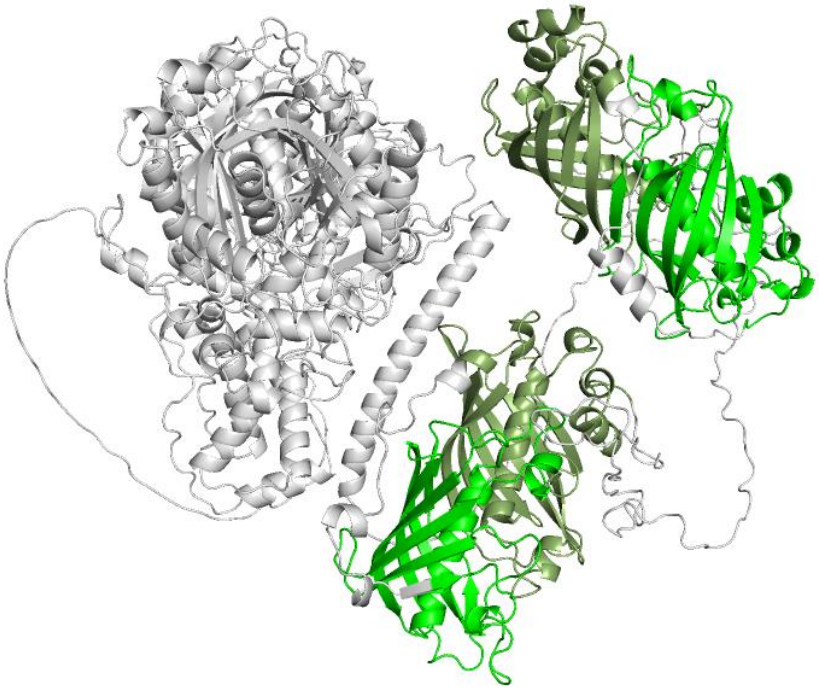
Structural characterization of PfaC: the key protein in PUFA synthesis



Paula Isla Gangoiti

Master's Degree in Molecular Biology and Biomedicine

Department of Molecular Biology, UC

Department of Microbiology and Genomics, IBBTEC

Academic year 2022-2023

Supervised by Dr. Gabriel Moncalián Montes and Dr. Maria Lucas Gay

Abstract

Polyunsaturated fatty acid synthases are large enzymatic complexes capable of producing polyunsaturated fatty acids (PUFAs) such as eicosapentaenoic acid (EPA) or docosahexaenoic acid (DHA). Both EPA and DHA are classified as omega-3 fatty acids and are essential for human health and nutrition. These PUFAs are naturally present in fish, seafood, including fish oils, as well as in other organisms like microalgae and proteobacteria from the deep sea. However, due to concerns regarding overpopulation and sustainability, there is a growing need to search for new alternatives for their production.

The marine bacteria *Moritella marin*a possesses an EPA and DHA producing cluster that consists of the *pfa* genes: *pfaA, pfaB, pfaC, pfaD* and *pfaE*. Therefore, deciphering the structure and function of these proteins is of great interest. In this work, we studied the PfaC protein, which consists of two ketosynthase-like domains and two dehydratases domains (DHs). We successfully expressed and purified the full-length *Moritella marin*a PfaC protein, as well as the C-terminal domain containing two dehydratases domains (DHs), using *E.coli* as the expression system. To accomplish this, we tested several overexpression conditions and purification tags. For the DHs domain, we conducted crystal screenings experiments.

To complement the experimental work, computer programs such as Alphafold2 were used to predict the protein structure of PfaC. This computational approach has allowed us to predict the positions of the active sites and understand the molecular mechanism of PfaC. We have gained insights into the specialized functions of the PfaC protein that presents domains preserved in other *pfa* genes but, in this case, is specialized first to elongate the chain above 18 carbon atoms and, secondly, to stabilize double bonds from trans to cis by isomerization.

Lo primero de todo, quiero agradecer a mis directores el Dr. Gabriel Moncalián Montes y la doctora Maria Jesús Lucas Gay, por permitirme formar parte de sus grupos de trabajo. Gabi, gracias por darme tanta autonomía. Las divagaciones e hipótesis que hemos hecho todos estos meses me han hecho mejorar mi pensamiento crítico y a intuir todas esas cosas que no vemos a simple vista. María, lo tuyo es de otro planeta, estás a todo y todo lo sacas adelante. Gracias por la paciencia cuando no entendía tus explicaciones y por estar siempre pendiente de cada paso que damos en el laboratorio. Gracias a ti, sabemos todos que, si en algún momento fallamos, tu vas a estar ahí con la mejor solución para sacarlo todo adelante.

También quiero agradecer a los miembros del labo 2.08 y actual, después de muchos meses de espera, 2.01; y alguna personita suelta de otros labos. Arancha, Antonio, Mariadel, Dani, Tjaša, Carlos, Miguel, Nahuel, Sara, Aurora, Iván, Vanessa… Si se me queda alguien, que se de por incluido que a estas alturas ya no tengo el cerebro para pensar más jajajajajaj. Muchas gracias a todos por estar ahí y alegrar los dias. La verdad es que es muy guay poder trabajar en lo que te gusta y además hacerlo rodeada de gente tan maravillosa como vosotros.

También agradecer al resto de IPs y miembros de grupo. Marta, Sheila, Mapi, Jorge, Ana, Santi, Matxalen, Elena, Iñaki, Rosa… Haber podido trabajar a vuestro alrededor ha sido todo un honor, y llevaré siempre conmigo vuestros consejos, tanto científicos, como sobre la vida.

Agradecer por la parte personal a dos grandes amigas que he hecho este año. Primero a Thalia, las tardes de tarta de queso, los guacamoles, y el simple hecho de estar ahí la una para la otra en los buenos y malos momentos, han hecho que te considere no solo una gran compañera de trabajo y estudios, si no una gran amiga. Nos separa un océano, literalmente, pero seguimos ahí la una para la otra. También quiero agradecer a otra personita especial, también del IBBTEC casualmente, Raquel. My gymsister favorita, las tardes de gimnasio contigo este año han sido lo mejor, gracias por ayudarme a ser mi mejor versión. Salir del laboratorio y entrenar se ha convertido en una rutina para mí, algo que llevaba intentando tiempo, y en parte es gracias a ti. A las dos os espera un futuro brillante como científicas y como mujeres, y espero estar ahí para veros avanzar.

Lo último, pero lo más importante de todo, quiero agradecer a mi familia. Esti, la mejor hermana mayor que cualquiera podría desear. Tu generosidad hacia los demás y sobre todo hacia nosotros es algo que jamás nadie podrá llegar a entender. Desde pequeñita me has impulsado y mira donde estamos las dos, lo hemos luchado mucho, pero mira lo que hemos conseguido… Y todo lo que nos queda todavía, hasta el infinito, pero siempre contigo de la mano. Papá. Mami. Qué os voy a decir que no sepáis ya. Os quiero con locura y me siento muy orgullosa de teneros como padres y de que me hayáis criado de la manera en la que lo habéis hecho, con valores y poniendo a la familia en el centro de todo. Me habéis apoyado en las malas y celebrado en las buenas, me habéis hecho resiliente. Vuestro apoyo es lo que me hace seguir hacia delante cada día. Papá, gracias por las canciones que nos envías por las mañanas, todos los dias, sin fallar uno, estas creando la banda musical de mi vida como un director de orquesta. Mamá, gracias por tu cariño, por tu delicadeza y por estar siempre pendiente de que no nos falte nada, siempre tienes un beso o una caricia para dar. Os quiero mucho a los tres (y a Lola también, jeje), ya lo sabéis.

Paula Isla Gangoiti

Index

Abbreviations

ACP – acyl carrier protein

ARA – arachidonic acid

AT – acyltransferase

CLF – chain length factor

Cm – chloramphenicol

DH – dehydratase

DHA – docosahexaenoic acid

DTT – dithiothreitol

EPA – eicosapentaenoic acid

ER – enoyl reductase

FAS – fatty acid synthases

Gm – gentamicin

GST – glutathione S-transferase

IPTG - isopropyl-β-D-thiogalactopyranoside

Kn – kanamycin

KR – ketoreductase

KS – ketosynthase

LC-PUFAs – long-chain polyunsaturated fatty acids

MAT – malonyl-CoA/acetyl-Co-A-ACP-transacetylase

MBP – maltose-binding protein

NCBI – National Center for Biotechnology Information

NIH – National Institutes of Health

NTA - nitrilotriacetic acid

ORFs – open reading frames

PCR – polymerase chain reaction

PDB – protein data bank

Pfa – polyunsaturated fatty acid

pLDDT – per-residue confidence score

PMSF – phenylmethylsulphonyl fluoride

Ppant – phosphopentetheinyl

PUFAs – polyunsaturated fatty acids

SDS-PAGE – sodium dodecyl-sulfate polyacrylamide gel electrophoresis

TRX - thioredoxin

Introduction

1. Chemical nature of long-chain polyunsaturated fatty acids

Long-chain polyunsaturated fatty acids (LC-PUFAs) are secondary metabolites that contain at least 18 or more carbons in length. In higher eukaryotes, they are essential membrane components contributing to fluidity, flexibility and selective permeability for its adaptation to specific conditions. These properties affect cell integrity and function. LC-PUFAs are the precursors to numerous lipid-derived signaling molecules that can be classified in two main families depending on the position of the first double bond closest to the methyl end known as ω, these families are ω-6 and ω-3 (Liu et al., 2023). ω-3 fatty acids are characterized by having their first double bond on the third carbon atom starting from the last (ω) methyl group. ω-6 fatty acids, on the other hand, have their first double bond located on the sixth carbon atom from the last carbon. Figure 1 shows canonical examples of ω-6 and ω-3 PUFAs.

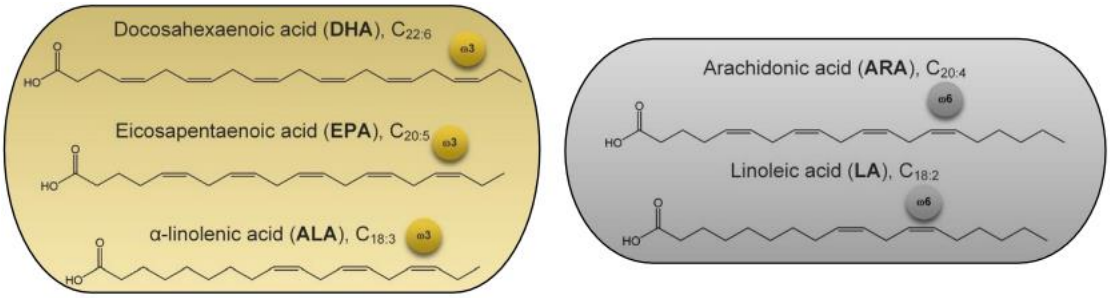


Figure 1. Canonical examples of omega-3 (ω-3) and omega-6 (ω-6) fatty acids. Extracted from Jovanovic et al., 2021.

Arachidonic acid (ARA), eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) belong to the category of LC-PUFAs.

2. Health impact and natural source of LC-PUFAs

LC-PUFAs are known to ease symptoms in patients suffering from chronic diseases and reduce the risk of developing them. DHA consumption is associated to antiarrhythmic and anti-inflammatory effects, modulation of autonomic function, decreased platelet aggregation, vasodilation, decreased blood pressure, plaque stabilization, decreased triacylglycerols and inhibition of synthesis of low-density lipoprotein (Lee et al., 2008).

LC-PUFAs also have beneficial effects on other diseases like asthma, arthritis, nephritis, lupus erythematosus and multiple sclerosis. It is noteworthy the important role of DHA in brain function and development due to its influence on neurotransmitter biosynthesis, signal transduction, uptake of serotonin, binding of β-adrenergic and serotonergic receptors, and monoamine oxidase activity (Abedy and Sahari, 2014).

For decades, the majority of LC-PUFAs have been extracted from fish and seafood, including fish oils. In the last years, microalgae have also been used to produce commercial LC-PUFAs. Nevertheless, there are other organisms such as proteobacteria from the deep sea that can produce big amounts of LC-

PUFAs but have not yet been optimized for large-scale industrial production. The increasing awareness of the health benefits associated with LC-PUFAs, such as EPA or DHA, combined with the global population growth, has led to a growing concern about insufficient production and environmental sustainability (PMID 35804698). Consequently, the production and purification of LC-PUFAs from bacterial sources presents an attractive alternative for meeting the growing demand.

The Food and Agriculture Organization (FAO) has raised their concerns around the environmental sustainability problem. Aquaculture is limited to human consumption even though it has been thought to be a possible strategy as an alternative to conventional fishing.

A solution that has been gaining importance in recent years has been the introduction of LC-PUFA synthetic pathways to bacteria, yeast and algae (Walsh et al., 2016). Bacteria from the genera *Colwellia*, *Moritella*, *Shewanella*, and *Photobacterium* are known to be capable of producing ω-3 fatty acids by using alternative synthetic pathways to synthetize LC-PUFAs from short substrates.

3. Overall mechanisms of fatty acid biosynthesis

The formation of a C4 fatty acid molecule through the condensation of two acetyl-CoA units is the first step of fatty acids biosynthesis. A series of rounds of carboxylation then ensues, incorporating the malonyl-CoA into the growing chain (Alberts and Vagelos, 1968).

Fatty acid synthases (FAS) can be classified into two distinct types called I and II (Figure 2). FAS type I proteins consist of large polypeptides with catalytic domains organized into single domains, forming huge megasynthases. On the other hand, FAS type II systems have each enzymatic activity organized in separated proteins, as is typically the case with bacterial FAS (Marrakchi, Zhang and Rock, 2002; Giordano et al., 2015).

Although the two FAS types are organized differently, they both have the same protein domains (KS, ketosynthase; AT, acyltransferase; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; ACP, acyl carrier protein), making the reactions needed to produce fatty acids quite similar. The main difference between the two FAS types is that FAS II proteins can produce a wide diversity of products needed for cellular metabolism, while FAS I systems are more specialized and can typically only produce palmitate, (Smith, Witkowski and Joshi, 2003).

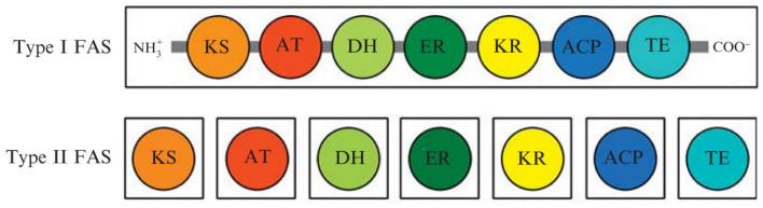


Figure 2. Domain organization of type I and type II FAS. Type I FAS proteins are large polypeptides, while Type II FAS are composed of numerous individual proteins. Extracted from Tsai and Ames, 2009.

The catalytic model for the synthesis of palmitate in *E.coli* is depicted in Figure 3 . The first step of the synthesis is the transfer of the acyl moiety of acetyl-CoA to the ACP domain catalyzed by malonyl-CoA/acetyl-Co-A-ACP-transacetylase (MAT). The next reaction in the cycle involves the NADPH-

dependent β-ketoacyl-ACP reductase, where the KS domain catalyzes the decarboxylative condensation of the acyl group with malonyl-ACP, forming a β-ketoacyl-ACP intermediate. Then, the KR domain processes the β-carbon, followed by the dehydration of β-hydroxyacyl-ACP to β-enoyl by the DH domain from FabA or FabZ enzymes. It is noteworthy that FabA also catalyzes the isomerization of trans-2-decenoyl-ACP to cis-3-decenoyl-ACP, introducing a cis double bond into the growing acyl chain. This is then reduced by a NADPH-dependent ER domain that forms a 4C acyl substrate. This substrate allows further elongation until reaching 16-18 carbon atoms. The final step of the cycle, catalyzed by the NADH-specific enoyl-ACP reductase, involves the release of the product from the ACP through the TE domain (Rock and Cronan, 1996; Maier, Jenni and Ban, 2006).

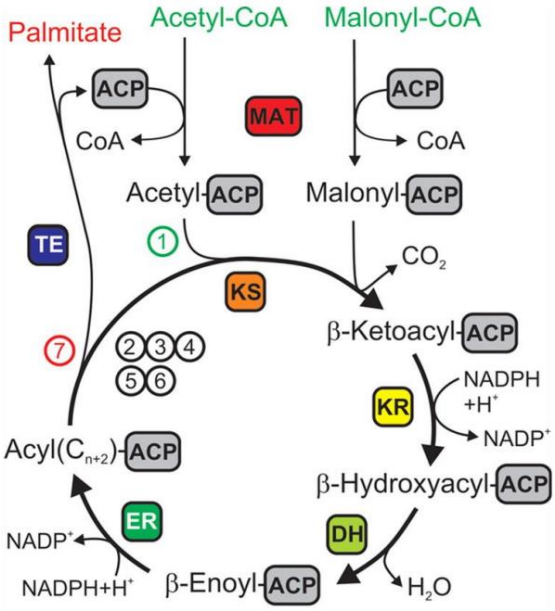


Figure 3. Catalytic cycle model for palmitate synthesis. Abbreviations: KS, ketosynthase; AT, acyltransferase; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; ACP, acyl carrier protein; MAT, malonyl-CoA/acetyl-CoA-ACP-transacylase. Extracted from Maier, Jenni and Ban, 2006.

4. Structural biology of fatty acid synthases (FAS)

Type I and II FAS differ in that type I present all domains in large multifunctional polypeptide, whereas type II FAS use monofunctional enzymes for fatty acid synthesis. The mechanisms for elongation and reduction are the same for both.

The structure shown in Figure 4 corresponds to Type I FAS, more specifically to the mammalian FAS. Its structure is formed by two identical homodimers forming a dimer, with the N-terminal catalytic domains (KS, MAT, DH) separated from those of the C-terminal (ER, KR, ACP) by a central region (Chirala et al., 2001).

In Figure 5, the domains that are part of the Type I mammalian FAS are represented separately. The KS domains (Figure 5A) are located at the bottom and have a structure very similar to the KS I (FabB) of *E. coli*. MAT domains (Figure 5B) are homologous to bacterial malonyl transferase (FabD). The DH domains (Figure 5C) are at the top of the FAS structure. They present a 'hot dog' folding very similar to

that of the bacterial dimeric dehydratases FabA and FabZ, as will be explained later. The ER domains, which are part of the medium-chain dehydrogenase family, (Figure 5D) are structurally more similar to a zinc-free quinone reductase. And finally, the KR domains, which are part of the short-chain dehydrogenase family, (Figure 5E) are found together with the ER domains in the FAS arms.

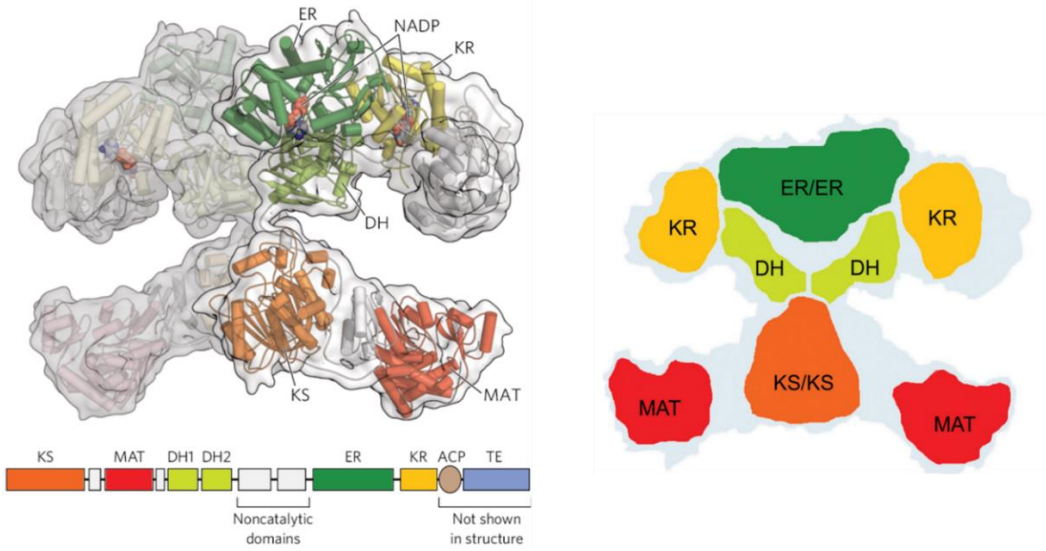


Figure 4. Structure of a fatty acid synthase Type I system. The domains are ketosynthase (KS), malonyl/acetyl-CoA–ACP transferase (MAT), dehydratase (DH), enoyl-reductase (ER), and keto-reductase (KR). Modified from Lehninger Principles of Biochemistry and Asturias 2006.

In addition, it should be noted that in the central part of the mammalian FAS there are about 650 residues that are assigned as 'core' or 'interdomain' to which no catalytic activity has been assigned and which are not part of any specific domain. These residues are believed to be involved in FAS dimerization (Maier, Jenni and Ban, 2006).

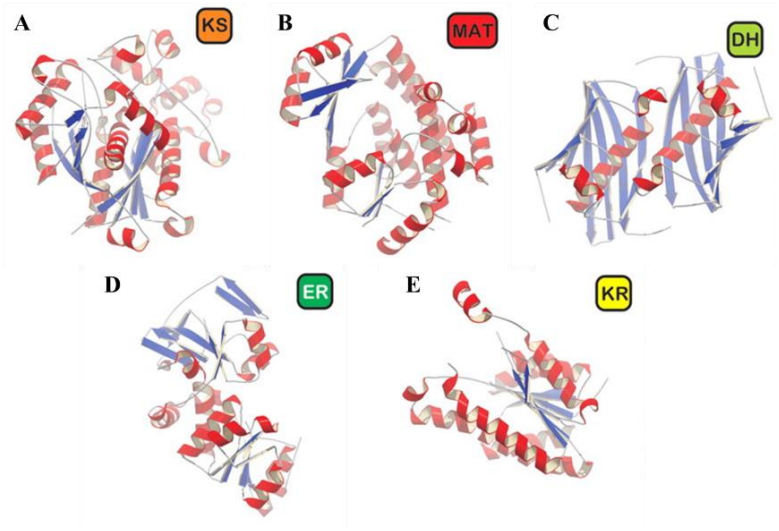


Figure 5. 3D structure of the separated domains that compose the mammalian FAS. **A**. β-ketoacyl-ACP synthase (KS). **B**. Malonyl/acetyl-CoA–ACP transferase (MAT). **C**. β-hydroxyacyl-ACP dehydratase (DH). **D**. Enoyl-ACP reductase (ER). **E**. β-ketoacyl-ACP reductase (KR). Modified from Maier, Jenni and Ban, 2006.

5. PUFA synthesis

Omega-3 fatty acids have been classified as secondary lipids because they are not involved in the classical metabolic pathways for LC-PUFA synthesis. These omega-3 LC-PUFAs contain multiple double bonds starting from the third carbon when counted from the methyl end of the molecule (Holub, 2002). The FAS system described above is responsible for LC-PUFA synthesis, but for the formation of the double bonds there are two different pathways: aerobic and anaerobic pathways.

5. 1. PUFA synthesis pathway

Desaturases and elongases are specialized enzymes that are present in plants and animals for LC-PUFAs synthesis. For the production of longer fatty acids (more than 18 carbon atoms), FAS synthases pathways are used to elongate the fatty acid chain and introduce more double bonds, thereby generating unsaturations (Figure 6).

