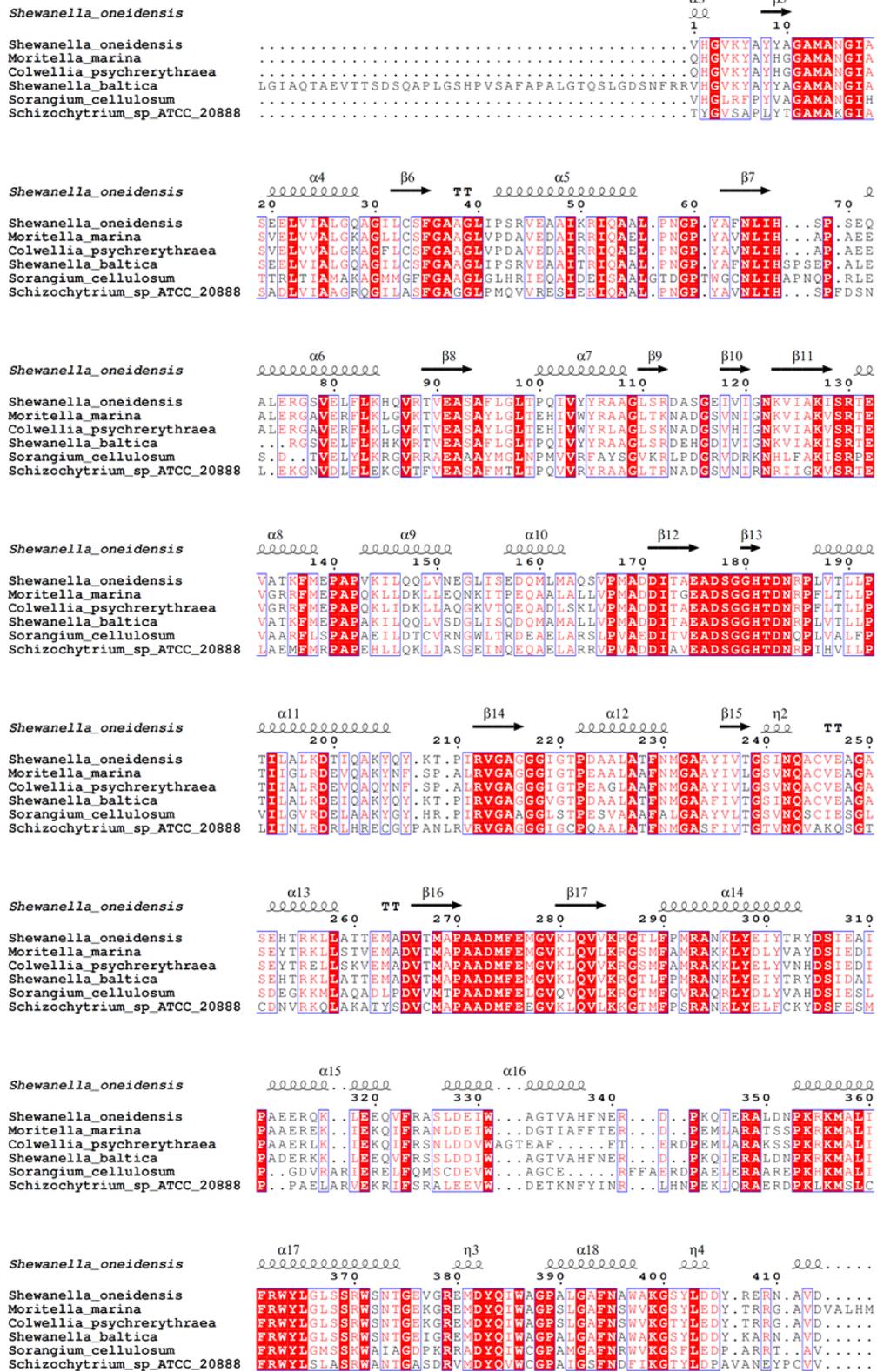
# PfaC KS-KS sequence alignment



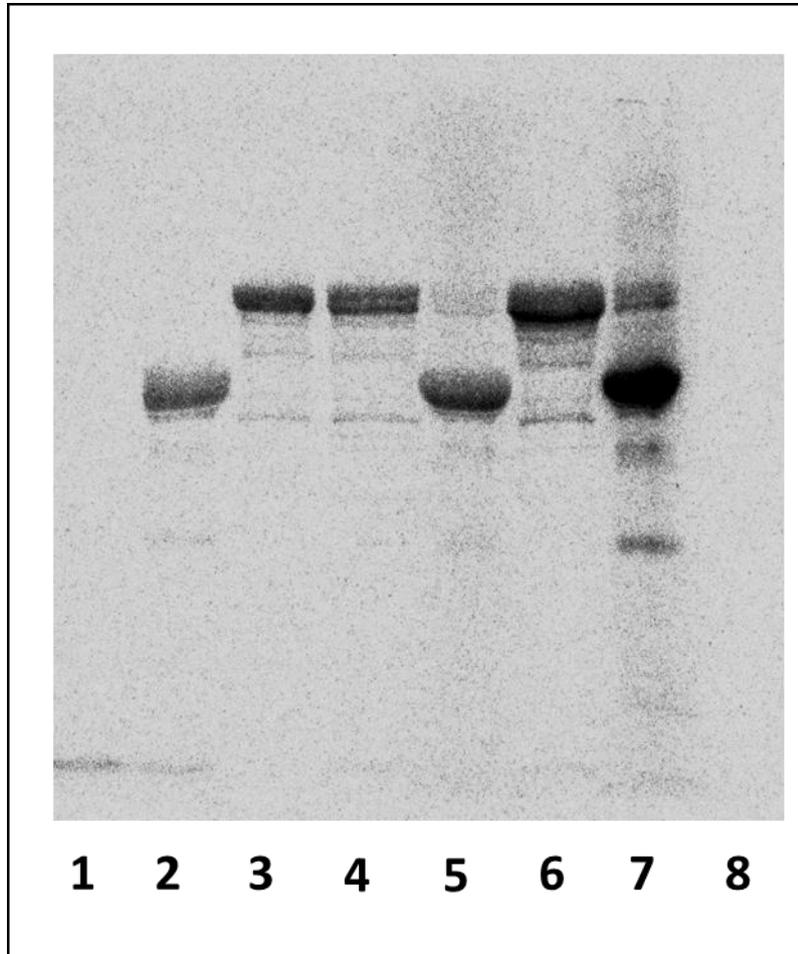
Figure\_sm 6



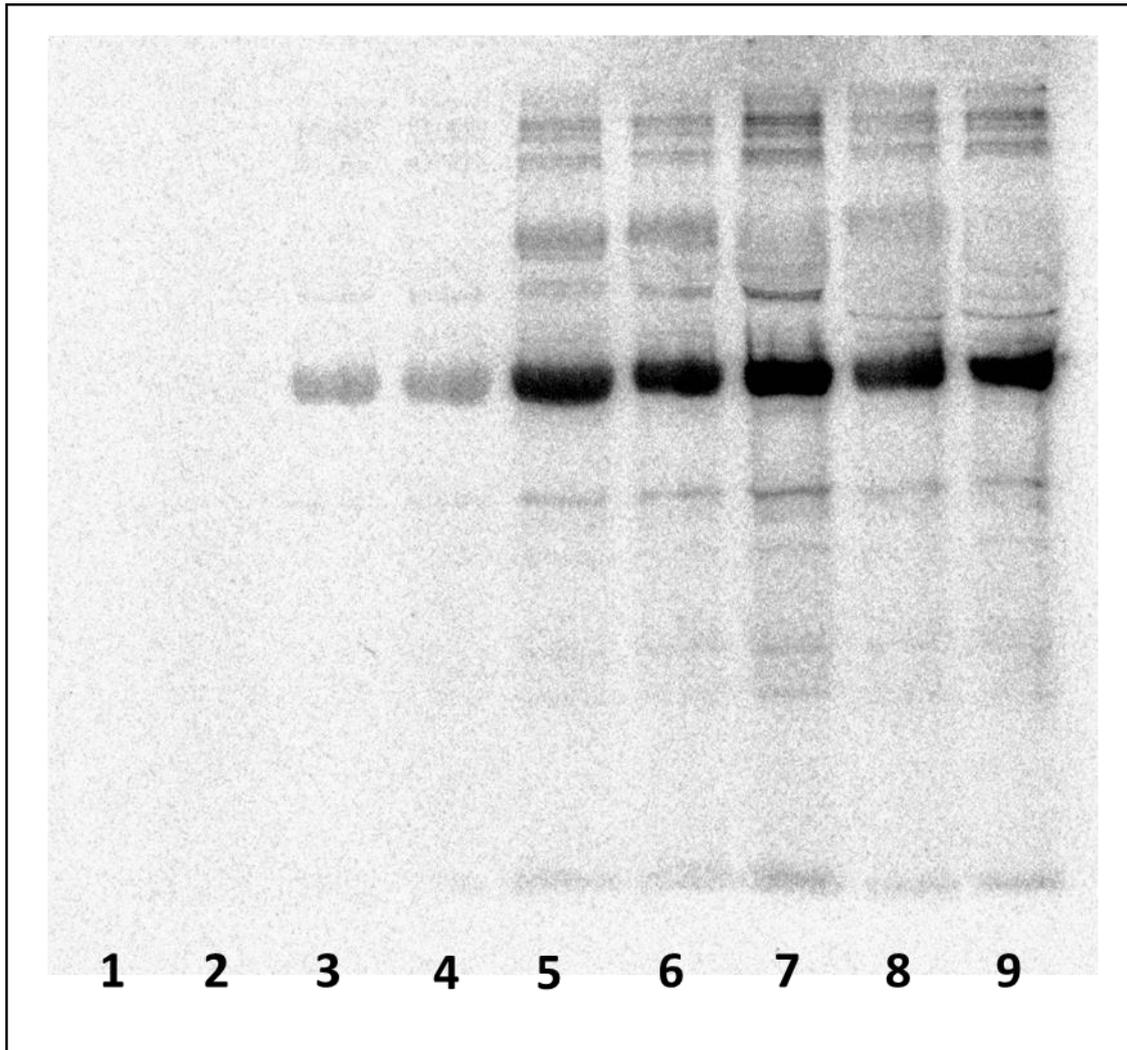
# PfaD sequence alignment



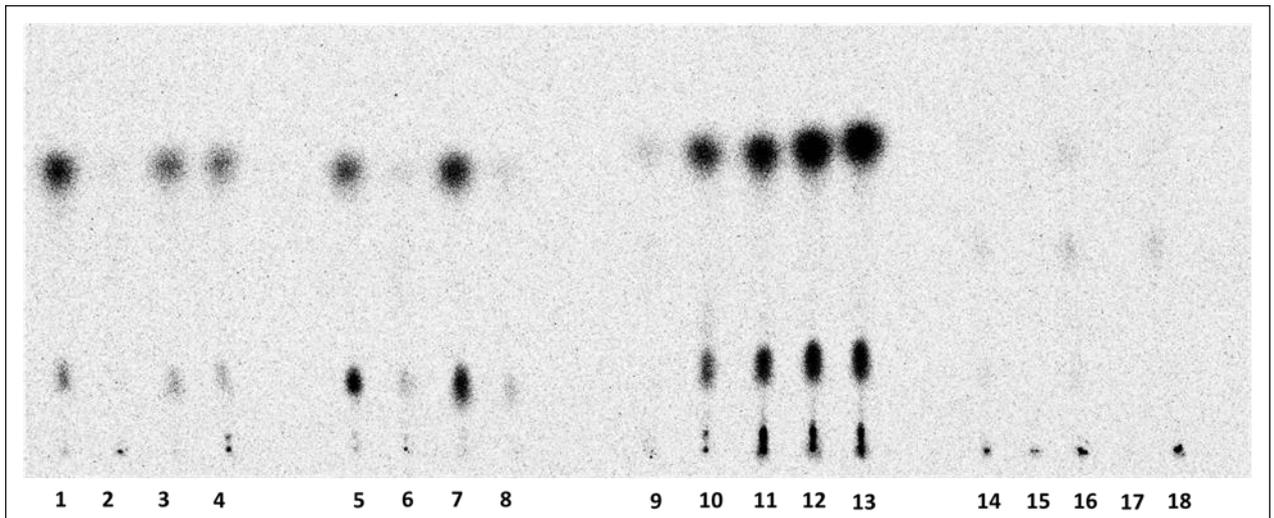
Figure\_sm 8



**Figure\_sm 9.** Binding enzyme assays of ACPs domains from *Colwellia* and ACPs, KSAT domains from *Moritella* using radiolabeled substrates. All samples were treated with malonyl-CoA. Line 1: apo-ACPcolwellia(c); 2: holo-ACPc; 3: KSATm; 4: apo-ACPmoritella(m)+KSAT; 5: holo-ACPc+KSAT; 6: apo-ACPmoritella(m)+KSAT; 6: holo-ACPc+ KSAT



**Figure\_sm 10.