# Elucidation of the biochemical factors governing the enzymatic desulfonation of 6:2 fluorotelomer sulfonate: purification and enzymatic characterization of *Escherichia* and *Gordonia* monooxygenases

**By**

**Garret William O'Connell**

**Bachelor of Science, Honors in Biology and Chemistry – University of New Brunswick**

**A thesis submitted in partial fulfillment of the requirements for the degree of:**

**Master of Science in Environmental Science**

**In the Department of Biological Sciences**

**Thesis examining committee:**

**Jonathan Van Hamme (PhD), Full Professor (TRU) and Thesis Supervisor, Biological Sciences**
**Eric Bottos (PhD), Assistant Professor (TRU and Committee Member, Biological Sciences**
**Sharon Brewer (PhD), Associate Professor (TRU) and Committee Member, Physical Sciences**
**Andrea Franzetti (PhD), Associate Professor (UNIMIB) and External Examiner, Earth and Environmental Sciences**

**April 2020**
**Thompson Rivers University**

# I. Abstract

Fluorotelomer sulfonate degradation is thought to be a rate limiting step in the degradation of fluorinated surfactants in activated sludge wastewater treatment. Aliphatic sulfonates are structurally similar to fluorotelomer sulfonates and are partially degraded by alkanesulfonate monooxygenases. To understand the metabolism of 6:2 fluorotelomer sulfonate (6:2 FTSA), two nitrilotriacetate monooxygenases (ISGA 1218 and 1222) from *Gordonia* NB4-1Y, along with the *Escherichia coli* alkanesulfonate monooxygenase (SsuD) were cloned into the pMAL-c2 and pET28b protein production vectors, purified, given the *E. coli* flavin reductase (Fre) as a source of reduced flavin production and challenged *in vitro* with octane sulfonate and 6:2 FTSA. A combination of gas chromatography, mass spectrometry and spectrophotometry revealed that ISGA 1218 and 1222 were inactive against octane sulfonate and 6:2 FTSA; SsuD, however, was active against both octane sulfonate and 6:2 FTSA, removing up to 120 μM of added octane sulfonate and up to 130 μM 6:2 FTSA during a two hour reaction at room temperature. These data are preliminary, however, suggest sulfite is produced as a product of 6:2 FTSA degradation and likely represents the first evidence of the biochemical transformation of fluorinated surfactants by a purified bacterial enzyme. Further, *Escherichia coli* BL21(DE3) was found to tentatively grow on 6:2 FTSA prepared in water as a sole source of sulfur reaching $OD_{660}$ of 0.32 as compared to 0.15 of the no sulfur treatment after 48 hours. Attempts to produce *Gordonia* NB4-1Y mutants via conjugation or electroporation with pK18mobsacB1218A or pK18mobsacB1222AB were unsuccessful. Transformation of 6:2 FTSA by SsuD suggests the *ssu* operon in *E. coli* is responsible for the desulfonation of fluorotelomer sulfonates and presents the *ssuABC* system as a potentially versatile model to study the *in vivo* import of fluorotelomer sulfonates. Furthermore, assignment of 6:2 FTSA degradation to the *ssu* operon suggests that fluorotelomer sulfonate degradation may be enriched under sulfur starvation conditions. Identifying the enzymes responsible for aliphatic sulfonate degradation in *Gordonia* NB4-1Y is paramount to understanding the metabolism of fluorinated surfactants by this bacterium. Here we present one luciferase-like class flavin-dependent oxidoreductase (ISGA 08960) as candidate for further biochemical analysis.

Keywords: monooxygenase, 6:2 fluorotelomer sulfonate (FTSA), alkanesulfonate monooxygenase (SsuD), *Gordonia* NB4-1Y, protein purification, cloning

# II. Table of Contents

# III. Acknowledgements

Foremost, I would like to thank my supervisor, Dr. Jonathan Van Hamme for the charismatic support throughout the completion of my thesis.

I would further like to thank Dr. Eric Bottos, Breanne McAmmond and all of the TRUGen Team for their continued support inside and outside the pub.

I would like to thank Dr. Kirsten Wolthers and Osei Boakye Fordwour from the University of British Columbia, Kelowna Campus for their generous support with regards to protein chromatography and providing expression vectors.

I would like to that Dr. Don Nelson for his continued support in my academic development and generously providing chromatography resin and expression vectors.

I would like to thank the Spanish research team at the *Estacion Experimental del Zaidin,* Dr. Pieter van Dillewijn, Jesus de la Torre, Ines Aguilar Romero for hosting me in their lab in Granada and their interest and support in furthering my academic career.