The specific function of elongase enzymes is to incorporate malonyl units into the fatty acid chain and elongate the chain. At the same time, desaturase enzymes introduce double bonds into the acyl chain. These two enzymes are essential for the production of LC-PUFAs, such as EPA or DHA. However, mammalian cells lack the ability to produce Δ12 or Δ15 desaturases needed to produce the first molecules of the ω-6 or ω-3 pathways using oleic acid (C18:1) as substrate. Δ12 desaturates oleic acid to produce α-linoleic acid (C18:2, ω-6) and Δ15 desaturates α-linoleic acid to produce linoleic acid (C18:3, ω-3). This is why DHA and EPA are essential fatty acids in mammals as they need to be ingested in the diet or with supplements. Moreover, there is an unbalanced proportion of ω-6 and ω-3 fatty acids in humans due to the excessive consumption of ω-6 fatty acids. As observed in Figure 6, both ω-6 and ω-3 pathways share the same enzymes and thus, if there are too much ω-6 fatty acids in our diet, the common desaturases and elongases are mainly used in the ω-6 pathway producing a lack in ω-3 LC-PUFA like DHA or EPA.

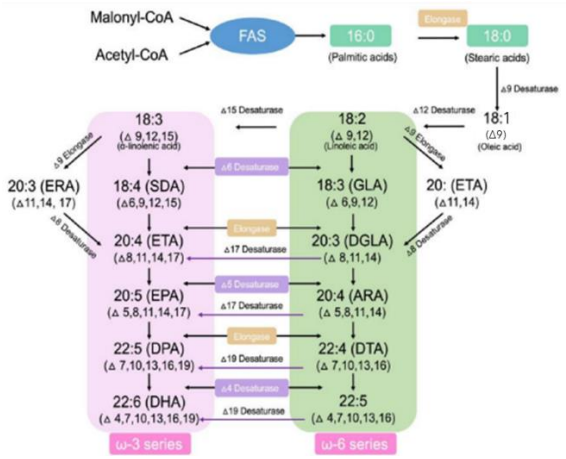


Figure 6. Aerobic pathway for desaturation and elongation. The FAS complex of enzymes uses acetyl-CoA and malonyl-CoA to begin the FA synthesis. Following that, the saturated FA stearic acid C18:0 is desaturated and elongated repetitively to form various LC-*PUFAs*. Extracted from Diao 2020.

This aerobic route employs desaturases as oxygenase enzymes to remove the hydrogen atoms from the acyl chain, using activated molecular oxygen.

5. 2. 'De novo' PUFA synthesis pathway

In nature, there is more than one pathway to synthesize LC-PUFAs. The anerobic pathway is capable of producing omega-3 FA from basic molecules. *Moritella marina* and other marine γ-proteobacteria have the ability to produce LC-PUFAs 'de novo' through dedicated PUFA megasynthases (Nichols et al., 1999). The gene clusters responsible for this synthesis codify for homologous enzymes involved in FA synthesis and consist of five open reading frames (ORFs) fundamental for *E. coli*'s production of EPA (Yawaza, 1996). Orisaka and his team discovered that *Moritella marina* has a gene cluster homologous to the one involved in the production of DHA when transformed in *E.coli* (Orisaka et al., 2006).

5. 1. 1. PUFA gene clusters

The cluster of genes responsible for the production of LC-PUFAs, known as *pfa* (polyunsaturated fatty acid) gene clusters, have been identified in different species of marine bacteria from the genera *Colwellia*, *Moritella* and *Shewanella*, and in a less abundant distribution in *Alteromonas* and *Photobacterium*. Figure 7A illustrates a generic organization of the five *pfa* genes, while Figure 7B shows the wide diversity of *pfa* clusters in different bacteria. Figure 5C presents the specific organization of the *pfa* cluster in *Moritella marina*.

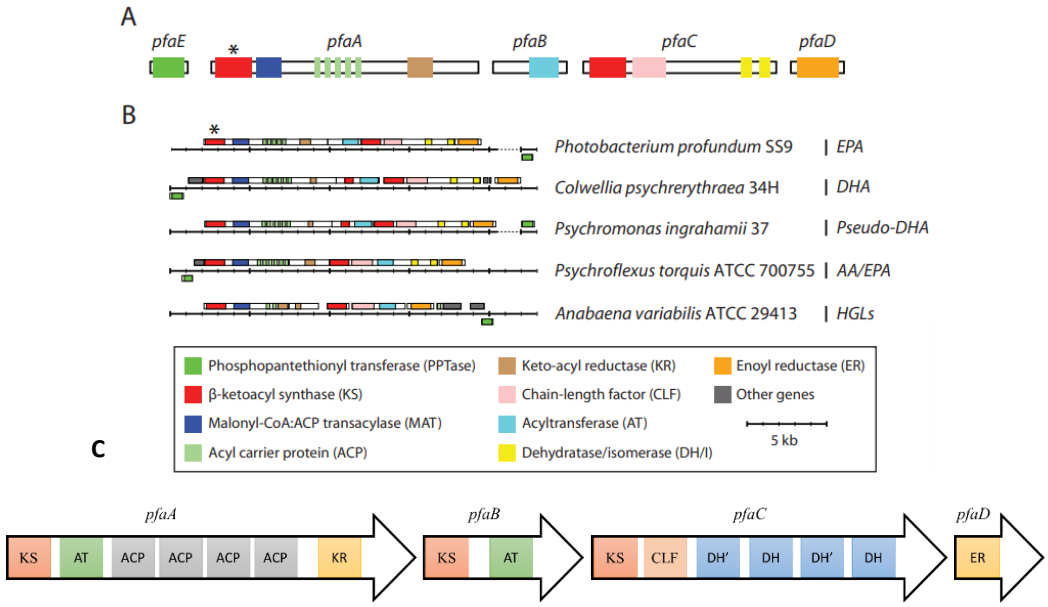


Figure 7. Diversity of *pfa*-like gene clusters. **A**. Organization of the five gene cluster, *pfaEABCD*, that participate in the synthesis of LC-PUFAs. **B**. Enzymatic domain architecture within *pfaEABCD* from representative organisms that produce EPA, DHA, ARA and the aglycone moiety of heterocyst glycolipids (HGLs). Extracted from Shulse and Allen, 2011. **C**. *pfa* gene cluster from *Moritella marina*. Modified from Gemperlein et al. 2014.

The iterative biosynthesis of LC-PUFAs involves several biochemical reactions that are determined by the functional domains of the *pfa* gene cluster, which share similarity to FAS and PKS systems. Nevertheless, researchers have not yet been able to elucidate the specific enzymatic reactions responsible for EPA or DHA production (Yoshida et al. 2016).

Over the last few years, efforts have been made to decipher and analyze the protein structure within the *pfa* cluster through various techniques, including X-ray crystallography. However, the crystallization process for these proteins presents significant difficulties and challenges. Using various computer programs, we have been possible to make predictions of the possible structural models of some of these proteins, though, further work remains to be done to fully understand the functionality of all these structures.

6. Characteristics of *pfa* gene cluster from *Moritella marina*

The pfa gene cluster has been previously described by Okuyama et al., 2007. This cluster contains five open reading frames called *pfaA, pfaB, pfaC, pfaD*, and *pfaE*. In the case of the bacterial species *Moritella marina*, the cluster gene only includes *pfaABCD* and lacks *pfaE*. These genes were the first to be cloned from DHA-producing bacteria.

This article brings out that *pfaA* encodes a multifunctional protein that includes domains 3-keto-acyl synthase (KS), malonyl co-A:acyl carrier protein (ACP) acyltransferase, five or six ACP repeats, and 3-ketoacyl–ACP reductase (KR). The *pfaC* gene encodes for two KS domains and two or three DH domains. The second domain of KS of *Moritella marina* was considered as chain length factor. As verified later in this work, they observed a structural homology between the DH of PfaC and FabA/FabZ. Finally, the *fabB* and *fabD* genes encode proteins with KS, AT, and ER domains. However, they found that the KS domain of PfaB in *Moritella marina* lacked an active site sequence. The exact mechanism for the production of DHA is not described yet. However, it is speculated that there may be some interaction between the domains of different Pfa performing the function together, rather than as individual proteins.

6. 1. Experimental structural biology of Pfa proteins

Knowledge of the structural organization and folding of the Pfa proteins responsible for the biosynthesis of fatty acids in marine bacteria is essential to be able to optimize the synthesis mechanism of these fatty acids in the laboratory. So far, few publications have conducted an in-depth structural analysis. Only the structure of the PfaD of *Shewanella piezotolerans*, which acts as a trans-acting enoyl reductase (ER), and the KS domain of the PfaC of *Moritella marina* are known to date.

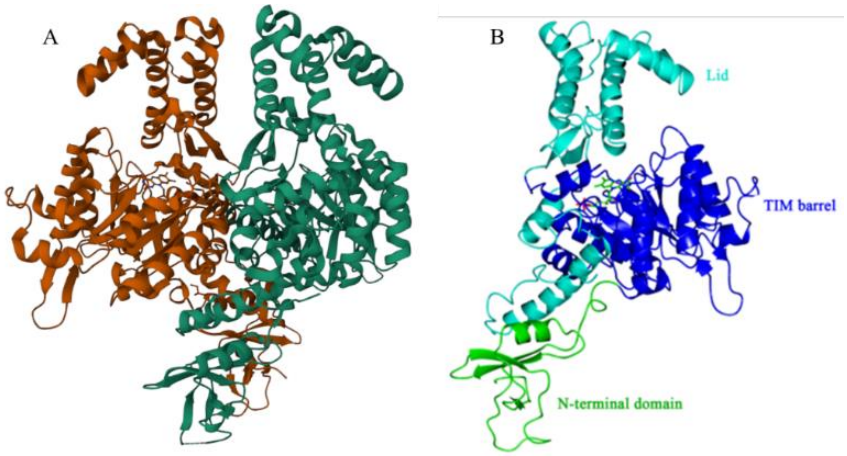


Figure 8. Crystal structure of PfaD. **A**. PfaD homodimer . **B**. Domain prediction for each monomer of PfaD. Modified from Zhang et al., 2021; PDB entry: 6LKZ.

PfaD presents a homodimer structure in which each subunit is consisted of three domains: the N-terminus has one α-helix and five β-sheets; the central domain has the classic triose phosphate isomerase barrel (TIM barrel) composed of eight stranded parallel β-sheets surrounded by eight α-helices; lastly, in the C-terminal domain it constitutes a lid above the TIM barrel (Zhang et al., 2021; Figure 8).

As for the structure of the KS domain of PfaC, it is composed of two domains tightly packed along a 2-fold axis. Like other KS domains involved in FA and polyketide biosynthesis, it is composed of two β-sheets sandwiched between a-helices. The two domains have a sequence identity of 27%, with practically identical structures; the cores of the β-sheets of both domains are identical, with slight differences around the α-helices (Santin et al., 2020; Figure 9). PfaC is the second largest *pfa* cluster and has multiple domains (Figure 9). It contains two KS domains (aminoacids 2-455 and 467-883) and four dehydratase domains (aminoacids 1152-1331, 1358-1546, 1630-1814 and 1845-2010).

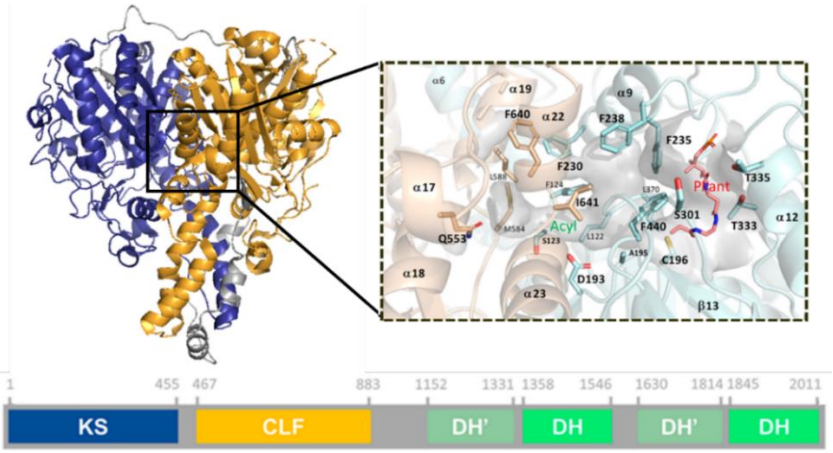


Figure 9. Crystal structure of KS domain and domain prediction for PfaC. Modified from Santin et al., 2020; PDB entry 6RIW.

This work will be focused on the structural study of the rest of the domains present in PfaC, the dehydratase domains, as well as the structure of the entire PfaC protein (KS-CLF didomain and DH domains) and the possible interactions between the domains or the subunits within the same domain.

6. 2. The condensing domains of PfaC: KS and CLF domains

Condensing domains have the capability to conduct condensation reactions independently. Within this classification are included the domains KS and CLF (KS-like domain). The first, which is included in Figure 10, displaces the sequence alignment of these domains, including the two KS domains that exhibit homology with known LC-PUFA synthases (Santin et al., 2020).

In previous investigations (Santin et al., 2020) the active sites of the KS-CLF didomain were predicted through comparison with known sequences. Notably, the conserved triad C196-H331-H368, found in other synthases, is responsible for catalyzing the hallmark decarboxylative condensation reaction of all KS domains. Also noteworthy are S301-T333/335 residues, as they enable the interaction with the phosphopentetheinyl (Ppant) arm of CoA, which is attached to the acyl carrier protein domains. Additionally, the conserved Q553 residue forms the floor of the substrate pocket.

In the KS-CLF didomain, there are additional residues that are part of the active sites and are not commonly found in other synthases. One such residue is K295, which is the primary docking site for the electronegative acyl carrier protein domains. Another set of residues, including L122-F124-F230-F235-F238-L588-F640-I641-I644, allows the formation of the mKS-CLF lipophilic tunnel.

A characteristic feature of the KS-CLF didomain is the presence of a helical bundle located at the top, which contributes to the tightly packed structure of this didomain. Furthermore, this helical bundle may interact with other components of the PFAS complexes to facilitate the strong association of the KS and CLF domains (Santin et al., 2020).

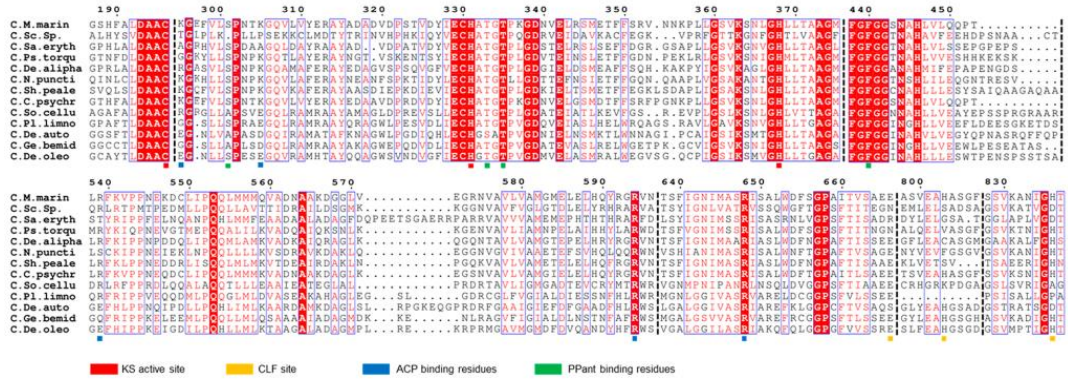


Figure 10. Alignment of KS-CLF didomains from known LC-PUFA synthases. The residues thar are identically conserved are colored in white with a red background, while other residues with high rate of conservation are colored in red and framed in blue. Extracted from Santin et al., 2020.

6. 3. The modifying domains of PfaC: DH domains

The dehydratase domains are responsible for conducting the fatty acid elongation step. In some cases, it may be followed by cis isomerization of the double bond (*fabA* gene). The dehydratase domains of PKS have been studied less compared to other domains. Two different genes *fabA* and *fabZ* encode for the DH domains in the Type II FAS system of *E. coli*. FabZ metabolizes saturated and unsaturated fatty acids and FabA is more efficient at processing 10-carbon acyl chains to introduce unsaturations.

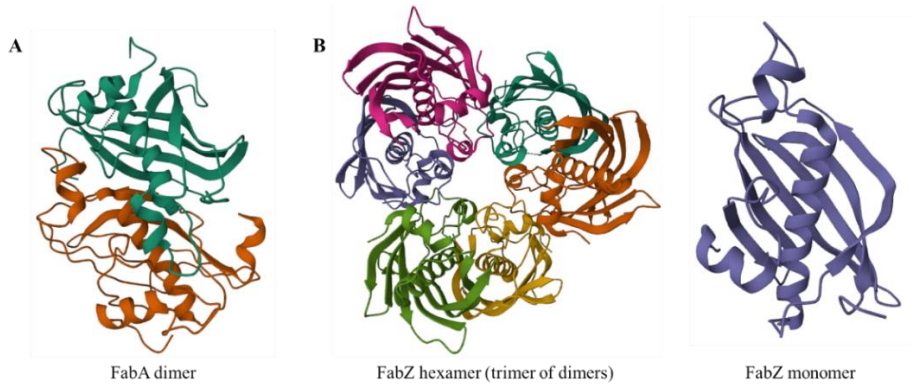


Figure 11. **A**. Structure of 3-hydroxydecanoyl-Acyl Carrier Protein Dehydratase (FabA) from Pseudomonas aeruginosa; Moynié et al., 2016; PDB entry: 4CL6. **B**. Structure of (3R)-hydroxyacyl-ACP dehydratase (FabZ) from *Pseudomonas aeruginosa*; Kimber et al., 2004; PDB entry: 1U1Z.

The structure of FabA (Figure 11A) and FabZ (Figure 11B) is based on a 'hot dog' folding, in which each monomer has a long central α-helix and several antiparallel β-sheets. In FabA, the dimers are symmetric, with the two independent active sites located between the two subunits of the enzyme. This folding is characteristic of proteins associated with fatty acid biosynthesis and metabolism. In FabZ, the monomers from the three dimers are also symmetric.

His70 acts as a catalytic base to abstract a proton from C2 of the substrate, D84 promotes the removal of the hydroxyl group to generate the enoyl-ACP product. Substrate trans-2-decenoyl-ACP is then isomerized to cis-3-decenoyl-ACP with histidine adding a proton to C2 and aspartic abstracting a proton from C4. The tunneling structure of the active site is arranged by hydrogen bonding interactions with aspartic acid and amides of C80 and G79 (Figure 12, White et al., 2005).

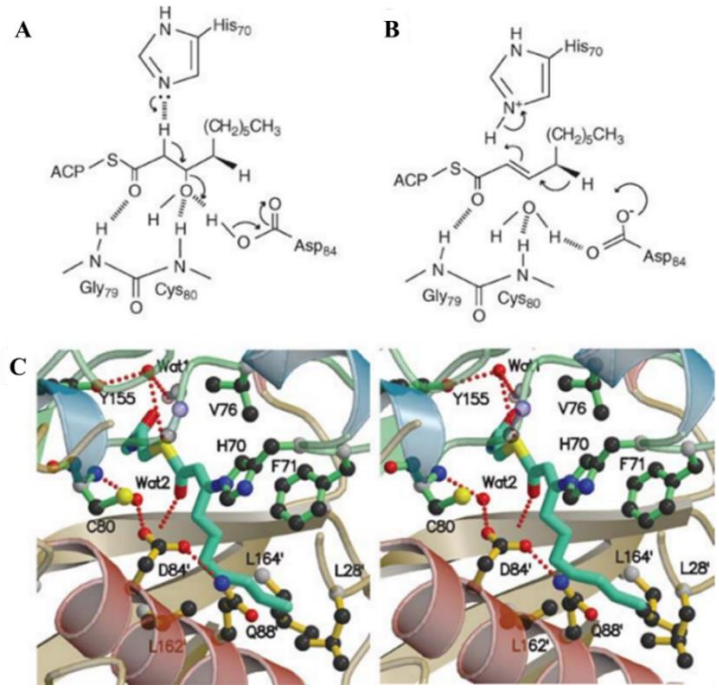


Figure 12. FabA mechanism. Extracted from White et al., 2005.

In the case of PfaC, it contains two DH'-DH didomains in which the first of the modules lacks an active site. These domains are responsible for controlling the regiochemistry in the formation of double bonds, ensuring the correct positioning of the double bonds in the fatty acid chain.

The DH domains catalyze dehydrations which, depending on the length of the carbon chain, are followed by hydrogenation of the double bond through the ER domain or give rise to cis bonds by isomerization of the α,β-unsaturated ACP-bound intermediate (Santin et al., 2020).

In this work, the structure and function of these PfaC DH domains will be further analyzed to gain a better understanding of their role in polyunsaturated fatty acid synthesis.

Objectives

Structural biology and protein engineering allow us to improve the production and productivity of enzymes with key functions in multiple biological systems, such as fatty acid synthesis. The objective of this Master's thesis has been the characterization of PfaC from *Moritella marina*, the key protein in the synthesis of long-chain PUFAs. This aim has been addressed by the following approaches:

- Structural characterization with computational modeling software *AlphaFold2* of full-length PfaC as well as of the individual DH domain.
- Confirmation of PfaC's dehydratase function by comparing its 3D model with previously described structures of proteins known to possess dehydratase activity.
- Improvement of PfaC production in *Moritella marina* using *E.coli* as an expression system.
- Enhancement of PfaC's solubility and stability to facilitate higher-yield purification.
- Determination of PfaC 3D structure through x-ray crystallography.

Materials and Methods

1. Bacterial strains

The bacterial strains used to express the protein PfaC were the following: *E. coli BL21(DE3)*, which has inserted in the chromosome the sequence for rRNA polymerase of the T7 phage under a promoter regulated by isopropyl-β-D-thiogalactopyranoside (IPTG); *E.coli C41(DE3)*, which presents increased membrane invaginations resulting in an expanded membrane surface; *E. coli Rosetta(DE3)*, to enhance the expression of eukaryotic proteins that contain rare codons in *E. coli*; and *E. coli ArcticExpress(DE3)*, optimized for low temperatures. The features for each strain are detailed in Table 1.

Table 1. Strains from *E. coli* used in this work.

Strain	Genotype	Source
BL21(DE3)	F- *ompT hsdS*$_B$ (r_B- m_B-) *gal dcm* (DE3)	Invitrogen
C41(DE3)	F – *ompT hsdS*$_B$ (r_B- m_B-) *gal dcm* (DE3)	Sigma-Aldrich
Rosetta(DE3)	F-*ompT hsdS*$_B$(r_B- m_B-) *gal dcm* (DE3) pRARE (Cam^R)	Sigma-Aldrich
ArcticExpress(DE3)	F– *ompT hsdS*(r_B – m_B –) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* Hte [*cpn10 cpn60* Gent^r]	Agilent
Top10	F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lacZ*ΔM15 Δ*lac*X74 nupG *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*E15 *gal*K16 *rps*L(StrR) *end*A1 λ-	Invitrogen

All the DE3 strains are lysogens of λDE3, carrying a chromosomal copy of T7 RNA polymerase gene under the control of the lacUV5 promoter. These strains are able to produce proteins from target genes cloned in pET vectors upon induction by IPTG.

2. Plasmids

The plasmids used in this work are listed in Table 2.

Table 2. Plasmids used in this work.

Name	Description	Promoter/Inductor	Size (bp)	Resistance	Ref.
pMLG31	pIA-GST-TEV	T7/IPTG	6661	Kn	Maria Lucas' laboratory
pMLG32	pIA-HisMBP-TEV	T7/IPTG	7247	Kn	
pMLG33	pIA-HisTRX-TEV	T7/IPTG	6404	Kn	
pET29c	pET Bacterial Expression Vector	T7/IPTG	5372	Kn	Novagen

3. Oligonucleotides

The oligonucleotides used for the construction of the plasmids are listed in Table 3. All oligonucleotides were purchased from Thermo Fisher Scientific.