** Binding enzyme assays of multiple domains from *Moritella marina* using radiolabeled substrates. All samples were treated with malonyl-CoA(M) or Acetyl-CoA(A). Line 1: KS-CLF+M; 2: KS-CLF+holo-ACP+A; 3: KS-CLF+holo-ACP+M; 4: : KS-CLF+holo-ACP+A+M; 5: KSAT+holo-ACP+M; 6: KSAT+holo-ACP+KS-CLF+A+M; 7: KSAT+holo-ACP+KS-CLF+KR-DH+A+M; 8: KSAT+holo-ACP+KS-CLF+KR-DH+PfaB+A+M; 8: KSAT+holo-ACP+KS-CLF+KR-DH+PfaB+PfaC\_DH2M+A+M



**Figure\_sm 11.** Thin layer chromatography of multiple domains from *Moritella marina* using radiolabeled substrates. All samples were treated with malonyl-CoA(M) or Acetyl-CoA(A). An \* Indicates that the domain used carry a mutation in its active site. Line 1: KS\*AT+A; 2: KS\*AT+M; 3: KS\*AT+holo-ACP+A; 4: KS\*AT+holo-ACP+M; 5: KSAT\*+A; 6: KSAT\*+M; 7: KSAT\*+holo-ACP+A; 8: KSAT\*+holo-ACP+M; 9:KS-CLF+A+M; 10: KS-CL+KSAT+A+M; 11: KS-CLF+KSAT+holo-ACP+A+M; 12: KS-CLF+KSAT+holo-ACP+KR-DH+A+M+NADPH; 13: KS-CLF+KSAT+holo-ACP+KR-DH+pfaB+A+M+NADPH; 14: PfaB+A; 15: pfaB+M; 16: pfaB+holo-ACP+A; 17: PfaB+holo-ACP+M; 18: PfaB+holo-ACP+A+M