I would like to thank Dr. Chelsea Vickers for her support and helpful comments regarding size exclusion chromatography and protein purification.

I would like to thank Dr. Nancy Flood for her help regarding the statistical analysis of my data.

I would like to thank Dr. Sharon Brewer and Trent Hammer for allowing me to use their gas chromatography systems and helpful support using and understanding mass spectrometry. Furthermore, I would like to thank Austin Pietramala for his help during GC-MS runs and data analysis.

Finally, I would like to thank Alana Babcock, Katheryn Haegedorn, Mathew Norman and Nicholas Chyzowski who, through designing and carrying out their own projects, helped me put together all the parts to this study.

# IV. Table of Figures

# V. Table of Tables

# 1.0 Introduction

The carbon-fluorine bond is one of the strongest bonds found in organic molecules and imparts partial charges on the carbon and fluorine atoms (O'Hagan, 2007). This gives fluorocarbons both lipo- and hydrophobic properties. Naturally occurring fluorine containing organic compounds are derived from Earth's geo-chemical processes or produced during specific biological process (Gribble, 2002 and O'Hagan and Harper, 1999); in addition to this, anthropogenic per- and polyfluoroalkyl substances (PFAS) such as perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA) have appeared in detectable quantities in the environment since the end of the 20$^{th}$ century (Moody *et al.*, 1999). An example of this class of molecules are the fluorinated surfactants, first put into mass production in the 1940s and now found in aqueous film forming foams (AFFF) for hydrocarbon-fire fighting, and durable wear repellents (Paul *et al.*, 2009 and Moody and Field, 2000). These anthropogenic molecules may be difficult to metabolize because of poor bioavailability, resistance to enzymatic biotransformation and toxicity (O' Loughlin *et al.*, 2009, Ochoa-Herrera *et al.*, 2016). The objective of this study is to design and develop protein production and mutagenesis vectors to assess the role two nitrilotriacetate monooxygenases (ISGA 1218 and 1222) have in the fluorinated surfactant degrading bacterium *Gordonia* NB4-1Y. Specific goals are to **1**) design and construct vectors for the production of ISGA 1218, 1222, alkanesulfonate monooxygenase (SsuD) and NADH:flavin oxidoreductase (Fre) with maltose binding protein (MBP) and hexa-histidine tags; **2)** develop an *in vitro* assay to test the activity of ISGA 1218, 1222 and SsuD against octane sulfonate and 6:2 fluorotelomer sulfonate (6:2 FTSA) with Fre; **3)** determine the kinetic properties of ISGA 1218, 1222 and SsuD against octane sulfonate and 6:2 fluorotelomer sulfonate; and, **4)** design and develop a markerless-mutagenesis vector to delete the genes encoding ISGA 1218 and 1222 from the *Gordonia* NB4-1Y genome.

## 1.1 Naturally occurring organic fluorine compounds

Naturally occurring compounds with carbon-fluorine bonds are not common in the environment, however, some processes may give rise to them. For example, fluoroalkanes may be produced by volcanoes and hydrothermal vents (Gribble, 2002) and some plant species, native primarily to Australia and Africa can produce fluorine containing metabolites such as fluorocitrate and fluorothreonine (O'Hagan and Harper, 1999). Naturally occurring organic fluorine compounds tend to contain a single fluorine atom; for example, the first to be identified was fluoroacetate (Marais, 1943) the toxic agent of *Gastrolobium*, the poison pea. Other examples include fluorocitrate, a product of fluoroacetate metabolism in eukaryotic cells and an inhibitor of the tricarboxylic acid cycle (TCA), nucleocidin, a broad-spectrum antibiotic produced by *Streptomyces clavus* (Morton *et al.*, 1969), fluorothreonine, a threonine analogue and anti-metabolite with antimicrobial properties biosynthesized by *Staphylococcus cattleya* (Hamilton *et al.*, 1997), and fluorooleic acid, the toxic agent found in the Western African *Datura toxicarium* (Peters *et al.*, 1960). O'Hagan and Harper (1999) presented the first overview of naturally occurring organic fluorine compounds. As of 2012, the five aforementioned compounds remain the only known naturally occurring organic fluorine compounds (Chan and O'Hagan, 2012).