Table 3. List of oligonucleotides. F: Forward/Start; R: Reverse/End.

Name	Description	Sequence (5'>3')
O-PI-1	PfaC-pIA-F	GCGAAAACCTGTACTTCCAGGGCATGGAAAATATTG CAGTAGTAGGTATTGC
O-PI-2	PfaC-pIA-R	GCTTTGTTAGCAGCCGGATCTCATTACGCTTCAA CAATACTTAAAACGATG
O-PI-3	PfaC-1112-pIA-F	GCGAAAACCTGTACTTCCAGGGCACAAACCTAA CCAGTACAGAAGC
O-PI-4	PfaC-seq1-F	GTAAGGCGACCGCTGTTCAAATGG
O-PI-5	PfaC-seq2-F	ATGCCGATGCAGATGTTGACCC
O-PI-6	PfaC-seq3-F	GTTGCGGTATTAGTAGCGATGGGC
O-PI-7	PfaC-seq4-F	GCAAGCGCATCAAGTTGCACC
O-PI-8	PfaC-seq5-F	GCTGATTTGGTTGAATTCGCTGAAGG
O-PI-9	PfaC-seq6-F	GCTGGTTCGTTGATGTCGGAAGG
O40	Seq-GST-F	GTATATAGCATGGCCTTTGCAGGGCTGGC
O94	pIA-TEV-iso-R	GCCCTGGAAGTACAGGTTTTCGCCGCTGC
O109	Seq-pIA-F	GATCTTCCCCATCGGTGATGTCGGCG
O110	Seq-pIA-R	GGGTTATGCTAGTTATTGCTCAGCGGTGGC
O161	Seq-T7-R	CTAGTTATTGCTCAGCGGTGGC
O206	Seq-MBP-2	CAACCTGCAAGAACCGTACTTCACCTGG
O284	pIA-isov2-F	TGAGATCCGGCTGCTAACAAAGCCCGAAAGG

4. Microbiological methods

4. 1. General culture conditions

The medium used for culturing E. coli was Luria-Bertani (LB) medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl). For bacterial growth in solid medium plates, LB was supplemented with 1.5% (w/v) bacterial agar. For the growth of bacterial strains with plasmids containing antibiotic resistance genes, the medium was supplemented with specific antibiotics: Kanamycin (Kn) at 50 µg/mL, Gentamicin (Gm)

at 20 µg/mL and Chloramphenicol (Cm) at 25 µg/mL. Cultures were incubated at 37ºC and 180 rpm (overnight).

4. 2. Bacterial glycerol stock

Cell culture stocks were prepared mixing 750 µL of the overnight culture with 750 µL of glycerol-peptone solution (glycerol 50% (v/v), peptone 0.15 % (w/v)). Glycerol stocks were conserved at -80ºC. LB agar plates were stored at 4ºC for a few weeks.

4. 3. Bacterial growth measurement

Bacterial growth was systematically measured by taking 1 mL samples of the bacterial cultures and making optical density measurements at 600 nm (OD_{600}) in a *Nanodrop2000c* spectrophotometer (Thermo Fisher Scientific) with cuvettes.

4. 4. Transformation methods

Two different methods were used for the transformation of competent cells: electroporation and transformation by heat shock.

Electroporation Transformation

100 ng of DNA were added to 100 µL of ice-cold competent cells and the mixture was poured into a 0.2 cm electroporation cuvette (*Gene Pulser*, BioRad) that was kept on ice throughout the process. A single pulse was applied with a field strength of 2.5 kV, resistance of 200 Ω and capacitance of 25 µF, using the *MicroPulser TM electroporator* (BioRad), as illustrated in Figure 13.

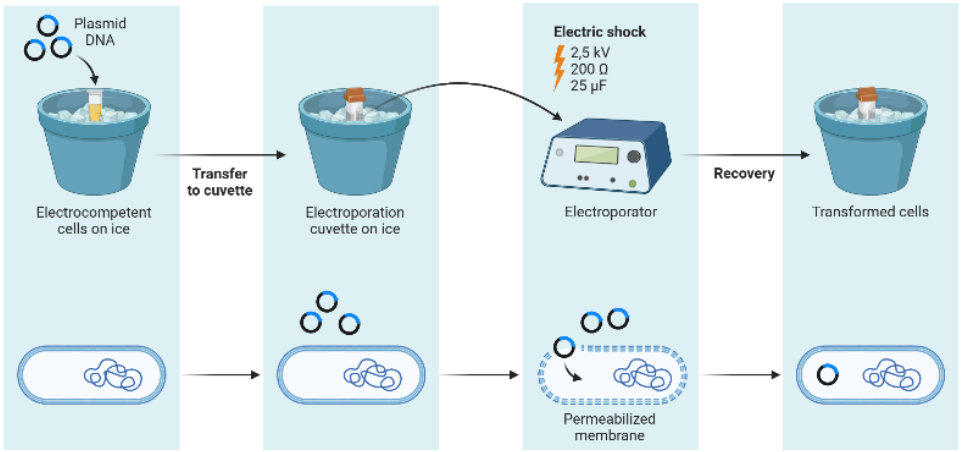


Figure 13. Electroporation transformation method. Created in BioRender.

Immediately after the electrical pulse, 1 mL of LB or SOC medium was added to the cuvette and the cells were incubated at 37°C for 45 min with shaking at 180 rpm. Following the incubation, some or all of the transformation mixture was plated onto a 10 cm LB agar plate containing the appropriate antibiotic and the plates were incubated at 37ºC overnight.

Heat Shock Transformation

100 ng of DNA were added to 100 µL of ice-cold competent cells and the mixture was incubated on ice for 30 minutes. The transformation tubes were heat shocked by placing the bottom of the tube into a 42ºC water bath for 45 seconds, and then immediately placing the tubes back on ice for 2 minutes, as shown in Figure 14.

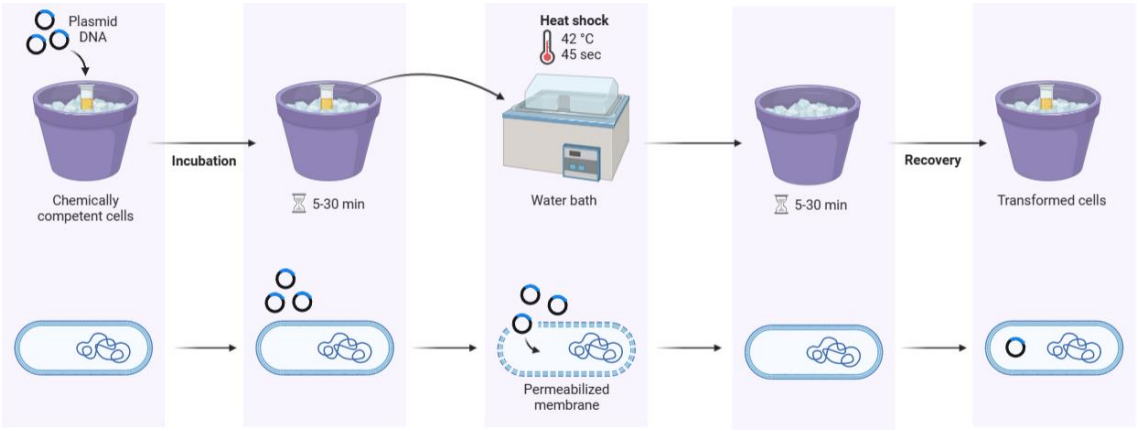


Figure 14. Heat Shock transformation method. Created in BioRender.

Immediately after, 500 µL of LB medium was added and the cells were incubated at 37°C for 45 minutes with shaking at 180 rpm. Subsequently, some or all of the transformation mixture was plated onto a 10 cm LB agar plate containing the appropriate antibiotic and the plates were incubated at 37ºC overnight.

4. 5. Preparation of chemical competent cells

Cells used for transformation were prepared to be chemically competent for heat shock transformation. The ultracompetent cells were prepared as follows: a colony was selected from a plate with the desired strain, inoculated into a flask and grown overnight at 37°C with shaking. The resulting culture was diluted 1:20 and further grown with shaking util it reached an OD600 of 0.5. After reaching the optical density, the culture was cooled on ice for 15 minutes. The cells were harvested by centrifugation at 3000 rpm at 4 °C for 10 minutes, and the resulting pellet was resuspended in TFBI buffer (100 mM KCl, 50 mM MnCl2, 30 mM K-Acetate, 10 mM CaCl2, 15% glycerol, ddH2O). The centrifugation and resuspension process were repeated a second time, this time using TFBII buffer (10 mM MOPS, 10 mM KCl, 75 mM CaCl2, 15% glycerol, ddH2O). Finally, the cells were aliquoted into 100 µl fractions and frozen at -80°C.

5. Molecular biology methods

5. 1. DNA extraction and purification from E. coli

Plasmid DNA was extracted from 10 mL LB cell cultures using the *GeneJET Plasmid Miniprep Kit* (Thermo Fisher Scientific) following the protocol described in the kit. DNA elution was performed using 50 µL of ultra-pure Millipore Milli-Q (*mq*) water at 60ºC.

5. 2. Nucleic acids quantification

DNA concentration was measured using a *Nanodrop2000c* spectrophotometer (Thermo Fisher Scientific) taking 1 or 2 µL samples. DNA purity was confirmed by an A260/280 nm ratio greater than 1.8.

5. 3. PCR

The DNA fragments were amplified using the polymerase chain reaction (PCR) in a C1000 TouchTM thermocycler (BioRad). Two different types of PCR were performed: one of them for the vectors and inserts, and another, to verify that the generation of the recombinant plasmid was correct.

In the first case, for the vectors and inserts PCRs, the high-fidelity enzyme Phusion Polymerase HF (Thermo Fisher Scientific) was used in final volumes of 50 µL. The PCR mix for these cases includes 10 µL of 5X Phusion HF Buffer, 1 µL of each oligo (forward and reverse), 0.5 µL of Phusion DNA polymerase (Thermo Fisher Scientific), 35.5 µL of water (mq) and variable volumes of plasmid template. The standard procedure was: 30 seconds of initial denaturation at 98°C, 30 cycles of amplification, including steps of denaturalization for 10 seconds at 98°C, annealing for 30 seconds at the corresponding annealing temperature (depending on the primers) and elongation for the appropriate time (30 s/Kb for PCRs less than 2000 nucleotides, and 20 s/Kb for PCRs more than 2000 nucleotides); and a final elongation step of 5 minutes at 72°C. After completion of the reaction, the samples were stored at 4°C for short-term conservation.

For the second case of PCR aimed to verify that the correct recombinant plasmid was generated, TAQ polymerase was used. The PCR mix included 10 µL of TAQ polymerase mix 2x (ABM), 1 µL of each oligo at 10 ng/µL (forward and reverse), 8 µL of water (*mq*) and a colony of the transformed bacteria. The standard procedure in this case was as follows: 3 seconds of initial denaturation at 94°C, 30 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 (60 s/Kb); and a last step of 5 minutes of final elongation at 72°C. After the completion of the reaction, samples were kept at 4°C for short-term conservation.

5. 4. Isothermal Assembly

Constructions were conducted by the Isothermal assembly Gibson method (Figure 15). The T5 exonuclease dilution buffer contained 50 % glycerol, 50 mM Tris 7.5, 0.1 mM EDTA, 1 mM dithiothreitol (DTT), 0.1 M NaCl and 0.1 % of Triton 100X. The isothermal reaction buffer was prepared by mixing 3 mL of 1M Tris-HCl pH 7.5, 150 µL of 2 M MgCl2, 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 water (*mq*) to a final volume of 6 mL. 1.5 µL of diluted T5 exonuclease 1U/ µL was then mixed with 40 µl the isothermal reaction buffer 5x, 2.5 µL of Phusion polymerase (Thermo Fisher Scientific) and 10 µL Taq Ligase (ABM), resulting in a final volume of 100 µL. The isothermal mixture was aliquoted in PCR tubes, with each tube containing 15 µl of the mixture.

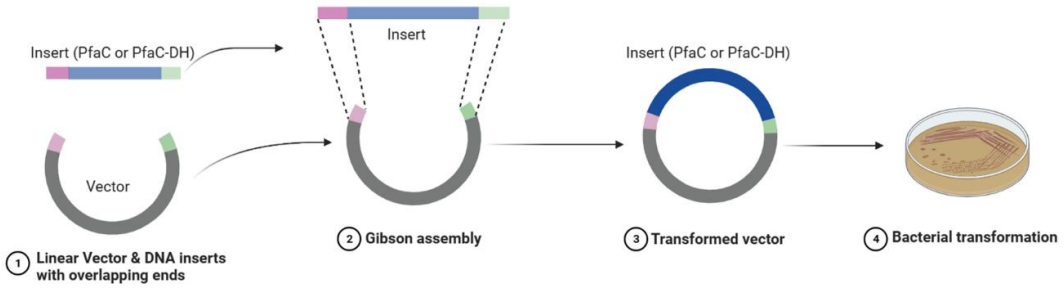


Figure 15. Gibson Assembly method. Created in BioRender.

Isothermal assembly reactions were performed by mixing 15 µl of the isothermal mix with a corresponding volume for 50-100 ng of the vector PCR and a 2X molar excess of the insert PCR. Reactions was conducted for 60 min at 50ºC in a the *C1000 Touch™* thermal cycler (BioRad).

5. 5. DNA electrophoresis in agarose gel

To verify that the size of the PCR products was correct, DNA electrophoresis was conducted in agarose gels. Solid agarose was dissolved in TAE 1X (45 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8.2) to a final concentration of 1% (w/v). 3 µL of *SYBR Safe* (Thermo Fisher Scientific) was added to 30 mL of 1% agarose for small gels and 12 µL to 120 mL for large gels. *GeneRuler 1Kb Plus DNA Ladder* (Thermo Fisher Scientific) was used as a molecular marker. Electrophoresis was performed at 120 V for 45 minutes using a horizontal electrophoretic running system (BioRad) and TAE 1X buffer. UV Gel Doc 2000 (BioRad) was used for gel visualization.

5. 6. Sanger sequencing

All plasmids were sequenced by *Eurofins Genomics* (Milano, Italy) using the Sanger method. Plasmids were sequenced using the primers described in Table 3.

6. Protein methods

6. 1. Protein overexpression and cell lysis

The proteins were purified for future experiments. For this, the overexpression was conducted with the following protocol.

Expression tests

Cultures were inoculated in 50 mL LB flasks with 50 µL of the corresponding antibiotic 1000x and 500 µL of the 10 mL pre-inoculum grown overnight at 37ºC with shaking. The culture was incubated at 37°C until reaching an OD_{600} of 0.6-0.7 and left on the bench at room temperature until reaching an OD_{600} of 0.8 (10-15 minutes). Once cooled, protein expression is induced by adding 10 µL or 25 µL 1M IPTG for final concentrations of 0.2 mM or 0.5 mM, respectively. The bacterial cultures were incubated overnight at the appropriate temperature for the expression of the PfaC protein: 12ºC for the *ArcticExpress* strain, and 15 or 18ºC for the *Rosetta(DE3)*, *BL21(DE3)* and *C41(DE3)* strains.

Once the overexpression was conducted, the cells were lysed to subsequently proceed with the purification of the proteins. After leaving the cultures overexpress the protein overnight, they were centrifuged in 50 mL Falcon tubes for 10 minutes at 5000 rpm. The supernatant was discarded, and the pellets were frozen at -80ºC for preservation. This freezing step further serves as an extra to cell lysis. The pellets were resuspended in 1.5 mL of lysis buffer each (50 mM Tris pH 8, 300 mM NaCl, 20 mM imidazole, 1 mM DTT, 0.5 mM phenylmethylsulphonyl fluoride (PMSF), 5 mM benzamidine, 0.5 mg/mL lysozyme and DNAse). In the case of purification of the protein with the GST tag, imidazole was not added. The pH of the buffer was adjusted after stabilization at 4ºC, and the solution was filtered.

Cell lysis was achieved in two different ways. On the one hand, lysis was performed with a sonicator, which generates high-frequency waves that break the cells. The other lysis method involves using detergents, such as Triton 10%, which breaks the lipid barrier solubilizing proteins and interrupting the lipid-lipid, lipid-protein and protein-protein interaction. Lysis by sonication was achieved by 20 cycles of sonication (30 seconds on, 30 seconds off) with an amplitude of 60% conducted in an ice bath at 4°C. For detergent lysis, the pellet was first resuspended in 0,5 ml of lysis buffer and incubated it on ice for 30-45 minutes, to allow the action of lysozyme. Then 50 µL of Triton 10% was added to each sample.

Large-scale

The protocol followed was the same. 1L LB flasks containing 1 mL of the corresponding antibiotic 1000x, were inoculated with 20 mL of the pre-inoculum, prepared the day before. For overexpression, 200 µL or 250 µL of 1M IPTG (final concentration 0.2 mM or 0.25 mM) were added. Then cultures were centrifuged in 500 mL bottles with JA10 rotor at 4000 rpm for 15-20 minutes each, using *Avanti J-30I* centrifuge (Beckman Coulter). The pellets were stored in the freezer at -80ºC, following the same procedure as in the expression tests. Following the expression test protocol, but on larger scale, the pellets from 2L cultures were resuspended in 30-50 mL of lysis buffer. Cell lysis was achieved by sonication at 60% amplitude on ice, using 4 cycles (5 seconds on, 25 seconds off). The lysate was then centrifuged for 45 minutes at 20000 rpm and 4°C using the 30.1 rotor in the *Avanti J-30I* centrifuge (Beckman Coulter).

6. 2. Protein purification

The purification of the proteins was conducted using two methods: affinity and gel filtration chromatography. All buffers were filtered with a 0.45 µm-pore filter.

The proteins were labeled with four different tags. Three of them contain the polyhistidine tag (6His) in the N-terminal domain. Two of them also have the maltose-binding protein (MBP) or thioredoxin (TRX) tag, combined with the histidine tag. These tags have affinity for nickel ions (Ni^{2+}), which can be found in Ni-NTA chelating resins (NTA, nitrilotriacetic acid, minimizes metal leaching during purification), capable of binding up to 60 mg of protein per milliliter of resin (CubeBiotech, Ni-Indigo). The other tag used was glutathione S-transferase (GST), located also in the N-terminal domain of the protein. For this tag, glutathione-agarose resins were used, which can bind about 40 mg of protein per millimeter of resin (CubeBiotech).

Expression tests purification

Protein purification was conducted on ice using purification columns to which 100 µL of resin preserved in ethanol 20% (1:1) was added (Figure 16). Subsequently, the resin was washed with 500 µL of water (*mq*), followed by 500 µL of Tr300-20i buffer (50 mM Tris pH 8, 300 mM NaCl, 20 mM imidazole), for GST-tagged proteins, or Tr300+DTT buffer (50 mM Tris pH 8, 300 mM NaCl, 1 mM DTT) for the rest of the tags. Once the column containing the resin was equilibrated, the supernatant sample of the previously centrifuged lysed product, was added.

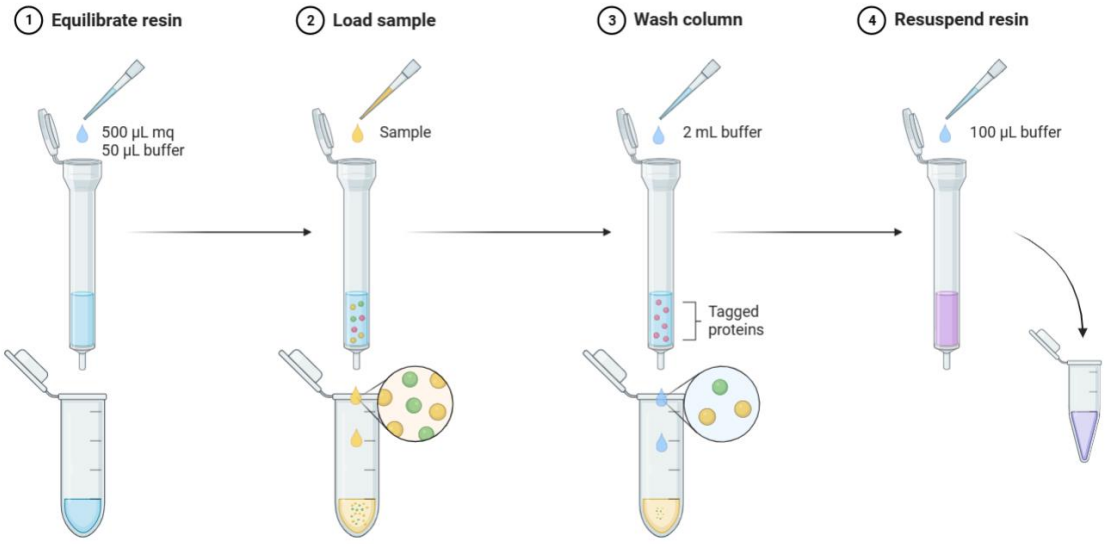


Figure 16. Protein purification for expression tests. Created in BioRender.

The flow-through was discarded and the resin was washed with 2 mL of Tr300 buffer (50 mM Tris pH 8, 300 mM NaCl) or Tr300-20i buffer. Once washed, the resin was resuspended in 100 µL of buffer and transferred to another Eppendorf tube for preservation.

Affinity Chromatography

For large-scale protein purification, containing the histidine tag, chromatography was conducted at 4°C using the *ÄKTA Start* protein purification system (GE Healthcare). The sample was loaded at a flowrate of 3 mL/min. The 5 mL *HisTrap HP* column (GE Healthcare) was first washed with 5 CV of water (*mq*) and then equilibrated with 5 CV of buffer A (Tr300-20i + DTT). Following the loading step, the column was washed with 5 CV of buffer A to remove nonspecific binding. The protein was eluted using a linear gradient of buffer B, Tr300-400i (50 mM Tris-HCl pH 8, 300 mM NaCl, 400 mM Imidazole, 1 mM DTT) from 0% to 100%. Fractions of 1.8 mL were collected during the elution process. In addition, in some cases it was necessary to perform a dialysis process overnight to change the salt and buffer composition of the sample. A dialysis membrane with a cutoff of *Spectrum™ Labs Spectra/Por™ 1 6-8 kD MWCO Standard RC Dry Dialysis Kits* (Spectrum Labs) was used. The HisMBP-PfaC$_{DH}$ protein was against 1 L of Tr300 buffer, at 4ºC with stirring. If further purification was needed, it was achieved via gel filtration chromatography using buffer Tr150 (25 mM Tris pH 7.5, 150 mM NaCl and 1mM DTT).

For proteins containing the GST tag, chromatography was conducted at 4°C using 5 mL of *Pure Cube Glutathione Agarose* Resin loaded into an empty column (10 mg/mL binding capacity). The Glutathione Agarose resin was first washed with 50 mL of water (*mq*) and then equilibrated with 50 mL of buffer Tr300+DTT. The resin was then mixed with the sample and incubated for 2 hours at 4ºC on a roller. Afterward , the mixture was loaded into an empty column, where the flowthrough was collected, and the resin washed with 400 mL of buffer Tr300+DTT. The resin was resuspended in 5 mL of buffer Tr300 and placed on a Falcon tube for overnight incubation at 4ºC, after adding 1 mL of the TEV protease. The following day, the mixture was poured back into the column and the flowthrough was collected. In this case, two 15 mL wash cycles were performed to maximize protein recovery. After this step, the flowthrough and the two wash samples were mixed and concentrated using an *Amicon® Ultra-15 Centrifugal Filter Unit* (50kDa MWCO) for its later purification via gel filtration chromatography with the buffer Hp200 (7.5) (25 mM Hepes pH 7.5, 200 mM NaCl, 1 mM DTT).

Gel filtration chromatography

Gel filtration chromatography is a type of exclusion chromatography that allows the separation of molecules based on their molecular size. It was performed on the *ÄKTA pure* protein purification system (GE Healthcare) with two columns, a *Superdex 200 Increase 10/300* column (GE Healthcare) and a *Superose 6 Increase 10/300 GL* column (GE Healthcare). The column was equilibrated with 1.5 CV of buffer. The protein sample was injected using a 500 µL loop and 0,5 mL fractions were collected at a flow rate of 1 mL/min. The elution profile was monitored by UV absorbance at 280 nm.

6. 3. Protein electrophoresis in denaturing conditions

Samples taken at different stages of overexpression (-IPTG, +IPTG), bacterial lysate and purification steps were subjected to SDS-PAGE (Sodium dodecyl-sulfate polyacrylamide gel electrophoresis) to confirm successful overexpression and protein purification. The samples were mixed with Loading Buffer 5X (250 mM Tris-HCl pH 6.8, 10% of SDS, 50% glycerol, 0.025% bromophenol blue dye and 500 mM DTT). All samples were incubated for 5 minutes at 90-100°C and loaded onto acrylamide gels (8-10%). *PageRuler™ Plus Prestained Protein Ladder* (Thermo Fisher Scientific) was used as molecular weight marker. Electrophoresis was performed at 200V for 45 minutes or 1 hour, depending on the acrylamide percentage of the gel, in SDS-PAGE 1X buffer (25 mM Tris, 192 mM glycine, 1% (w/v) SDS, pH 8.4). Finally, the gel was stained with BlueSafe.

6. 4. Protein concentration and quantification

Absorbance at 280 nm was the method employed for protein quantification. The absorbance at 280 nm was measured with the *Nanodrop 2000c* (Thermo Fisher Scientific). The concentration of the purified protein samples was estimated by applying the Beer-Lambert law, $A = \varepsilon c l$. The theoretical extinction coefficients were calculated using the *ProtParam ExPASy* tool.

7. X-ray crystallography

7. 1. Pre-crystallization test

Pre-crystallization tests were performed to determine the appropriate protein concentration for the crystallization screening. The reagents used are A1 (0.1 M TRIS hydrochloride pH 8.5, 2.0 M Ammonium sulfate), A2 (0.1 M TRIS hydrochloride pH 8.5, 0.2 M Magnesium chloride hexahydrate, 30% w/v Polyethylene glycol 4000), B1 (0.1 M TRIS hydrochloride pH 8.5, 1.0 M Ammonium sulfate) and B2 (0.1 M TRIS hydrochloride pH 8.5, 0.2 M Magnesium chloride hexahydrate, 15% w/v Polyethylene glycol 4000). From each reagent, 0.5 mL were placed on four different reservoirs of a EasyXtal 24-Well Plate. On the center of the screw cap 0.5 µL of protein sample were mixed with 0.5 µL of the corresponding reagent. Subsequently, the covers are placed back on the plate and the plate was left in an incubator at 21ºC for 1 hour (1-2 hours). After this time, the drops were observed under a microscope to evaluate the results. Once the appropriate precipitate is observed in the drops, crystallization screens were performed.

7. 2. Crystallization screens

To determine the optimal crystallization conditions of the PfaC protein (DH domain), several sitting-drop vapor diffusion commercially available screening kits were employed: JCSG-plus (Molecular Dimensions), Crystal Screen (Hampton Research), ProPlex (Molecular Dimensions) and Morpheus (Molecular Dimensions). These kits contain a wide variety of reagents, including a wide range of precipitants, organic compounds, and buffers at different pHs. 40 µl of each condition of the crystallization screen were dispensed into MRC 2 well crystallization plates (Molecular Dimensions) Each crystallization well was filled with 0.5 µL of protein and 0.5 µL of the corresponding precipitant, making a total of 96 different conditions per plate (96-well sitting drop plates).

8. Bioinformatic analysis

8. 1. Databases

The nucleic acid sequences described in this work were obtained from the NCBI database (*National Center for Biotechnology Information*, National Institutes of Health, NIH). The protein sequences were obtained from Uniprot and the protein structure files, were obtained from the PDB (*Protein Data Bank*, Research Collaboratory for Structural Bioinformatics, RCSB).

8. 2. Bioinformatic tools for protein and nucleotide analysis

For the alignment of sequences from different microorganisms confirmed to produce DHA or EPA, the tool used was *Multiple Sequence Alignment* from *Clustal Omega Server* (EMBL-EBI). *Dali* server was used for comparing 3D protein structures. For structure predictions, *ColabFold*, *SWISS-MODEL (ExPASy)* and *Phyre2* online servers were used. Visualization and manipulation of protein structures was performed using *PyMOL*. *CLC Sequence Viewer 8* (QIAGEN) was used for protein and genetic sequences

visualization and analysis. The *OligoAnalyzer* tool (IDT) was used to calculate specific features of oligonucleotides (length, melting temperature (Tm), GC content, etc.).

The *Blast* online server (NCBI) was used to identify homologous protein and DNA sequences. The *ProtParam* tool (ExPASy) was used to estimate protein parameters such as the molar extinction coefficient (ε), the isoelectric point (pI) and the molecular mass, shown in Table 4.

Table 4. Theoretical parameters calculated by *ProtParam* tool (EsPASy).

Protein (PfaC)	**Number of aminoacids**	**Molar extinction coefficient ($M^{-1}cm^{-1}$)**	**Molecular weight (KDa)**	**Isoelectric point (pI)**
Full length	2011	187715	218.4	5.52
1111-C_{term}	900	108360	100.3	5.05

Results and Discussion

1. Structural analysis

In this work we started analyzing the structure of PfaC from *Moritella marina*. The structure of the KS-CLF domains was known since the group of Dr. Moncalián experimentally solved the structure using X-ray crystallography. This structure is described in Santin et al., 2020, and is depicted in Figure 9 of the *Introduction* section. However, there is currently no available structural data for the full-length PfaC protein or its C-terminal DH domains. To gain insights into the structure-function relationship of this protein, we employed the artificial intelligence program *AlphaFold2* to model the structure of full-length PfaC. The predicted structure of the full-length PfaC protein (KS-CLF and DH domains) is shown in Figure 17A. In Figures 17B and 17C, the KS-CLF and DH domains are shown separately, for better visualization.

In the predicted structure no interactions are observed between the KS-CLF and DH domains, nor between the two DH domains themselves. This lack of interaction was further supported by our attempts to purify the complete protein, which proved unsuccessful. Instead, we mainly obtained degradation products, leading us to hypothesize that the domains exist as individual subunits and do not form a tightly folded complex.

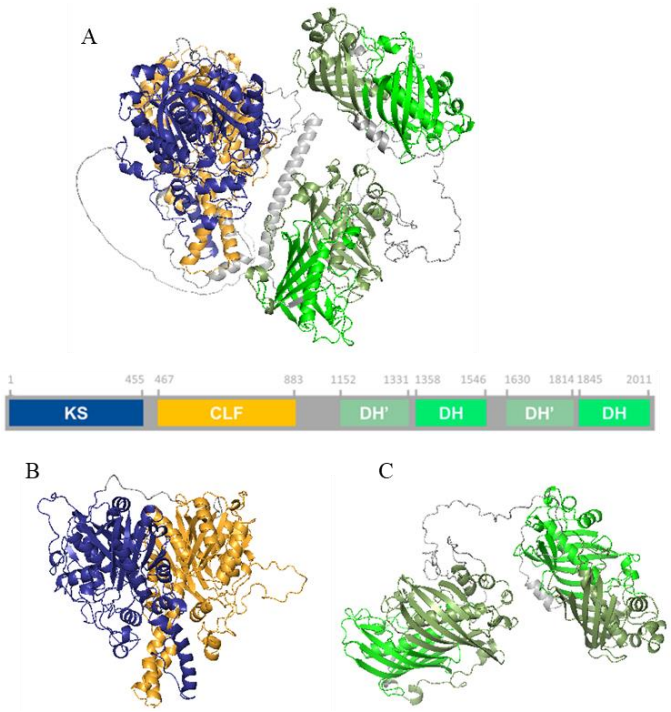


Figure 17. Domain prediction for PfaC. Color coding of domains: KS, dark blue; CLF, yellow; DH', dark green; DH, light green. **A**. PfaC from *Moritella marina* prediction model by *AlphaFold2*. **B**. Expanded view of KS domains. **C**. Expanded view of DH domains (complete PfaC protein prediction).

Moreover, several fragments of the protein (colored in grey in Figure 17A) are not part of any specific domain and have no assigned function. *AlphaFold2* predicts these regions as loops or α-helixes, but with no apparent interaction with the well-defined domains of the protein. Figure 18A displays the

structure of the complete PfaC protein predicted by *AlphaFold2*, with each residue colored according to the per-residue confidence score (pLDDT, predicted local distance difference test) between 0 and 100. The color scheme indicates the accuracy of the prediction, where dark blue represents high accuracy (> 90), light blue corresponds to slightly lower accuracy (70-90), yellow indicates low accuracy (70-50) and orange represents very low accuracy (< 50). From the image we can see that the confidence of the structural prediction of *AlphaFold2* for the PfaC protein is mostly very high, except in the loops previously mentioned.

We have also examined the PAE (predicted absolute error) plot corresponding to the PfaC *AlphaFold2* model. This plot provides a pairwise measure of the predicted error in residue-residue spacing. In the predicted structure, the PAE plot reveals three distinct regions with low PAE scores corresponding to the KS-CLF, DH1'-DH1 and DH2'-DH2 domains. These regions indicate a higher confidence in the accuracy of the predicted structure for these domains. The PAE plots also suggest that PfaC has a modular organization, with distinct domains that function independently and potential unstructured regions that do not participate in domain-domain interactions. The electrostatic potential surface is calculated using *PyMOL*. The surface representation shows positively charged areas (blue),negatively charged areas (red) and residues with a neutral charge (white) (Figure 18C). This electrostatic potential distribution may play a crucial role in specific interactions with other Pfa proteins through electrostatic interactions.

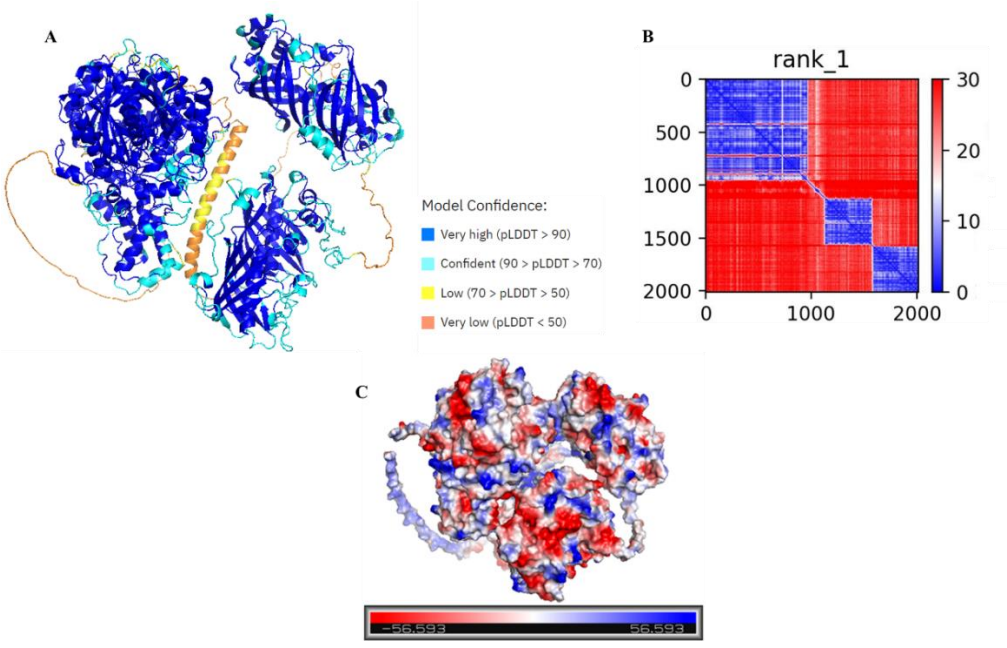


Figure 18. **A**. *AlphaFold2* predicted structure of PfaC protein colored by per-residue confidence score (pLDDT). B. Residue-residue alignment confidence (PAE) plot, representing the expected position error for each residue in the sequence. **C**. Electrostatic potential surface of PfaC.

After studying the model of full-length PfaC we decided to focus on the DH domains, for which the structure has not been experimentally determined (Figure 19). To address this, we employed *AlphaFold2* again to predict the structure of the isolated DH domains from the PfaC protein. In the predicted structure, the DH domains exhibit a similar organization to the described in *Photobacterium profundum* where the domains follow a DH1'-DH1- DH2'-DH2 pattern (Oyola-Robles et al., 2013).

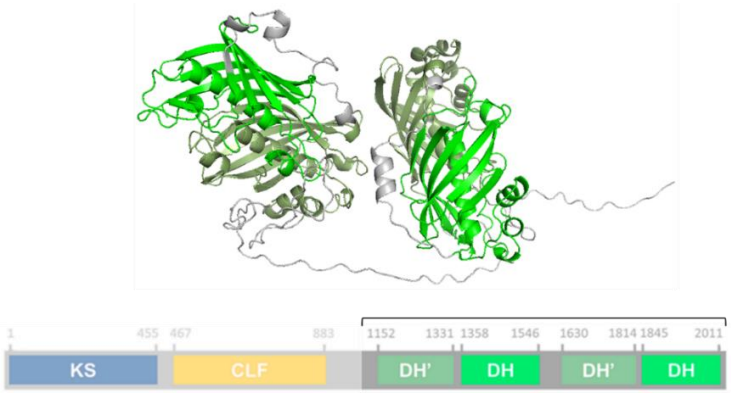
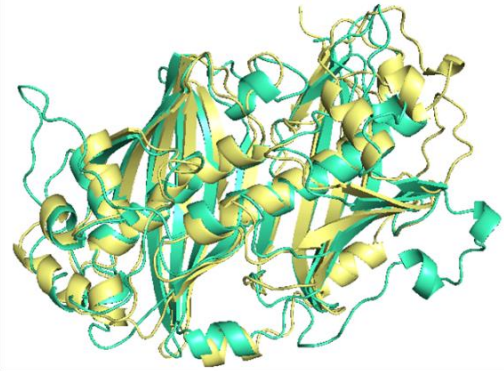


Figure 19. PfaC DH domains from *Moritella marina* predicted by *AlphaFold2*.

The folding shown by these domains corresponds to the 'hot dog' fold, found in a large number of proteins associated with the biosynthesis of fatty acids and their regulation in metabolism. Dimers with a 'hot dog' fold are formed by the association of β-sheets of both subunits, resulting in a continuous antiparallel β-sheet. The two helices lie antiparallel to each other at the dimer interface. Enzymes with this dimeric 'hot dog' fold, have active sites composed of residues from both subunits. In addition, it has been described that the two domains may present a sequence similarity (Pidugu et al., 2009). By aligning the two subunits, we see this to be true for the DH domains of PfaC (Figure 20). The sequence alignment reveals a sequence identity of 24.06%. However, the structure alignment has an average RMSD of 5.659 Å (calculated with *PyMOL*) that reveals a high structural similarity.



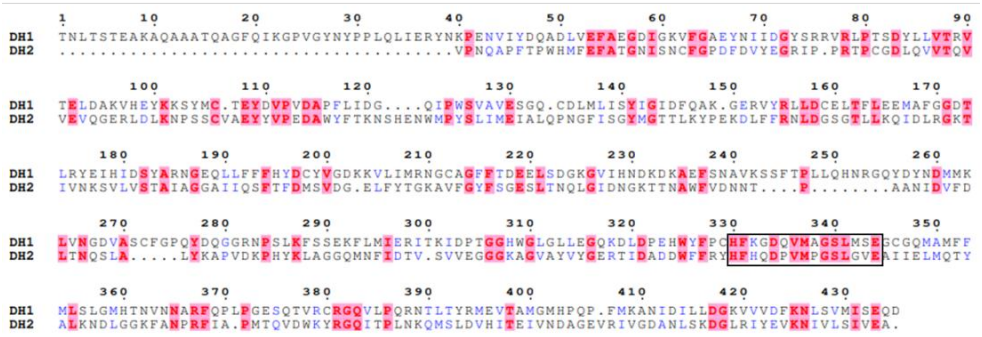


Figure 20. Alignment of the DH domains of PfaC. The DH1'-DH1 domain (residues 1152-1546) is colored yellow, and the DH2'-DH2 domain (residues 1630-2011) is colored green. The alignment reveals a sequence identity of 24.06%, calculated by the *Clustal Omega* tool. Residues that remain identical are colored red on a pink background, and conserved residues are colored blue. Highlighted with a black box, the area of the sequence in which the residues that form the active site of the protein are found.

As observed in the predicted structure, there is minimal interaction between the two DH domains. In addition, we later verified this lack of interaction when trying to purify the $PfaC_{DH}$ and obtaining many degradation products that we hypothesize are the separated subunits of the domains. The DH domains of the PfaC protein predicted by *AlphaFold2* were colored according to the per-residue confidence score (pLDDT). Additionally, the residue-residue alignment confidence (PAE) plot indicates that, although the two domains are individually well defined, the spatial relationship between them has not been accurately predicted, because the PAE is high for residue pairs x, y from two different domains. The electrostatic potential surface represented indicates that near the surface of the protein there are positively charged areas (blue) and also negatively charged areas (red). The white color corresponds to residues with a neutral charge. The possible interactions between the two domains are marked in black. They correspond to those areas in which a positively charged surface of one domain faces one negatively charged of the opposite domain (Figure 21B). This possible interaction should be further confirmed experimentally.

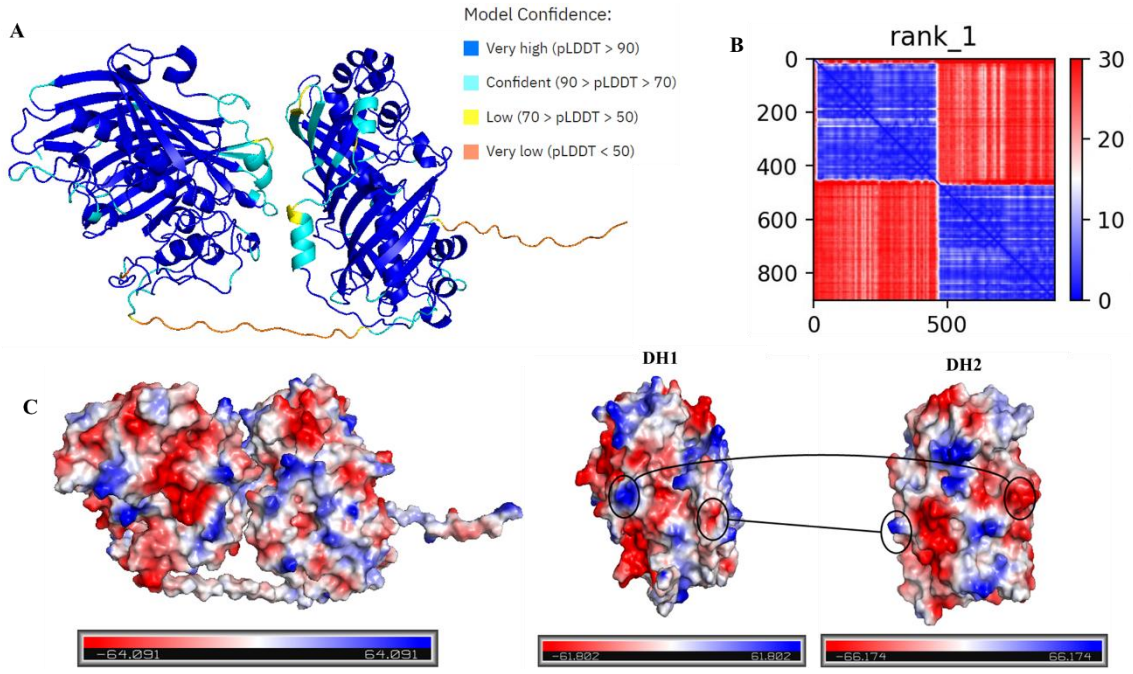


Figure 21. **A**. *AlphaFold2* predicted structure of $PfaC_{DH}$ protein colored by per-residue confidence score (pLDDT). B. Residue-residue alignment confidence (PAE) plot, representing the expected position error for each residue in the sequence. **C**. Electrostatic potential surface of $PfaC_{DH}$. The parts of each domain that are facing each other are shown on the right, with possible interactions between the two domains marked in black.

The structure of PfaC from *Moritella marina* or other marine bacteria is not currently solved so the analysis of its structure-function relationship is challenging. To locate the active sites in the DH domains of the PfaC protein and describe their function, we compared the structure predicted by *AlphaFold2* with several other previously described proteins using tools such as *Dali Server*, *Phyre2*, and *Clustal Omega*. After many alignments with different proteins, we found the protein 3-hydroxyacyl-ACP dehydratase (FabZ) from *Yersinia pestis* (McGillick et al., 2016), which aligns with the DH domains of PfaC from *Moritella marina* align. This finding has enabled us to proceed with the study of the active site.

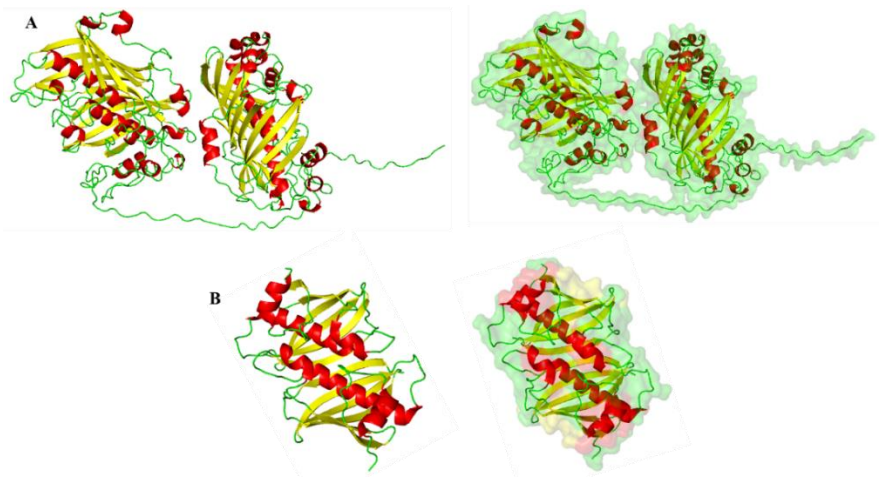


Figure 22. Proteins colored according to their secondary structure with α-helices in red, β-sheets in yellow, and loop regions in green. **A**. $PfaC_{DH}$ prediction by *AlphaFold2*. **B**. 3-hydroxyacyl-ACP dehydratase (FabZ) from *Yersinia pestis*.

At first glance it is observed a certain similarity between the structures of PfaC and FabZ proteins (Figure 22). However, when performing the alignment of the DH domains of PfaC with FabZ, we did not get a convenient alignment. Nevertheless, by aligning the subunits separately, a successful structural alignment was achieved (Figure 23).

We further analyzed the superposition of the known active site of FabZ with PfaC. The His54 and Glu68 residues of FabZ correspond to the His1441 and Glu1455 residues of one DH subunit; and the residues His1907 and Glu1921 from the other subunit (Figure 23, Supplementary Figure S1).

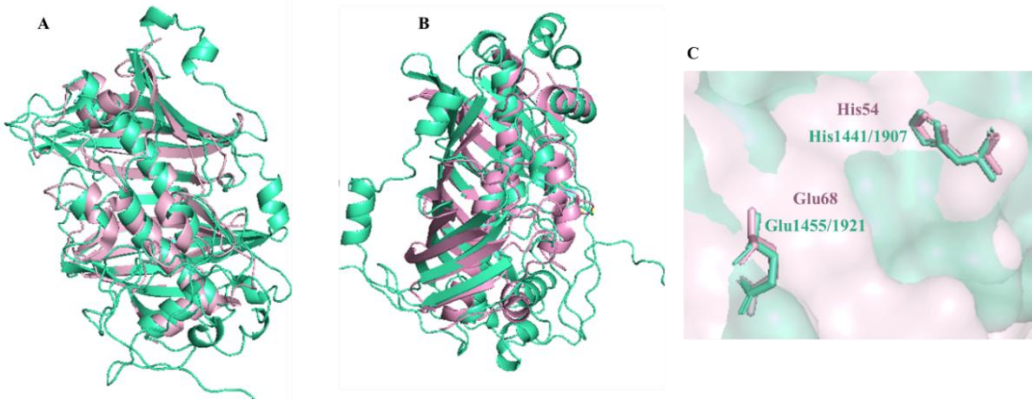


Figure 23. Structural alignment of PfaC DH domains (green) with FabZ (pink) from *Yersinia pestis*. **A**. Alignment of FabZ to DH domain (residues 1152-1546). **B**. Alignment of FabZ to DH domain (residues 1630-2011). **C**. Highlighted residues that form the catalytic stie.

As described in the article written by McGillick et al., 2016, these two residues of FabZ, His54 and Glu68, form a catalytic site, which is contributed by the presence of another His54 in the opposite monomer. In the case of the DH domains of PfaC, these residues are also conserved. The assigned function to this active site is the dehydration reaction of β-hydroxyacyl-ACP into α,β-unsaturated acyl-ACP, thanks to the action of one or two FabZ proteins. FabZ is a key dehydratase in the formation of saturated and unsaturated fatty acids.

The exact mechanism for the reaction that takes place at this active site of FabA from *Pseudomonas aeruginosa* (Moynié et al., 2016) is described in White et al., 2005. In this mechanism, histidine acts as a catalytic base, abstracting a proton from C2 of the substrate, while aspartic acid (or glutamic acid in $PfaC_{DH}$) promotes the removal of the hydroxyl group to generate the enoyl-ACP product. Substrate trans-2-decenoyl-ACP is then isomerized to cis-3-decenoyl-ACP with histidine adding a proton to C2 and aspartic acid (glutamic acid in $PfaC_{DH}$) abstracting a proton from C4. The tunneling structure of the active site is arranged by hydrogen bonding interactions with aspartic acid and amides of cysteine and glycine (or glutamic acid, serine and glycine in $PfaC_{DH}$) (Figure 24).

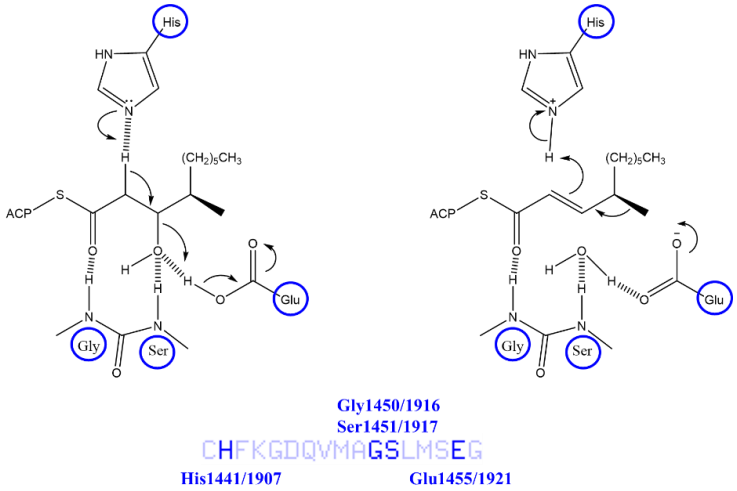


Figure 24. Mechanism described for the active site of the DH domain. Modified from White et al., 2005.

This is the first analysis of the *AlphafFold2* model of PfaC of *Moritella marina*. The study confirms the structural and sequence similarity of the DH domains of PfaC and those described for FabA and FabZ, confirming their dehydratase function. The position of the residues present in the active site of FabA and FabZ are conserved in PfaC, therefore confirming their functional significance. PfaC participates in the biosynthesis of fatty acids together with the other proteins encoded by the *pfa* gene cluster, PfaA, PfaB, PfaD and PfaE. Although solving its structure is a challenging task, it is of great interest to be able to elucidate all the details in LC-PUFA biosynthesis by *Moritella marina* and other marine bacteria.

2. Cloning of *pfaC* constructs

To achieve the characterization of PfaC, we generated several plasmids containing the *pfaC* gene and transformed them into *E.coli* strains for protein overexpression. Modifying the *pfa* gene cluster in its natural host (*Moritella marina*) is very challenging, that is why we modified and expressed it in *E.coli*. The Isothermal Assembly Gibson method was used for the construction of the vectors, obtaining six different ones. Three of these vectors contain the complete PfaC protein and three more contain only the PfaC DH domains. In addition, they were built with three different purification tags, to observe the differences in protein solubility once expressed. These three tags were GST-TEV, HisMBP-TEV, and HisTRX-TEV (Supplementary Figures S2-7).

To prepare inserts and vectors various conditions were used (Table 5), following the steps described in the *Materials and methods* section. The bands in the electrophoresis gels (agarose 1%), show that the

PCR results are in the theoretically calculated expected positions described in Table 5 (Figure 25). Once the results of these PCRs were confirmed and the fragments were purified, the next step was the Isothermal Gibson Assembly of inserts and vectors following the steps described in the *Materials and methods* section.

Table 5. PCR prepared for cloning PfaC and $PfaC_{DH}$.

	Insert		**Vector**		
	Full length PfaC	**$PfaC_{DH}$**	**pGST**	**pHisMBP**	**pHisTRX**
Number id.	1	2	1	2	3
Oligo 1	O1	O3	O284	O284	O284
Oligo 2	O2	O2	O94	O94	O94
Plasmid template	PfaC	PfaC	pMLG31	pMLG32	pMLG33
(ng)	20	1	10	5	15
PCR size	6082	2749	5905	6462	5460
Elongation time (s)	180	122	118	180	180
Annealing temperature (ºC)	55	64	72	72	65
Concentration (ng/µL)	61.1	27.7	67.1	29.1	63.2
260/280	1.96	1.82	1.85	1.6	1.84

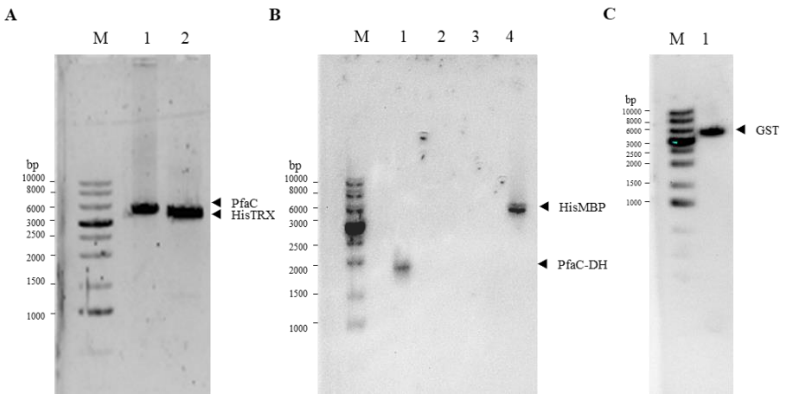


Figure 25. DNA electrophoresis gel of the PCR results for PfaC and $PfaC_{DH}$ cloning. Lane M shows the *GeneRuler 1Kb DNA Ladder* (Thermo Fisher Scientific). **A.** Lane 1, Insert 1 PCR (full-length PfaC, 6082 bp); lane 2, Vector 3 PCR (HisTRX, 5460 bp). **B.** Lane 1, Insert 2 PCR ($PfaC_{DH}$, 2749 bp); lane 4, Vector 2 PCR (HisMBP, 6462 bp). **C.** Lane 1, Vector 1 PCR (GST, 5905 bp).

Different conditions were evaluated in the isothermal assembly reaction (Table 6). After transformation into Top10 competent cells by heat-shock of the isothermal product colonies were obtained for all six constructs. To confirm the presence of recombinant plasmids colony PCR analysis we confirmed the presence of the recombinant plasmids with colony PCR

Table 6. Isothermal Assembly conditions.

	Full-length PfaC			**$PfaC_{DH}$**		
Purification tags	GST	HisMBP	HisTRX	GST	HisMBP	HisTRX
Id.	pPI1	pPI2	pPI3	pPI4	pPI5	pPI6
Vector	1	2	3	1	2	3
(ng)	100	76	97	100	76	98
Insert	1	1	1	2	2	2
Concentration (ng/µL)	45.1	71.2	39.8	36.2	41.0	26.8
260/280	1.85	1.81	1.88	1.86	1.82	1.75

3. Expression and purification of PfaC

Expression tests

For the overexpression and purification of the PfaC protein labeled with an histidine tag, we first used an already constructed PfaC histidine plasmid. To improve the purification yield of this protein we made several approaches with four strains of: *BL21(DE3), C41(DE3), Rosetta(DE3)* and *ArcticExpress(DE3)*; varying the overexpression temperatures with IPTG and using different lysis methods.

We experimented to find the protocol with which we would have a higher yield in obtaining the protein. To do this, we used two concentrations of IPTG (0.2 mM and 0.5 mM), and we varied the overexpression temperature (12, 15 and 18ºC). We determined that the best temperatures were: 15ºC as the lowest overexpression temperature or, in the case of the *ArcticExpress(DE3)* strain, 12ºC; and 18ºC as the highest (Figures 26 and 27). The expected size of the protein is 219 kDa.

Overexpression has been achieved regardless of the strain and the conditions used. The strains with the highest levels of overexpression are *BL21(DE3), Rosetta(DE3)* and *ArcticExpress(DE3).* It should be noted that in the case of the *C41(DE3)* strain the level of overexpression is low in comparison to the one obtained in the ones mentioned.

When analyzing the results obtained by performing small volume purification assays starting from 50 mL of culture (expression tests), we can see that for the *BL21(DE3)* strain, the protein is in the pellet and there is no protein bound to the resin. Only a small amount of protein is seen on the resin when overexpressed at 18°C and 0.2 mM IPTG. Overexpression levels of the *C41(DE3)* strain were low, but a

small amount of protein is observed on the resin. The experiments with this strain were repeated several times, always obtaining the same results (Figure 27B). In the *Rosetta(DE3)* strain we can see that the protein in all cases the protein is in the pellet and, although soluble protein is appreciated in the supernatant fraction, no protein is bound to the resin. This would indicate that the protein is not folded correctly or has lost the histidine tag in the purification process (Figure 27C). We see different results for *ArcticExpress(DE3)* strain, depending on the temperature.

Overexpressing at 18°C, the PfaC-His protein is not retained on the resin, as it aggregated and was lost in the pellet. It should be noted that, for this temperature, we can see some bands that correspond to chaperonin 60 (Cpn60, 57 kDa) which is expressed by this strain (Figure 27D). The chaperonin 60 expressed in the *ArcticExpress(DE3)* strain improves the overexpression of proteins at low temperatures, possibly increasing its yield and obtaining more soluble protein. For overexpression at 12ºC and with an IPTG concentration of 0.2 mM, a band was obtained with the expected size of the protein (219 kDa). The quality of the overexpression samples in the *ArcticExpress(DE3)* strain was better than in the other conditions where we observed proteins on the resin (*BL21(DE3)* and *C41(DE3)*). In addition, it presents fewer high molecular weight contamination bands. Therefore, we decided to use the *ArcticExpress(DE3)* strain under these conditions to scale up to higher volumes.

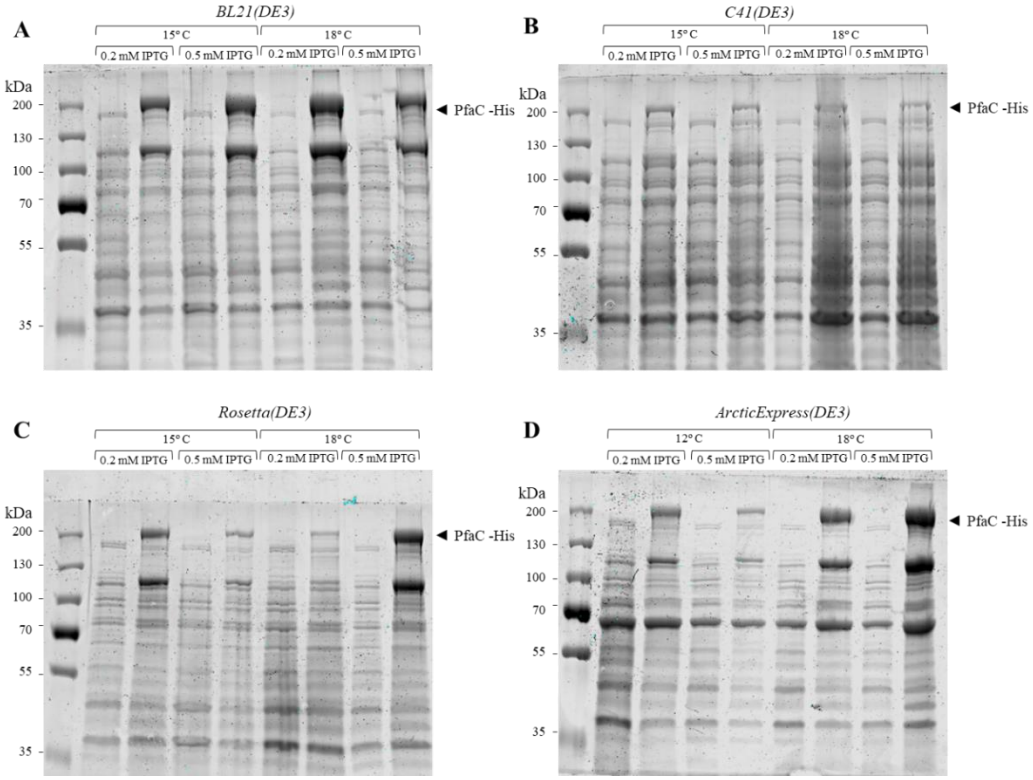


Figure 26. SDS-PAGE gel of the overexpression samples of PfaC-His (219 kDa). Lane M shows the *PageRuler™ Plus Prestained Protein Ladder* (Thermo Fisher Scientific). In all experiments, the samples are ordered in pairs as follows: -IPTG, +IPTG. **A.** *BL21(DE3)*. **B.** *C41(DE3)*. **C.** *Rosetta(DE3)*. **D.** *ArcticExpress(DE3)*.

It has been previously proven that PfaC can be produced in the *Rosetta (DE3)* strain of *E.coli* (Santin et al., 2020). In our study, we further confirmed that its production in the strains *ArcticExpress(DE3)* and *BL21(DE3)* also allows the expression of PfaC at similar levels.

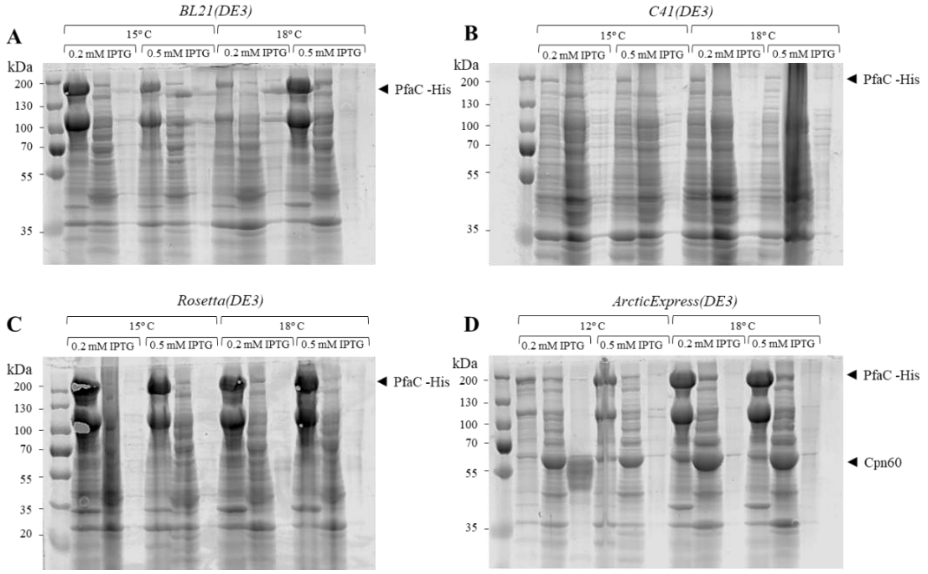


Figure 27. SDS-PAGE gel of the purification samples of PfaC-His. Lane M shows the *PageRuler™ Plus Prestained Protein Ladder* (Thermo Fisher Scientific). In all experiments, the samples are ordered in threes as follows: pellet, supernatant, resin. **A.** *BL21(DE3)*. **B.** *C41(DE3)*. **C.** *Rosetta(DE3)*. **D.** *ArcticExpress(DE3)*.

The purification of the PfaC protein with the GST, HisMBP and HisTRX tags was also tested. With this, an attempt was made to improve the solubility and stability of the protein, to prevent it from remaining in the insoluble fractions and to obtain more protein bound to the resin. PfaC overexpression and purification were not improved by the addition of the purification tags HisMBP, GST and HisTRX (Figure 30A; Figure 30B lanes 2-8). It should be noted that in the case of the HisMBP-PfaC an intense band can be observed, which could potentially correspond to overexpressed and purified maltose-binding protein (MBP, 40.2 kDa), without the presence of the PfaC protein (Figure 30A, lanes 2-8), This could be attributed to either cleavage of the linker between HisMBP and PfaC by proteases or contamination with the template plasmid pIA-HisMBP.

Large-scale purification

The large-scale overexpression of PfaC-His was conducted at 18ºC, with an IPTG concentration of 0.2 mM, using the strain *ArcticExpress(DE3)* in 2 liters of LB. Purifications were performed following the steps described in the *Materials and methods* section.

Once overexpressed, the PfaC-His protein was purified through affinity chromatography with a HisTrap HP column. Despite the fact that a high overexpression was obtained, the majority of the protein is found in the insoluble fraction, the pellet. The amount of soluble protein obtained was very low. Affinity chromatography with the HisTrap column does not provide a pure PfaC sample as there are several lower molecular weight bands co-eluting with the PfaC-His fractions (219 kDa, Figure 29A-B). For this reason,

we conducted a second chromatographic separation, this time of the gel filtration type. The *Superdex S-200 16/60* column used for the gel filtration chromatography was able to separate the large (100-200 kDa) from the small (50-70 kDa) fragments in the sample, but it was not possible to separate the 219 kDa PfaC-His protein from the 100 kDa fragments (Figures 29C-D, Figure 28).

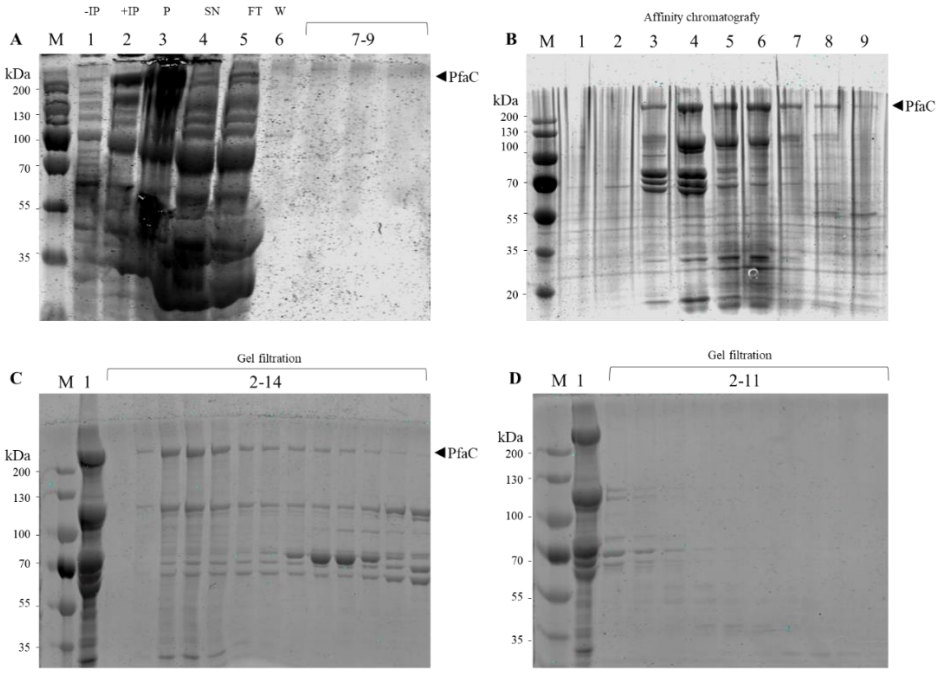


Figure 28. SDS-PAGE gel of the expression and purification of PfaC-His. Lane M shows the *PageRuler™ Plus Prestained Protein Ladder* (Thermo Fisher Scientific). **A.** Lane 1, -IPTG; 2, +IPTG; 3, pellet sample; 4, supernatant sample; 5, flowthrough sample; 6, wash sample; lanes 7-8-9, *HisTrap HP* column elution fractions 25-30-48. **B.** Lane 1, *HisTrap HP* column elution fraction 8; lane 2, Fr. 3; lane 3, Fr. 5; lane 3, Fr. 7; lane 4, Fr. 9; lane 5, Fr. 11; lane 6, Fr. 13; lane 7, Fr. 15; lane 8, Fr. 17; lane 9, Fr. 19. **C.** Lane 1, injection; lanes 2-14, *Superdex S-200 16/60* column elution fractions 4-16. **D.** Lane 1, injection; lanes 2-11, *Superdex S-200 16/60* column elution fractions 18-30.

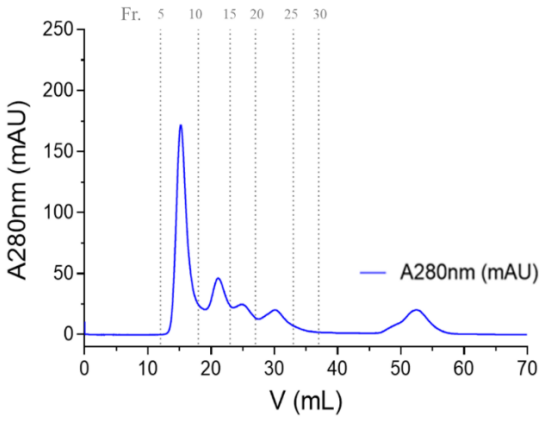


Figure 29. Chromatogram with *Superdex S-200 16/60* column from gel filtration shown in blue.

Among the degradation products obtained in this experiment, it is worth highlighting the one corresponding to 100 kDa. We think that it could correspond to the DH domains of PfaC, since they are 100.3 kDa in size and the histidines are in the C-terminus, so this degradation product will have been

retained in the Ni-Indigo column. But more tests would be necessary to verify it (MS (MALDI-Q-TOF) or LC-MS/MS for example). Due to the challenges faced in purifying the full-length PfaC protein, including low solubility and not being able to separate it from other fragments, we decided to stop these experiments and focus on purifying only the DH domains of PfaC individually.

4. Expression and purification of $PfaC_{DH}$

As mentioned in the previous section, once the complete purification of PfaC was not achieved, we built the constructs containing only the PfaC DH domains ($PfaC_{DH}$, residues 1111-2011).

Expression tests

In the previous section the parameters and techniques for PfaC expression and purification were optimized. We followed the optimized parameters for the expression of the proteins GST-$PfaC_{DH}$-, HisMBP-$PfaC_{DH}$ and HisTRX-$PfaC_{DH}$. Therefore, the transformation was conducted to *BL21(DE3)* competent cells, performing out the expression at 18ºC with an IPTG concentration of 0.2 mM, following the steps described in the *Materials and methods* section. The lysis of the cells was performed by adding 1% Triton and incubating the cells on ice for 45 minutes.

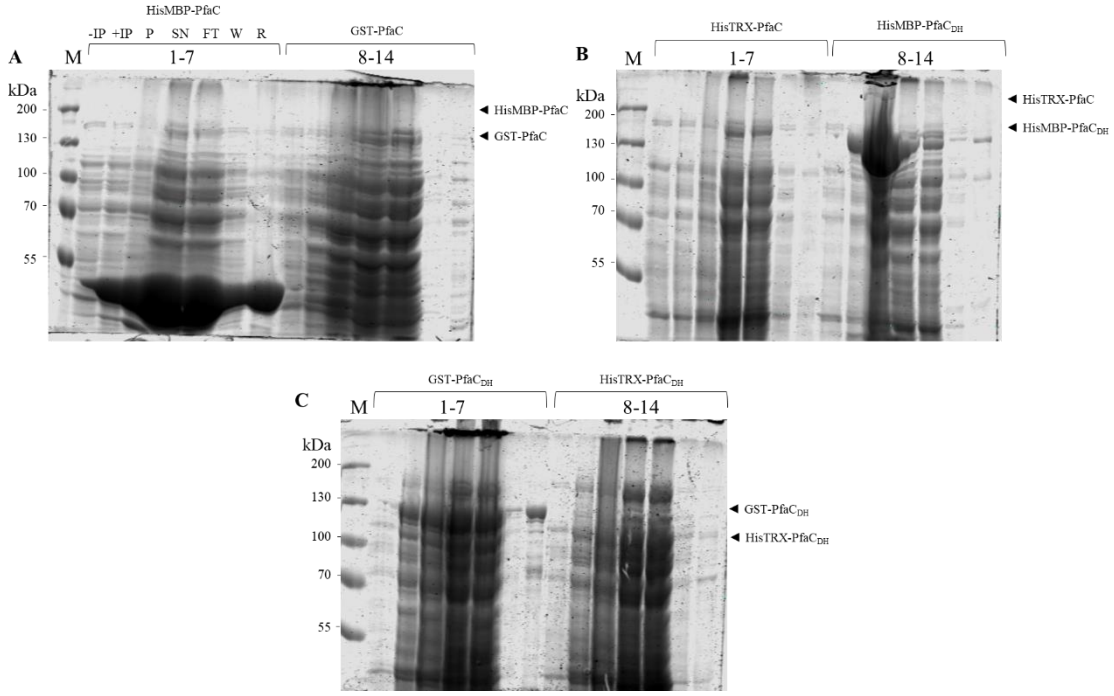


Figure 30. SDS-PAGE gel of the expression and purification of PfaC and $PfaC_{DH}$. Lane M shows the *PageRuler™ Plus Prestained Protein Ladder* (Thermo Fisher Scientific). In all experiments, the samples are ordered as follows: -IPTG, +IPTG, pellet, supernatant, flowthrough, wash, resin. **A.** Lanes 1-7, HisMBP-PfaC (258.6 kDa); lanes 8-14, GST-PfaC (243.9 kDa). **B.** Lanes 1-7, HisTRX-PfaC (230.1 kDa); lanes 8-14, HisMBP-$PfaC_{DH}$ (140.5 kDa). **C.** Lanes 1-7, GST-$PfaC_{DH}$ (125.8 kDa); lanes 8-14, HisTRX-$PfaC_{DH}$ (112.0 kDa).

The amount of protein retained in the GST-$PfaC_{DH}$ resin is the highest of all those tested, but a considerable amount is also observed for HisMBP-$PfaC_{DH}$. The amount of HisTRX-$PfaC_{DH}$ bound to the resin was very low. (Figure 30B, lanes 8-14; Figure 30C).

Large-scale purification

The overexpression was conducted at 18ºC and with an IPTG concentration of 0.2 mM, taking a colony from the plates previously prepared for the expression tests of the plasmids GST-PfaC$_{DH}$ and HisMBP-PfaC$_{DH}$ in *BL21(DE3).* The protein was overexpressed, since an intense band at 125.8 kDa appears after IPTG induction, corresponding to the GST-PfaC$_{DH}$ protein (Figure 31A, lanes 1-2). Once expressed, protein purification was necessary for the crystallization trials. Both purifications were performed following the steps described in the *Materials and methods* section. The GST-PfaC$_{DH}$ protein was purified with affinity chromatography using glutathione agarose resin (CubeBiotech) and gel filtration of the protein after cutting the GST tag. Despite the fact that a high overexpression was obtained, protein is found in the insoluble fraction, the pellet. The amount of soluble protein obtained was low.

The TEV protease cut the tag as we can see from the difference in the bands corresponding to the samples from before and after adding the TEV protease and incubating overnight. A band can be observed around 130 kDa, corresponding to the PfaC$_{DH}$ protein bound to the GST tag (125.8 kDa); while a new band around 100 kDa is shown, corresponding to the PfaC$_{DH}$ protein (100.3 kDa) with the GST tag already cut (Figure 31A, lanes 7-8). It is also noteworthy that there is a band with very low concentration around 130 kDa corresponding to the resin sample after the addition of the TEV protease. This tells us that there is a protein attached to the GST that has not been cut and, therefore, has remained attached to the resin. At this point, the yield has slightly decreased because the protein remained in the insoluble fractions and did not attach to the resin (Figure 31A, lane 11).

From the second affinity chromatography, after adding the TEV protease, three samples were extracted: flowthrough, wash 1 and wash 2 (Figure 31A, lanes 8-10). The concentration of these samples was measured using the wash elution buffer as a blank and concentrations of 0.60 mg/mL, 0.19 mg/mL and 0.04 mg/mL were obtained, respectively. Glutathione affinity chromatography does not provide a pure PfaC$_{DH}$ sample as there are several lower molecular weight bands co-eluting with the PfaC$_{DH}$ fractions (100.3 kDa, Figure 31A-B). The three samples mentioned before were pooled and concentrated to 500 µL to a final concentration of 12.98 mg/mL for its purification by gel filtration with the *Superose 6 Increase 10/300 GL* column (Figure 31B, lane 3).

In addition, the yield has also dropped when centrifuging the concentrated protein sample before injecting it for gel filtration. In the laboratory, the formation of a precipitate was observed which, when centrifuged, remained in the form of a pellet. To check what had precipitated, we resuspended the pellet in 40 µL of 1% SDS and loaded it onto the gel (Figure 31B, lanes 1-2). Several bands can be seen, including a band at 100 kDa, telling us thar part of the PfaC$_{DH}$ protein degraded.

Table 7. Measured concentrations of fractions 20-30 extracted on gel filtration.

Fraction	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
Concentration (mg/mL)	0.155	0.207	1.178	2.959	1.123	0.763	0.196	0.122	0.144	0.269	0.104

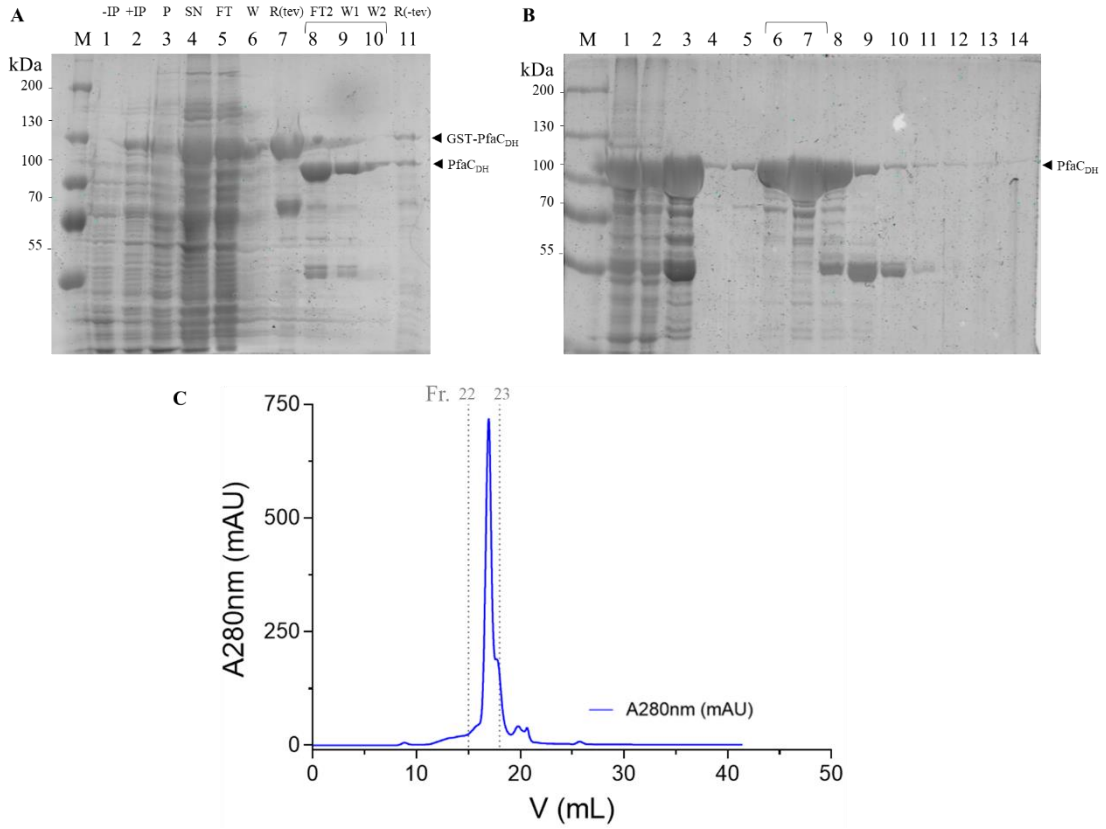


Figure 31. SDS-PAGE gel of the expression and purification of GST-PfaC$_{DH}$ and chromatogram from purification by gel filtration. Lane M shows the *PageRuler™ Plus Prestained Protein Ladder* (Thermo Fisher Scientific). **A.** Lane 1, -IPTG; 2, +IPTG; 3, pellet sample; 4, supernatant sample; 5, flowthrough sample; 6, wash sample; 7, resin sample before adding TEV protease (125.8 kDa); 8, flowthrough 2 (after TEV protease, 100.3 kDa); 9, wash 1 (after TEV protease); 10, wash 2 (after TEV protease); 11, final resin sample. **B.** Lanes 1-2, pellet sample after centrifugation of protein; lanes 3, injection; lanes 4-14, *Superose 6 Increase 10/300 GL* column elution fractions 20-30. **C.** Chromatogram from gel filtration shown in blue. Fractions 22 and 23 were concentrated for the crystallization trials.

The concentrations of the fractions corresponding to the region of interest in the gel filtration chromatogram were measured (Table 7, the two samples that were concentrated for future crystallization have been highlighted in green). For these crystallization trials, fraction 22 was brought to a final concentration of 1.20 mg/mL, while fraction 23 reached 3.20 mg/mL. It is important to note that in both cases protein formed aggregates that precipitated.

For the purification of the HisMBP-PfaC$_{DH}$ protein, two affinity chromatographies were performed (first without the addition of TEV protease, and another after its addition and dialysis overnight) and a subsequent gel filtration. The protein has been overexpressed, since a new band at 140.5 kDa appears, corresponding to the HisMBP-PfaC$_{DH}$ protein (Figure 32A).

There was a high concentration of protein bound to the HisMBP tag in the pellet, reducing the total yield of final protein that we obtained (Figure 32A).

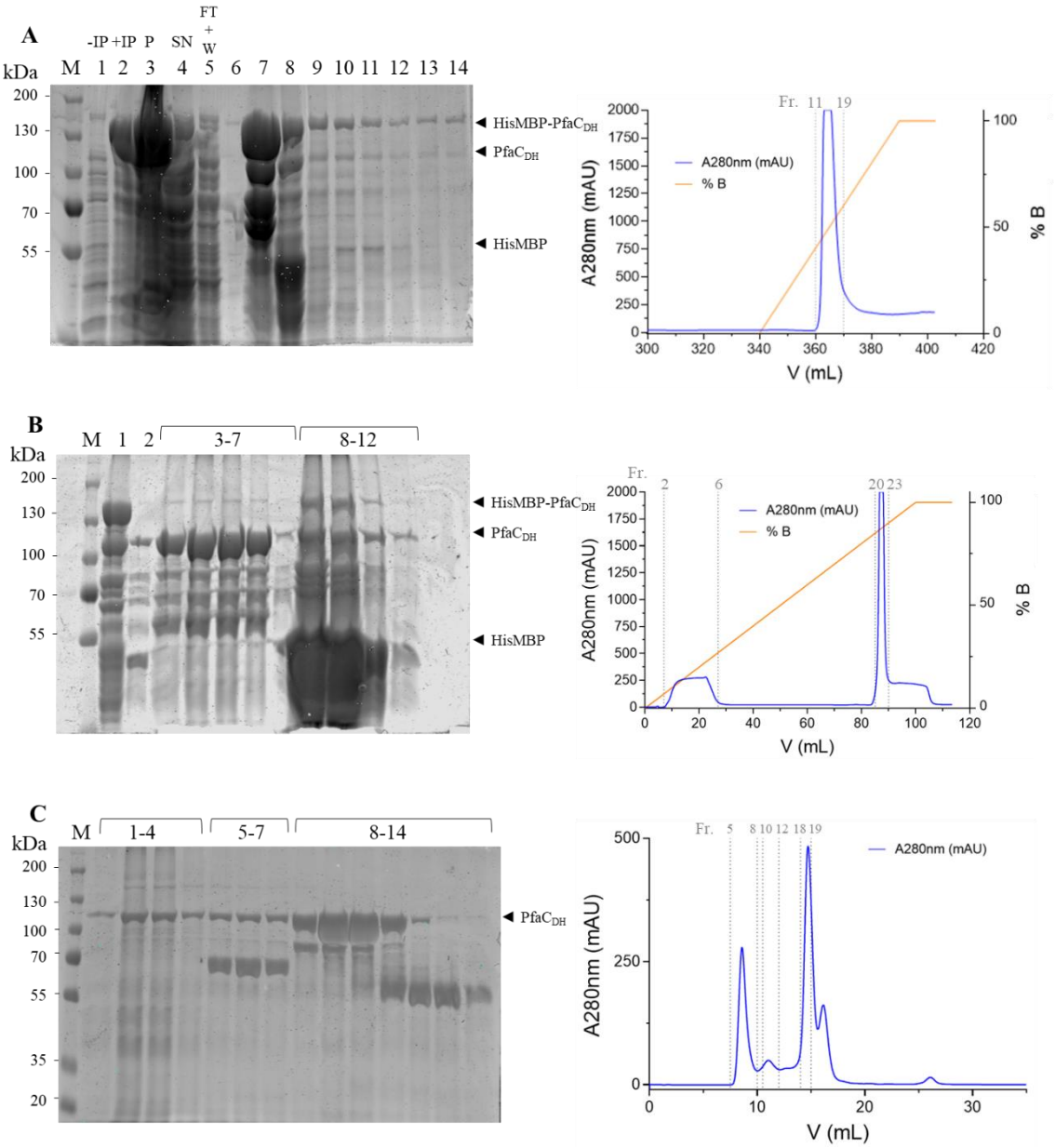


Figure 32. SDS-PAGE gel of the expression and purification of HisMBP-PfaC$_{DH}$; chromatograms by affinity chromatography (A, B) and gel filtration (C). Lane M shows the *PageRuler™ Plus Prestained Protein Ladder* (Thermo Fisher Scientific). **A.** Lane 1, -IPTG; 2, +IPTG; 3, pellet sample; 4, supernatant sample; 5, flowthrough+wash; lanes 6-14, *HisTrap HP* column elution fractions 11-19. Chromatogram from affinity chromatography, absorbance shown in blue and buffer B in orange (imidazole gradient from 20 mM to 400 mM). **B.** Lane 1, sample before adding TEV protease; 2, sample before adding TEV protease and incubate overnight; lanes 3-7, *HisTrap HP* column elution fractions 2-6; lanes 8-14, *HisTrap HP* column elution fractions 20-23. Chromatogram from affinity chromatography, absorbance shown in blue and buffer B in orange (imidazole gradient from 20 mM to 400 mM). **C.** Lanes 1-4, *Superdex 200 Increase 10/300 GL* column elution fractions 5-8; lanes 5-7, *Superdex 200 Increase 10/300 GL* column elution fractions 10-12; lanes 8-14, *Superdex 200 Increase 10/300 GL* column elution fractions 17-23. Chromatogram from gel filtration shown in blue. Fractions 10-12 and 18-19 were concentrated for the crystallization trials.

The first chromatography is prior to the addition of TEV protease, so the bands correspond to HisMBP-PfaCDH, (140.5 kDa). Comparing the data from the SDS-PAGE gel with the chromatogram, bands of very high concentration are observed corresponding to fractions 12-13 extracted from the chromatography. The chromatogram confirms that they have a high concentration, as a saturation peak is shown, exceeding 2000 mAU (maximum value measured by the detector). Fractions 12-18 were pooled, and TEV protease added. Dialysis was conducted overnight, and the next day, affinity chromatography was performed again. There was a difference between the sample taken before adding TEV protease and after adding it and performing dialysis (Figure 32B, lanes 1-2). In the first case, a band at 140.5 kDa is observed, corresponding to the PfaCDH protein attached to the HisMBP tag; while in the second case two differentiated bands are seen, one at 100.3 kDa of the PfaCDH and another one at 40.2 kDa of the loose HisMBP tag (Figure 32B). It should be noted that a stain is observed in fractions 20-23, indicating that there is a high concentration of MBP tag (40.2 kDa). This is confirmed by comparing the SDS-PAGE gel with the chromatogram to its right, as these fractions saturated the detector, exceeding 2000 mAU (maximum value measured by the detector) (Figure 32B, lanes 8-12).

Fractions 2-6 extracted from the second affinity chromatography were concentrated to 500 µL to a final concentration of 12.13 mg/mL for its purification by gel filtration with the Superdex 200 Increase 10/300 GL column (Figure 32C). Although various bands can be seen, the one with the highest concentration compared to the others is the one that corresponds to 100.3 kDa, the PfaCDH protein (Figure 32B). For the crystallization tests, fractions 10-12 were pooled and brought to a final concentration of 1.90 mg/mL, while with fractions 18-19 it was possible to reach 3.70 mg/mL. It is important to note that, in both cases, protein was lost upon concentration due to the formation of aggregates that precipitated.

Experimentation has led us to the conclusion that the PfaC protein, as well as its individual DH domains, are structurally unstable. This is probably due to the lack of interaction between the domains, which causes the protein to break into its domains or the subunits of the same domain. In addition, in several steps of the lysis and purification, we obtained aggregates and precipitates, causing even lower yield when obtaining the protein.

5. Estimation of the oligomeric states of $PfaC_{DH}$

The results obtained in the purifications indicate that the DH domains of PfaC can present several oligomeric states. The molecular weight, and with this, the oligomeric state of each of the results obtained in the fractions of the purification by gel filtration were estimated using a calibration curve calculated with molecular weight standards (Figure 33).

When analyzing the results of the DH domains that were purified through gel filtration containing the GST tag, with the *Superose 6 Increase 10/300 GL* column. A single peak was obtained which, according to calculations, corresponds to a protein fragment of 117.29 kDa (Table 8). Experimentally, a size slightly greater than 100 kDa is also seen on the SDS-PAGE gel. The protein in this case eluted according to its theoretical molecular weight. Therefore, we can conclude that the DH domains of PfaC in this case behave as a monomer, with the two subunits that compose it joined. (Figure 33B).

Table 8. Calculation of the theoretical and experimental sizes of $PfaC_{DH}$ and prediction of their oligomeric states.

Tag of purification	Elution volume (mL)	K_{av}	Theoretical size (kDa)	Experimental size (kDa)	Oligomerization state
GST	16.96	0.546	101.22	117.29	Monomer
HisMBP	8.59	0.006	-	2151.61	Aggregate
	11.04	0.164	-	649.34	GroEL contamination
	14.71	0.401	101.22	107.92	Monomer
	16.12	0.492	56.68	54.16	DH domain

In the case of the results of the gel filtration performed on the DH domains of the PfaC protein containing the HisMBP tag, we see four main peaks (Table 8).

- Peak I, with an elution volume of 8.59 mL, a protein size of 2151.61 kDa is obtained when calculating its experimental size with the calibration curve. This indicates that the protein has aggregated.
- Peak II, with an elution volume of 11.04 mL, and an experimental size of 649.34 kDa. This may be due to the formation of hexamers of the entire DH domain or dodecamers of subunits of the DH domain (607.32 kDa). This hexameric structure has already been described for the 'hot dog' fold in the FabZ (Kimber et al., 2004).
- Peak III with an elution volume of 14.71 mL, we obtain the complete DH domains, since a protein fragment with a theoretical size of 101.22 kDa and an experimental size of 107.92 kDa is obtained. Furthermore, we can conclude that it behaves as a monomer.
- Peak IV, with an elution volume of 16.12 mL, we obtain a protein fragment with a and an experimental size of 54.16 kDa. By checking the sequence of the DH domains of the PfaC protein, we have confirmed that it corresponds exactly to one of the subunits of this DH domain (theoretical size of 56.68 kDa).

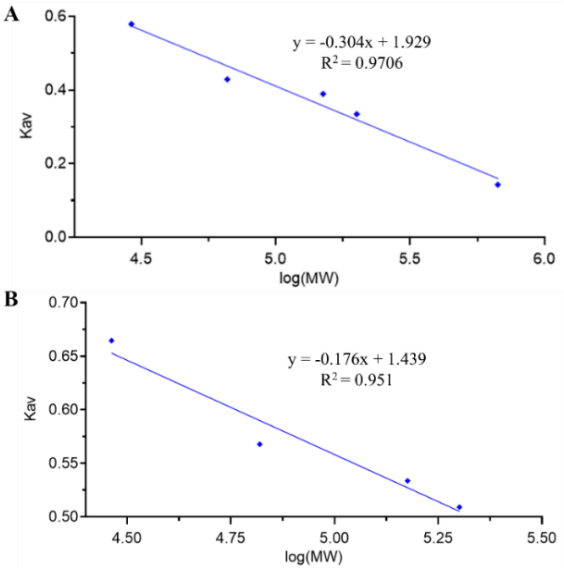


Figure 33. Plots of the average distribution constant (K_{av}) vs logarithm of the molecular weights. $K_{av} = (v_e - v_0)/(v_t - v_0)$; where V_e: elution volume, V_0: column void volume (8,5 mL), V_t: column volume (24 mL). Proteins used for the calibration curve are represented with diamonds. **A**. Calibration curve for *Superdex 200 Increase 10/300 GL* column. **B**. Calibration curve for *Superose 6 Increase 10/300 GL* column.

To confirm our results, we sent samples 2, 3 and 4, boxed in red in Figure 34A, to be analyzed with LC-MS/MS at the Proteomics Unit of the UPV/EHU. The results of the analysis with LC-MS/MS confirm what we had predicted from the experimental results. In case 2 there has been contamination with the *E.coli* chaperonin GroEL. However, the results still allow us to determine that in this case the protein is complete, with all the PfaC DH domains attached, since the 14 unique peptides found match the beginning and end of the sequence (Supplementary FigureS8A). In case 3, the complete $PfaC_{DH}$ protein is also obtained. In this case, 48 unique peptides were found. The identified sequence goes almost to the end with many peptides identified in that area so it can be assumed that the entire protein is there (Supplementary FigureS8B). In the last case, 28 unique peptides were obtained, all corresponding to the initial part of $PfaC_{DH}$. When checking the structure and the sequence, we saw that the cut has occurred between the two DH domains, separating them and purifying the first one (residues 1111-1616) (Supplementary FigureS8C). In the supplemental information the results of the unique peptides found in each case are shown, as well as a representation of the structure and the separation between domains.

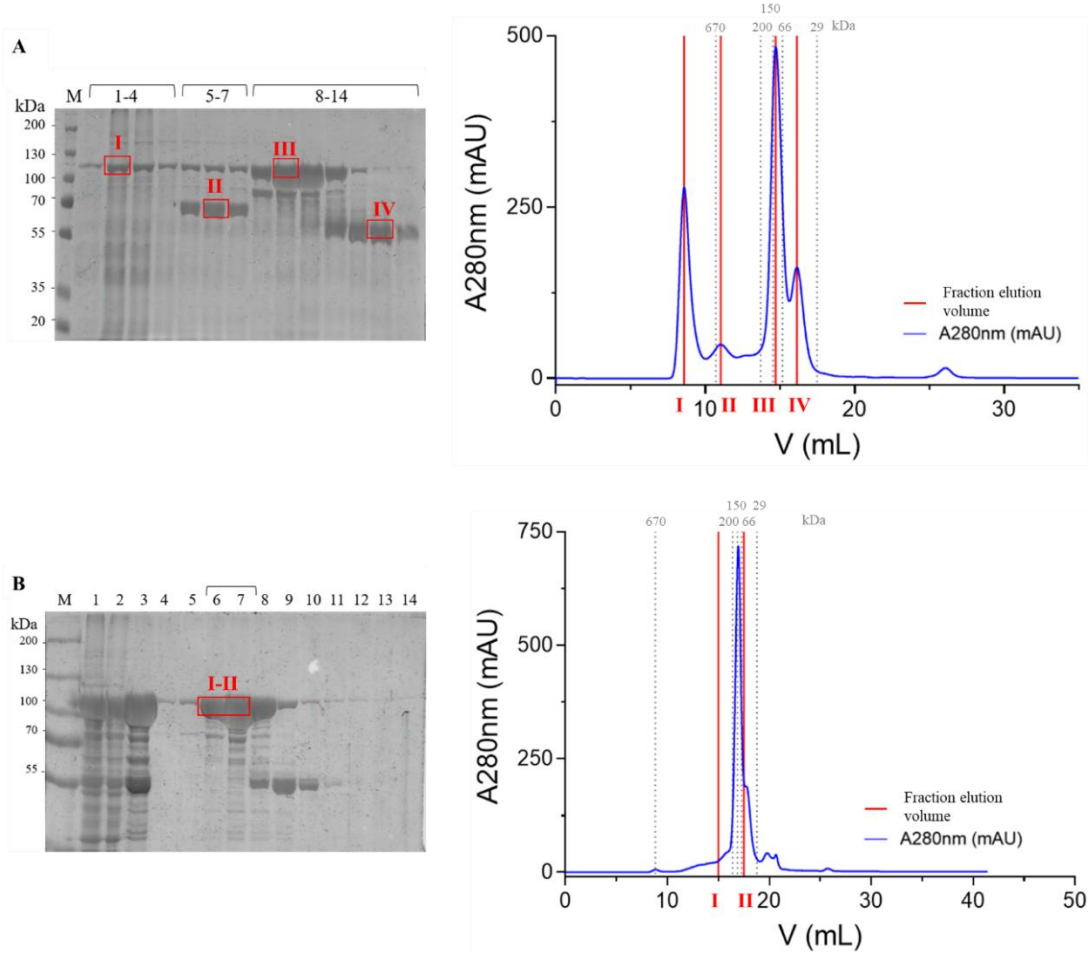


Figure 34. Chromatograms and SDS-PAGE gels of purification of $PfaC_{DH}$ by gel filtration chromatography. Marked with vertical lines in gray are the expected elution volumes for the indicated molecular sizes, calculated with the molecular weight standards. The red vertical lines indicate the elution volumes of the analyzed samples, also boxed in red on the SDS-PAGE gel. A. Gel filtration results for HisMBP-$PfaC_{DH}$ purification with *Superdex 200 Increase 10/300 GL* column. B. Gel filtration results for GST-$PfaC_{DH}$ purification with *Superose 6 Increase 10/300 GL* column.

6. Crystallization trials of $PfaC_{DH}$

In order for a protein to crystallize and generate diffraction-quality crystals, it is necessary to have a protein of very high purity (at least 90%) and concentration (depending on the crystallization kit used, 5-20 mg/mL).

The Pre-Crystallization Test was performed to select the appropriate protein concentration for crystallization screens, particularly Crystal Screen (Hampton Research). Samples that are too concentrated can result in an amorphous precipitate, while samples that are too dilute can result in clear droplets. Precipitate and clear droplets are typical crystallization screen results for reagent conditions that do not promote crystallization and are a part of each crystallization screen. However, by optimizing the protein concentration for the screen, the number of clear and precipitated results can often be reduced, which in turn results in more efficient sample utilization while simultaneously improving the chances of crystallization. In our case, with all concentrations tested, we obtained drops with light granular precipitate with PCT reagent A1/B1 and clear drops with PCT reagent A2/B2 (Figure 35A-D). As indicated by the PCT commercial house, these conditions are adequate to proceed with a screen.

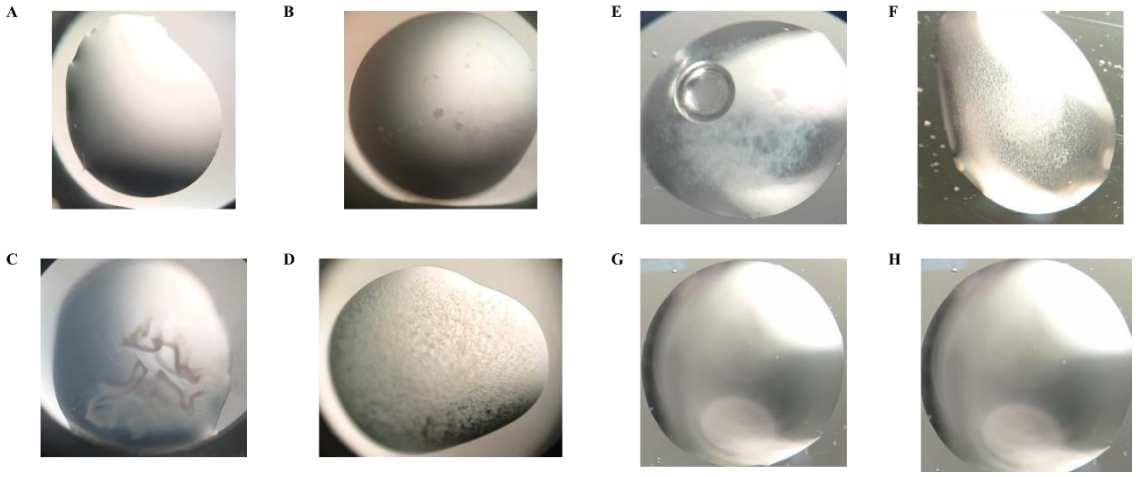


Figure 35. Images with examples of results obtained in crystallization trials and examples of results obtained in crystallization trials. **A-B.** Light granular precipitate. **C**-**D**. Clear drops. **E.** Clear drop. **F**. Aggregate formation. **G**. Precipitate formation. **H**. Crystalline precipitate formation.

Various concentrations of purified protein and various available screening kits were used: Crystal Screen (Hampton Research), ProPlex, JCSG-plus and Morpheus (Molecular Dimensions). All of these kits contain a wide variety of reagents, precipitants, and pH. Plates were prepared following the protocol described in the Materials and Methods section. The samples used for the crystallization tests and the kit used in each case are summarized in Table 9.

Unfortunately, in none of the kits and with none of the protein concentrations did we obtain any protein crystals. Different results were obtained during the crystallization trials. All these results were events of phase separation, formation of precipitates, crystalline precipitates or aggregates (Figure 35E-H).

Table 9. Concentrations of $PfaC_{DH}$ used for crystallization screening. In brackets volume ratio (protein:buffer).

Purification kit	Protein sample (purification tag)	Concentrations (mg/mL)
Crystal Screen	$PfaC_{DH}$ (GST)	1.20 / 3.20
	$PfaC_{DH}$ (GST)	6.70 (1:1) / 6.70 (2:1)
Proplex	$PfaC_{DH}$ (GST)	1.20 / 3.20
Morpheus	$PfaC_{DH}$ (HisMBP)	1.90 / 3.70
JCSC+	$PfaC_{DH}$ (HisMBP)	1.90 / 3.70

Discussion and Future Research

Polyunsaturated fatty acid synthases are large enzymatic complexes able of producing polyunsaturated fatty acids (PUFAs) such as eicosapentaenoic acid (EPA) or docosahexaenoic acid (DHA). Both EPA and DHA are classified as omega-3 fatty acids and are essential for human health and nutrition (Liu et al., 2023). *Moritella marina* possesses the gene cluster, necessary for their production, which consists of the pfa genes: *pfaA, pfaB, pfaC,* and *pfaD* (Okuyama et al., 2007). The domains in all the proteins encoded by these genes have been identified. However, in-depth studies of their 3D structures have not been conducted. So far, only the structure of the KS domain of PfaC has been solved for the Pfa proteins from *Moritella marina* (Santin et al., 2020). This work has focused on the structural study and production of the dehydratase (DH) domain, as well as the complete PfaC protein.

The computational modeling of the structure of PfaC of *Moritella marina* in *AlphaFold2* reveals that its two DH domains present a 'hot dog' fold, a characteristic feature of dehydratases associated with fatty acid biosynthesis and metabolism (Pidugu et al., 2009). We have also verified that the position of the residues that are part of the active site is conserved when compared to other proteins with 3-hydroxyacyl-ACP dehydratase (FabA and FabZ) function (Moynié et al., 2016; Kimber et al., 2004). Additionally, our analysis determines that there are no interactions between the domains present in PfaC, nor between the subunits of the DH domain.

It has been previously proven that PfaC can be produced in the *Rosetta (DE3)* strain of *E.coli* (Santin et al., 2020). In our study, we further confirmed that its production in the strains *ArcticExpress(DE3)* and *BL21(DE3)* also allows the expression of PfaC at similar levels. Additionally, we independently cloned the DH domains of PfaC with three different tags, overexpressed and purified them, improving their solubility. With the purified $PfaC_{DH}$ protein, we conducted several crystal screenings to find the optimal crystallization conditions. Despite our efforts, we did not obtain any crystals.

With the experimental data obtained in the laboratory and the results extracted from the analysis by LC-MS/MS, we performed a study of the oligomeric state of the PfaC's DH domains. We observed that $PfaC_{DH}$ behaves as a monomer with two DH subunits. Each of the subunits has a similar structure to the one described for FabA (Moynié et al., 2016.). From the *AlphaFold2* prediction and the experimental results we can deduce that there is no strong interaction between the DH1 and DH2 subunits of the dehydratase domain. This explains the different degradation products experimentally observed by LC-MS/MS in section 5 of the results.

Therefore, all this information has led us to the conclusion that the PfaC protein, as well as its individual DH domains, are structurally unstable when overproduced in *E. coli*. It is important to highlight that the presence of two or three dehydratase domains in PfaC from *Moritella marina* was predicted by Okuyama et al., 2007, with structures homologous to those already described for FabA or FabZ (Moynié et al., 2016; Kimber et al., 2004). However, in this work we have confirmed what was previously stated by Santin et al., 2020. PfaC undergoes a separation of their subunits, forming DH'-DH didomains in which the first of the modules (DH') lacks the active site. Each DH'-DH didomain remains united forming a single

structural unit with a 'hot dog' type folding, similar to the one described in *Photobacterium profundum* where the domains follow a DH1'-DH1- DH2'-DH2 pattern (Oyola-Robles et al., 2013). Therefore, it would be interesting for future research to clone each of the DH'-DHs, called DH1 and DH2 in this work, separately. This could improve the purification of the protein since we expect no degradation products. Once each subunit is purified, we can proceed to study its 3D structure through x-ray crystallography.

In the structural comparisons made in this work as well as in the referenced literature, it is evident that polyunsaturated fatty acid synthases and fatty acid synthases share a common structural organization of their domains. This supports the idea that both systems may have similar functions on a biochemical level. However, biochemical studies of these PfaC have not been conducted to provide an in-depth understanding of their role in fatty acid synthesis. As future research, it would be necessary to conduct studies of the enzymatic activity presented by these DH domains. To this end, for example, mutagenesis assays could be conducted in the active site described for PfaC.

The computational modelling and the oligomeric state study of PfaC made in this work represent an advance on the understanding of its role in the omega-3 fatty acid synthesis pathway conducted by *Moritella marina* and other marine bacteria

Conclusions

The computational modeling of the structure of the DH domains of PfaC of *Moritella marina* in *AlphaFold2* confirms that:

- PfaC consists of KS and DH domains joined by a flexible linker with no apparent interaction between the different domains.
- The two DH domains of PfaC present a 'hot dog' fold, characteristic of proteins associated with fatty acid biosynthesis and metabolism.
- PfaC has the potential to exhibit a dehydratase function because the position of the residues that are part of the active site is conserved when compared to other proteins with dehydratase function (FabA and FabZ).

From the experimental results, we can conclude that:

- The overexpression of PfaC in the *ArcticExpress(DE3)* strain at low temperatures (12ºC) and with low IPTG concentrations (0.2 mM) enhances its production.
- The addition of the purification tags HisMBP and GST improves $PfaC_{DH}$ solubility.
  $PfaC_{DH}$ behaves as a monomer in the conditions tested.
- In summary, the computational modeling of PfaC's DH domains has shed light on its structure and function. By improving solubility, stability, and purification techniques, as well as determining the complete structure using experimental methods, in a near future, researchers could gain a deeper understanding of PfaC's role in the synthesis of omega-3 fatty acids and its potential applications in producing valuable fatty acids.

References

Abedi, E., & Del Rio, D. (2014). Long-chain polyunsaturated fatty acid sources and evaluation of their nutritional and functional properties. *Food Science and Nutrition*, *2*(5), 443–463. https://doi.org/10.1002/fsn3.121

Alberts, A. W., Majerus, P. W., Talamo, B. R., & Vagelos, P. R. (1964). Acyl-Carrier Protein. II. Intermediary Reactions of Fatty Acid Synthesis. *Biochemistry*, *3*(10), 1563–1571. https://doi.org/10.1021/bi00898a030

Allen, E., & Bartlett, D. H. (2002). Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium Photobacterium profundum strain SS9 The GenBank accession numbers for the sequences reported in this paper are AF409100 and AF467805. *Microbiology*, *148*(6), 1903–1913. https://doi.org/10.1099/00221287-148-6-1903

Asturias, F. J. (2006). Mammalian Fatty Acid Synthase: X-ray Structure of a Molecular Assembly Line. *ACS Chemical Biology*, *1*(3), 135–138. https://doi.org/10.1021/cb6001448

Benatti, P., Peluso, G., Nicolai, R., & Calvani, M. (2004). Polyunsaturated Fatty Acids: Biochemical, Nutritional and Epigenetic Properties. *Journal of the American College of Nutrition*, *23*(4), 281–302. https://doi.org/10.1080/07315724.2004.10719371

Chirala, S. S., Jayakumar, A., Gu, Z., & Wakil, S. J. (2001). Human fatty acid synthase: Role of interdomain in the formation of catalytically active synthase dimer. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(6), 3104–3108. https://doi.org/10.1073/pnas.051635998

Diao, J., Song, X., Guo, T., Wang, F., Chen, L., & Zhang, W. (2020). Cellular engineering strategies toward sustainable omega-3 long chain polyunsaturated fatty acids production: State of the art and perspectives. *Biotechnology Advances*, *40*, 107497. https://doi.org/10.1016/j.biotechadv.2019.107497

Gemperlein, K., Rachid, S., Garcia, R. G., Wenzel, S. C., & Müller, R. (2014). Polyunsaturated fatty acid biosynthesis in myxobacteria: different PUFA synthases and their product diversity. *Chemical Science*, *5*(5), 1733. https://doi.org/10.1039/c3sc53163e

Giordano, D., Coppola, D., Russo, R., Denaro, R., Giuliano, L., Lauro, F. M., Di Prisco, G., & Verde, C. (2015). Marine Microbial Secondary Metabolites. In *Elsevier eBooks* (pp. 357–428). https://doi.org/10.1016/bs.ampbs.2015.04.001

Holub, B. J. (2002). Clinical nutrition: 4. Omega-3 fatty acids in cardiovascular care. *PubMed*, *166*(5), 608–615. https://pubmed.ncbi.nlm.nih.gov/11898942

Holub, D. J., & Holub, B. J. (2004). Omega-3 fatty acids from fish oils and cardiovascular disease. *Molecular and Cellular Biochemistry*, *263*(1/2), 217–225. https://doi.org/10.1023/b:mcbi.0000041863.11248.8d

Jiang, H., Zirkle, R., Metz, J. M., Braun, L., Richter, L., Van Lanen, S. G., & Shen, B. (2008). The Role of Tandem Acyl Carrier Protein Domains in Polyunsaturated Fatty Acid Biosynthesis. *Journal of the American Chemical Society*, *130*(20), 6336–6337. https://doi.org/10.1021/ja801911t

Jovanović, S., Dietrich, D., Becker, J. V., Kohlstedt, M., & Wittmann, C. (2021). Microbial production of polyunsaturated fatty acids — high-value ingredients for aquafeed, superfoods, and pharmaceuticals. *Current Opinion in Biotechnology*, *69*, 199–211. https://doi.org/10.1016/j.copbio.2021.01.009

Kimber, M. S., Martin, F. L., Lu, Y., Houston, S., Vedadi, M., Dharamsi, A., Fiebig, K. M., Schmid, M. B., & Rock, C. O. (2004). The Structure of (3R)-Hydroxyacyl-Acyl Carrier Protein Dehydratase (FabZ) from Pseudomonas aeruginosa. *Journal of Biological Chemistry*, *279*(50), 52593–52602. https://doi.org/10.1074/jbc.m408105200

Lee, J. D., O'Keefe, J. H., Lavie, C. J., Marchioli, R., & Harris, W. H. (2008). Omega-3 Fatty Acids for Cardioprotection. Mayo Clinic Proceedings, 83(3), 324–332. https://doi.org/10.4065/83.3.324

Liu, Y., Shen, N., Xin, H., Yu, L., Xu, Q., & Cui, Y. (2023). Unsaturated fatty acids in natural edible resources, a systematic review of classification, resources, biosynthesis, biological activities and application. *Food Bioscience*, 102790. https://doi.org/10.1016/j.fbio.2023.102790

Madeira, F., Pearce, M., Tivey, A., Basutkar, P., Lee, J. S., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*, *50*(W1), W276–W279. https://doi.org/10.1093/nar/gkac240

Maier, T., Leibundgut, M., & Ban, N. (2008). The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science*, *321*(5894), 1315–1322. https://doi.org/10.1126/science.1161269

Maier, T., Jenni, S., & Ban, N. (2006). Architecture of Mammalian Fatty Acid Synthase at 4.5 Å Resolution. *Science*, *311*(5765), 1258–1262. https://doi.org/10.1126/science.1123248

Marrakchi, H., Zhang, Y., & Rock, C. O. (2002). Mechanistic diversity and regulation of Type II fatty acid synthesis. *Biochemical Society Transactions*, *30*(6), 1050–1055. https://doi.org/10.1042/bst0301050

McGillick, B. E., Adler, M., Vieni, C., & Swaminathan, S. (2016). β-Hydroxyacyl-acyl Carrier Protein Dehydratase (FabZ) from Francisella tularensis and Yersinia pestis: Structure Determination, Enzymatic Characterization, and Cross-Inhibition Studies. *Biochemistry*, *55*(7), 1091–1099. https://doi.org/10.1021/acs.biochem.5b00832

Moi, I. M., Leow, T. C., Ali, M. T., Rahman, R. N. Z. R. A., Salleh, A. B., & Sabri, S. (2018). Polyunsaturated fatty acids in marine bacteria and strategies to enhance their production. *Applied Microbiology and Biotechnology*, *102*(14), 5811–5826. https://doi.org/10.1007/s00253-018-9063-9

Moynié, L., Hope, A., Finzel, K., Schmidberger, J. W., Leckie, S. M., Schneider, G., Burkart, M. D., Smith, A., Gray, D., & Naismith, J. H. (2016). A Substrate Mimic Allows High-Throughput Assay of the FabA Protein and Consequently the Identification of a Novel Inhibitor of Pseudomonas aeruginosa FabA. *Journal of Molecular Biology*, *428*(1), 108–120. https://doi.org/10.1016/j.jmb.2015.10.027

Nelson, D. R., & Cox, M. M. (2021). *Lehninger Principles of Biochemistry*, *7th ed*. http://dspace.vpmthane.org:8080/xmlui/handle/123456789/9026

Nichols, D. E., Bowman, J. L., Sanderson, K., Nichols, C., Lewis, T., McMeekin, T. A., & Nichols, P. D. (1999). Developments with Antarctic microorganisms: culture collections, bioactivity screening, taxonomy, PUFA production and cold-adapted enzymes. *Current Opinion in Biotechnology*, *10*(3), 240–246. https://doi.org/10.1016/s0958-1669(99)80042-1

Pidugu, L. S., Maity, K., Ramaswamy, K., Surolia, N., & Suguna, K. (2009). Analysis of proteins with the "hot dog" fold: Prediction of function and identification of catalytic residues of hypothetical proteins. BMC *Structural Biology*, *9*(1). https://doi.org/10.1186/1472-6807-9-37

Okuyama, H., Orikasa, Y., Nishida, T., Watanabe, K., & Morita, N. (2007). Bacterial Genes Responsible for the Biosynthesis of Eicosapentaenoic and Docosahexaenoic Acids and Their Heterologous

Expression. *Applied and Environmental Microbiology*, *73*(3), 665–670. https://doi.org/10.1128/aem.02270-06

Orikasa, Y., Nishida, T., Yamada, A., Yu, R., Watanabe, K., Hase, A., Morita, N., & Okuyama, H. (2006). Recombinant production of docosahexaenoic acid in a polyketide biosynthesis mode in Escherichia coli. *Biotechnology Letters*, *28*(22), 1841–1847. https://doi.org/10.1007/s10529-006-9168-6

Oyola-Robles, D., Gay, D. C., Trujillo, U., Sánchez-Parés, J. M., Bermúdez, M., Rivera-Díaz, M., Carballeira, N. M., & Baerga-Ortiz, A. (2013). Identification of novel protein domains required for the expression of an active dehydratase fragment from a polyunsaturated fatty acid synthase. *Protein Science*, *22*(7), 954–963. https://doi.org/10.1002/pro.2278

Rock, C. O., & Cronan, J. E. (1996). Escherichia coli as a model for the regulation of dissociable (type II) fatty acid biosynthesis. *Biochimica Et Biophysica Acta*, *1302*(1), 1–16. https://doi.org/10.1016/0005-2760(96)00056-2

Santín, O., Yuet, K. P., Khosla, C., & Moncalián, G. (2020). Structure and Mechanism of the Ketosynthase-Chain Length Factor Didomain from a Prototypical Polyunsaturated Fatty Acid Synthase. *Biochemistry*, *59*(50), 4735–4743. https://doi.org/10.1021/acs.biochem.0c00785

Schrödinger, L., & DeLano, W. (2020). *PyMOL*. Retrieved from http://www.pymol.org/pymol

Shulse, C. N., & Allen, E. (2011). Widespread Occurrence of Secondary Lipid Biosynthesis Potential in Microbial Lineages. *PLOS ONE*, *6*(5), e20146. https://doi.org/10.1371/journal.pone.0020146

Smith, S. F., Witkowski, A., & Joshi, A. (2003). Structural and functional organization of the animal fatty acid synthase. *Progress in Lipid Research*, *42*(4), 289–317. https://doi.org/10.1016/s0163-7827(02)00067-x

Tsai, S., & Ames, B. D. (2009). Chapter 2 Structural Enzymology of Polyketide Synthases. In *Elsevier eBooks* (pp. 17–47). https://doi.org/10.1016/s0076-6879(09)04602-3

Walsh, T. A., Bevan, S., Gachotte, D. J., Larsen, C. M., Moskal, W. A., Merlo, P., Sidorenko, L. V., Hampton, R. E., Virginia, S., Pareddy, D., Anthony, G., Bhaskar, P. B., Marri, P. R., Clark, L., Chen, W., Adu-Peasah, P. S., Wensing, S., Zirkle, R., & Metz, J. M. (2016). Canola engineered with a microalgal polyketide synthase-like system produces oil enriched in docosahexaenoic acid. *Nature Biotechnology*, *34*(8), 881–887. https://doi.org/10.1038/nbt.3585

White, S., Zheng, J., Zhang, Y. N., & Rock, C. O. (2005). The Structural Biology of Type II Fatty Acid Biosynthesis. *Annual Review of Biochemistry*, *74*(1), 791–831. https://doi.org/10.1146/annurev.biochem.74.082803.133524

Yazawa, K. (1996). Production of eicosapentaenoic acid from marine bacteria. *Lipids*, *31*(1), S297–S300. https://doi.org/10.1007/bf02637095

Yoshida, K., Hashimoto, M., Hori, R., Adachi, T., Okuyama, H., Orikasa, Y., Nagamine, T., Shimizu, S., Ueno, A., & Morita, N. (2016). Bacterial Long-Chain Polyunsaturated Fatty Acids: Their Biosynthetic Genes, Functions, and Practical Use. *Marine Drugs*, *14*(5), 94. https://doi.org/10.3390/md14050094

Zhang, M., Zhang, H., Li, Q., Gao, Y., Guo, L., He, L., Zang, S., Guo, X., Huang, J., & Li, L. (2021). Structural Insights into the Trans-Acting Enoyl Reductase in the Biosynthesis of Long-Chain Polyunsaturated Fatty Acids in Shewanella piezotolerans. *Journal of Agricultural and Food Chemistry*, *69*(7), 2316–2324. https://doi.org/10.1021/acs.jafc.0c07386

Supplemental Information

```
      1        10        20        30        40        50        60        70        80        90
PfaC  TNLTSTEAKAQAAATQAGFQIKGPVGYNYPPLQLIERYNKPENVIYDQADLVEFAEGDIGKVFGAEYNIIDGYSRRVRLPTSDYLLVTRV
FabZ  ..................................................................................................

            100       110       120       130       140       150       160       170       180
PfaC  TELDAKVHEYKKSYMCTEYDVPVDAPFLIDGQIPWSVAVESGQCDLMLISYIGIDFQAKGERVYRLLDCELTFLEEMAFGGDTLRYEIHI
FabZ  ..................................................................................................

            190       200       210       220       230       240       250       260       270
PfaC  DSYARNGEQLLFFFHYDCYVGDKKVLIMRNGCAGFFTDEELSDGKGVIHNDKDKAEFSNAVKSSFTPLLQHNRGQYDYNDMMKLVNGDVA
FabZ  ..................................................................................................

            280       290       300       310       320       330       340       350       360
PfaC  SCFGPQYDQGGRNPSLKFSSEKFLMIERITKIDPTGGHWGLGLLEGQKDLDPEHWYFPCHFKGDQVMAGSLMSEGCGQMAMFFMLSLGMH
FabZ  ....................HTLHIEEI..................................................................

            370       380       390       400       410       420       430       440       450
PfaC  TNVNNARFQPLPGESQTVRCRGQVLPQRNTLTYRMEVTAMGMHPQPFMKANIDILLDGKVVVDFKNLSVMISEQDEHSDYPVTLPSNVAL
FabZ  ..................................................................................................

            460       470       480       490       500       510       520       530       540
PfaC  KAITAPVASVAPASSPANSADLDERGVEPFKFPERPLMRVESDLSAPKSKGVTPIKHFEAPAVAGHHRVPNQAPFTPWHMFEFATGNISN
FabZ  ..................................................................................................

            550       560       570       580       590       600       610       620       630
PfaC  CFGPDFDVYEGRIPPRTPCGDLQVVTQVVEVQGERLDLKNPSSCVAEYYVPEDAWYFTKNSHENWMPYSLIMEIALQPNGFISGYMGTTL
FabZ  ..................................................................................................

            640       650       660       670       680       690       700       710       720
PfaC  KYPEKDLFFRNLDGSGTLLKQIDLRGKTIVNKSVLVSTAIAGGAIIQSFTFDMSVDGELFYTGKAVFGYFSGESLTNQLGIDNGKTTNAW
FabZ  ..................................................................................................

            730       740       750       760       770       780       790       800       810
PfaC  FVDNNTPAANIDVFDLTNQSLALYKAPVDKPHYKLAGGQMNFIDTVSVVEGGGKAGVAYVYGERTIDADDWFFRYHFHQDPVMPGSLGVE
FabZ  ............LDLLPHR.........FP.FLL.......VDRVLDFEE.....GKFLRAVKNVSFNEPFFQGHFPGKPIFPGVLILE

            820       830       840       850       860       870       880       890
PfaC  AIIELMQTYALKNDLGGKFAN..PRFIAPMTQVDWKYRGQITPLNKQMSLDVHITEIVNDAGEVRIVGDANLSKDGLRIYEVKNIVLSIV
FabZ  AMAQATGILAFKSR..GKLEPGELYYFAGIDEAR..FKRPVVPG.DQMIMEVEF..VKERRGLTRFTGVAKVDGEI..VCTATMMCAR..

      900
PfaC  EA
FabZ  ..
```

Figure 1. Sequence alignment of PfaC DH domains with FabZ from *Yersinia pesti.*

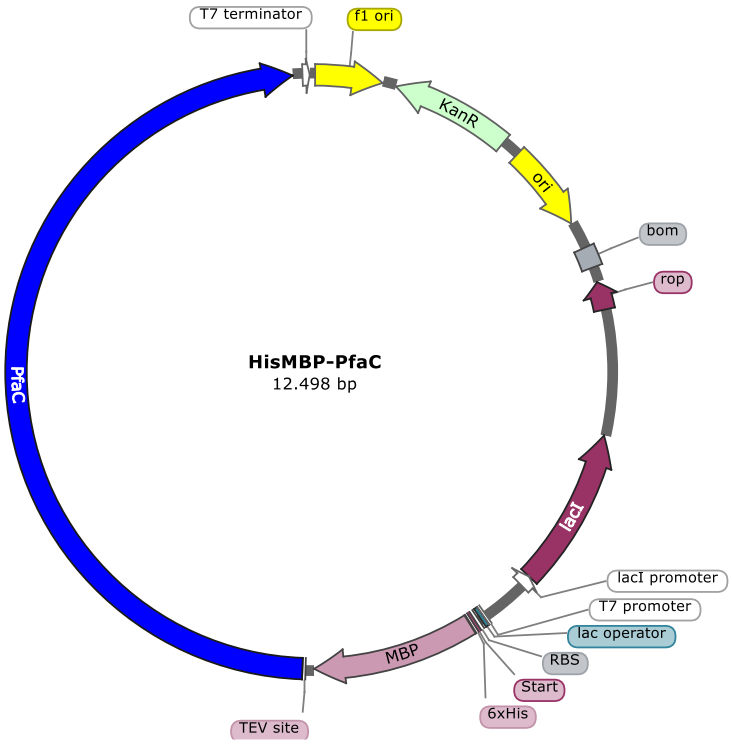


Figure 2. Expression vector pPI1 carrying the *pfaC* gene. Insert: HisMBP-PfaC.

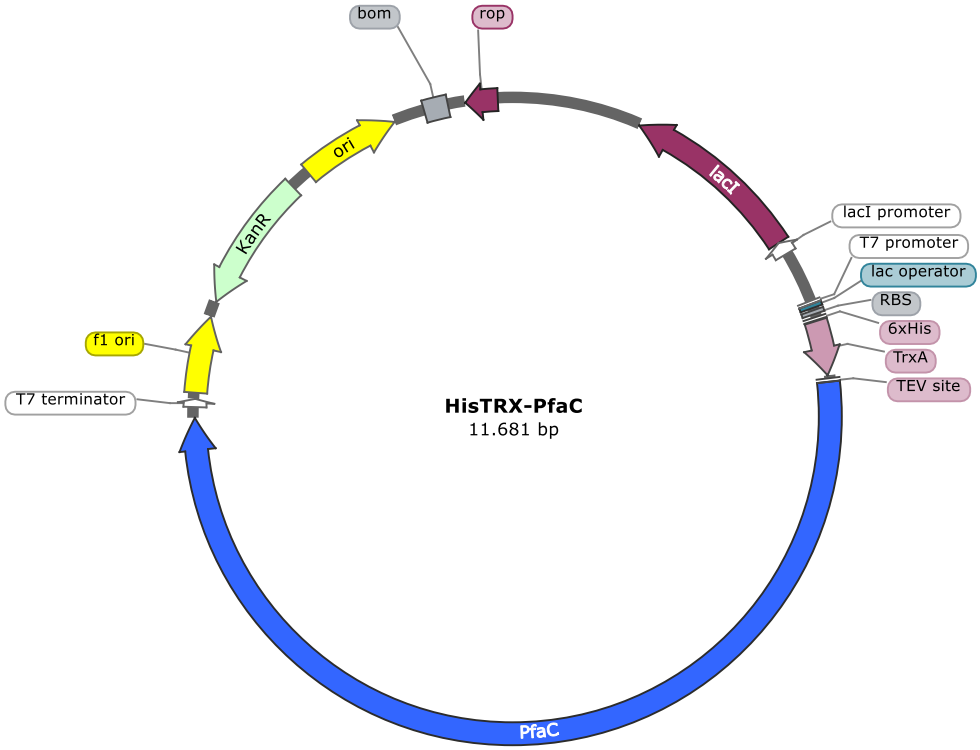


Figure 3. Expression vector pPI2 carrying the *pfaC* gene. Insert: HisTRX-PfaC.

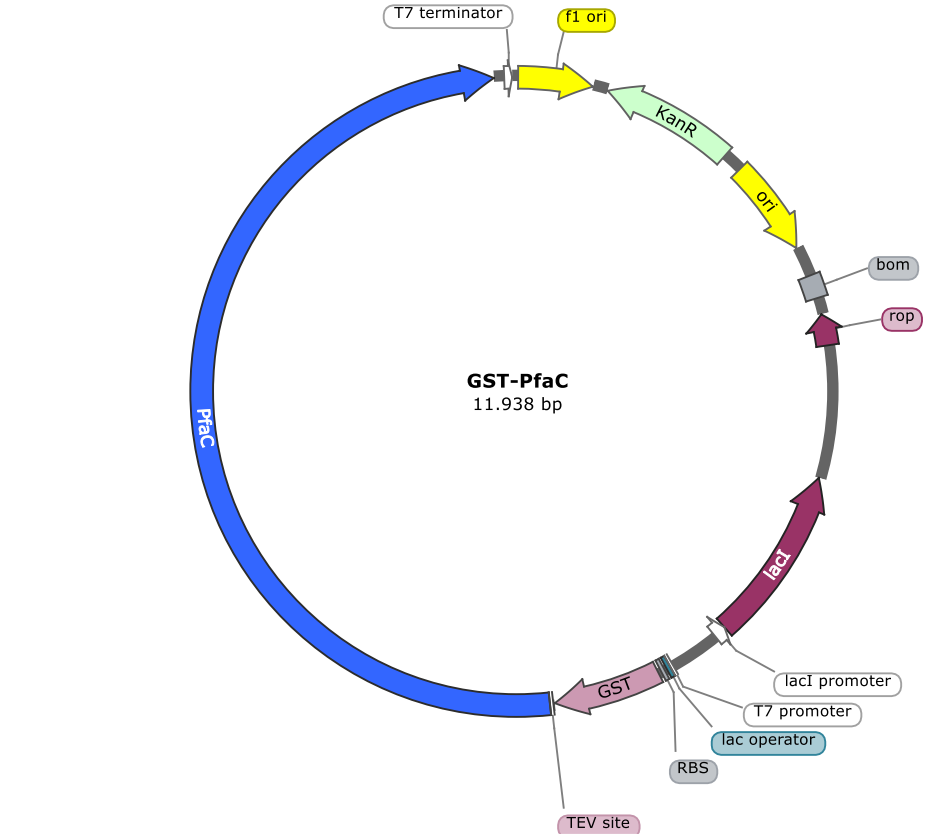


Figure 4. Expression vector pPI3 carrying the *pfaC* gene. Insert: GST-PfaC.

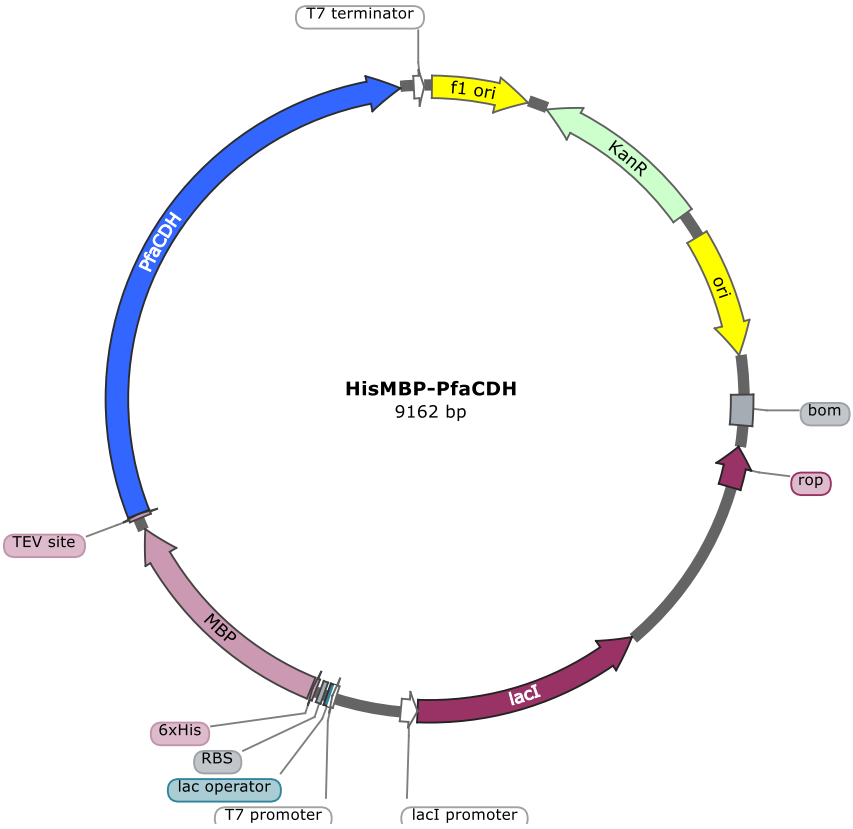


Figure 5. Expression vector pPI4 carrying the *pfaC* gene. Insert: HisMBP-PfaC$_{DH}$.

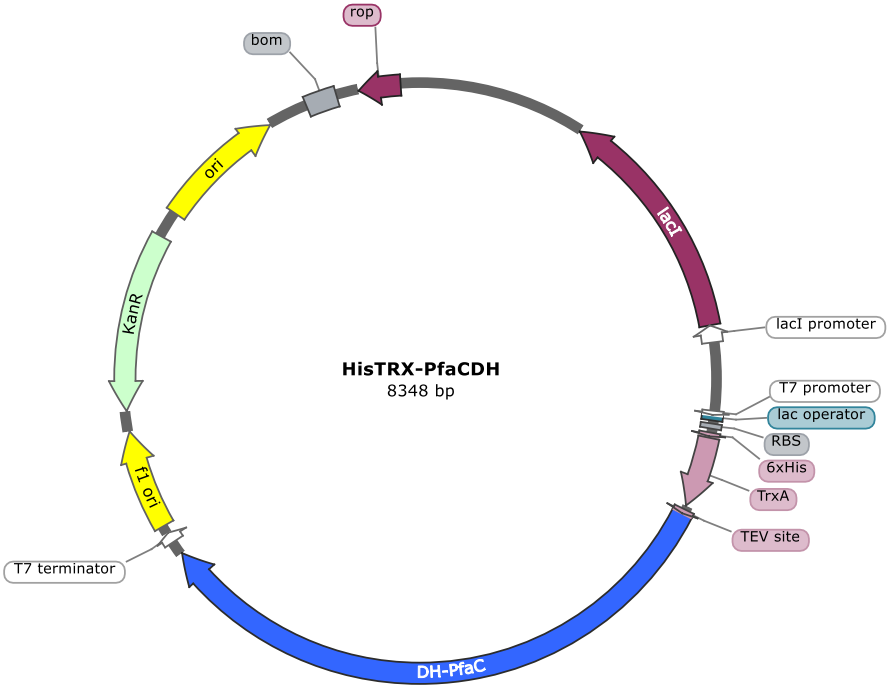


Figure 6. Expression vector pPI5 carrying the *pfaC* gene. Insert: HisTRX-PfaC$_{DH}$.

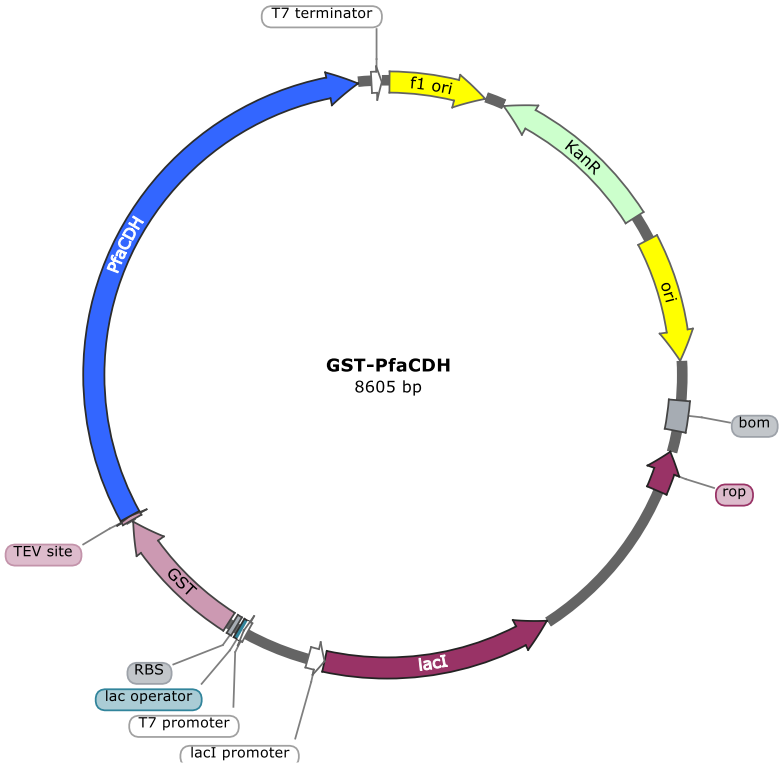


Figure 7. Expression vector pPI6 carrying the *pfaC* gene. Insert: GST-PfaC$_{DH}$.

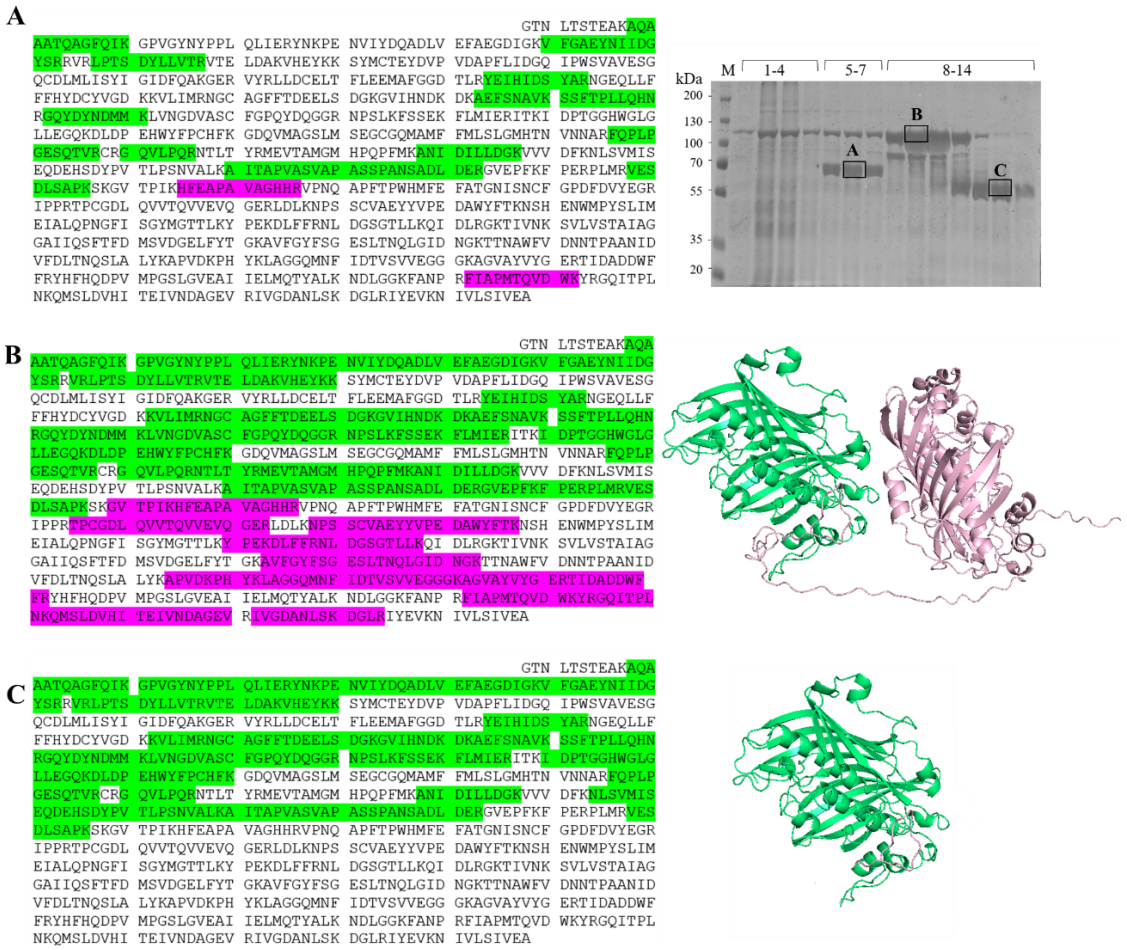


Figure 8. Gel filtration and MALDI-TOFF analysis of the PfaC-DH protein. Colored in green the residues from DH1 (1111-1629) and in pink from DH2 (1630-2011).