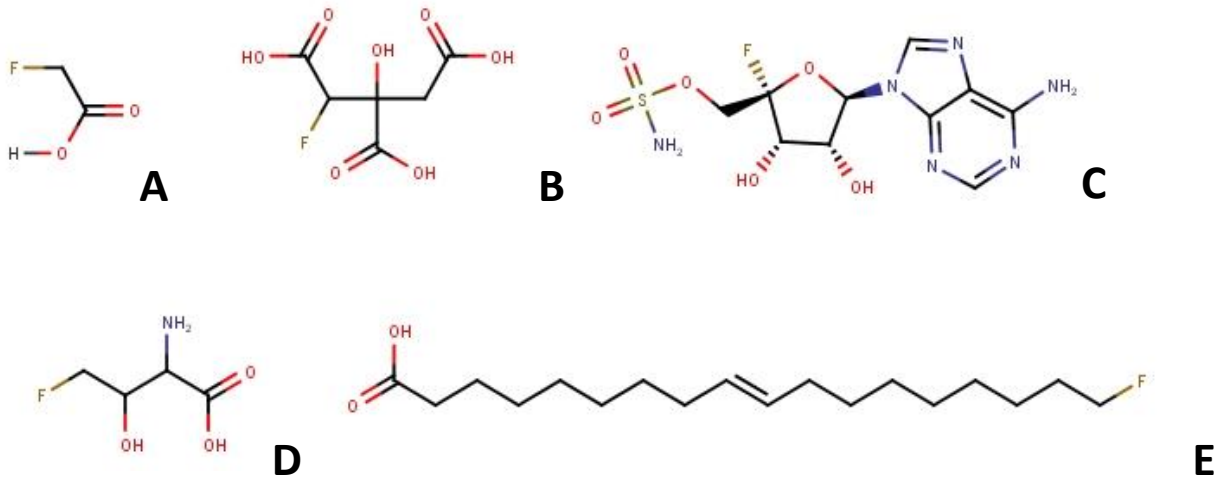


Figure 1. Naturally occurring organic fluorine compounds produced by biological systems. **A**: Fluoroacetate, **B**: Fluorocitrate, **C**: Nucleocidin, **D**: Fluorothreonine, **E**: Fluorooleic acid.

## 1.2 Anthropogenic fluorocarbons

Anthropogenic PFAS are found in catalysts, drugs and surfactants. Trifluoromethyl groups may be added to organic compounds using trifluoroacetate, a reagent that can modify ketone functional groups to a trifluoromethyl and hydroxyl functional groups (Chang, 2005). The carbon-fluorine bond is an important staple in the pharmaceutical industry. O'Hagan and Isanbor reported that as of 2006, 18% of drugs on the US market contained one or more fluorine atoms and, as of 2016, Atorvastatin was the third most prescribed medication in the United States (ClinCalc DrugState Database, 2019). Predicting the effect a carbon-fluorine bond will have on drug bioactivity can be challenging and can require extensive structure-activity relationship studies. Carbon-fluorine bonds primarily act on an organic structure by affecting the acidity or basicity of nearby hydrogen atoms (Wang *et al.*, 2013 and Morgenthaler *et al.*, 2007); in addition, carbon-fluorine bonds can decrease drug metabolism by inhibiting cytochrome P450 activity (Morgenthaler *et al.*, 2007 and Purser *et al.*, 2008) or increase the lipophilicity of the compound if positioned next to a pi-bond or on an aromatic ring (Smart, 2001 and Purser, 2008). Modulating acidity, basicity or lipophilicity can have a profound effect on drug potency and bioavailability by allowing stronger receptor interactions, facilitating passive diffusion across cell membranes or increasing storage in lipids. For example, Fluoxetine, otherwise known as Prozac, contains a trifluoromethyl group which imparts a nearly 6-fold increase in potency when incorporated in the *para*-position of the phenoxy ring over the non-fluorinated parent compound (Wong *et al.*, 1995 and Purser *et al.*, 2007). Contrarily, incorporation of a trifluoromethyl group in the *ortho*- or *meta*-position decreases potency by up to 14-fold (Wong *et al.*, 1995 and Purser *et al.* 2007). Pinpointing the exact effect of a carbon-fluorine bond on pharmacokinetics or bioavailability can be difficult and will depend on the mode of action, lipophilicity and conformation of the drug. Fluorinated drugs are not a concerning source of fluorocarbons in the environment due to the presence of at most three fluorines on any given fluorinated drug. However, the synthesis and

use of fluorinated surfactants such as perfluorooctane sulfonyl fluoride (POSF), PFOS and PFOA (Kissa, 2001) is concerning. These compounds are used for aqueous film forming foams for firefighting, omniphobic stains and non-stick finishes; with global production reaching up to 4500 tonnes in 2000 (Paul *et al.*, 2009) and global emissions predicted to reach as high as 450 tonnes per year of perfluorocarboxylic acids (PFCA) alone in 2020 (Wang *et al.*, 2014).

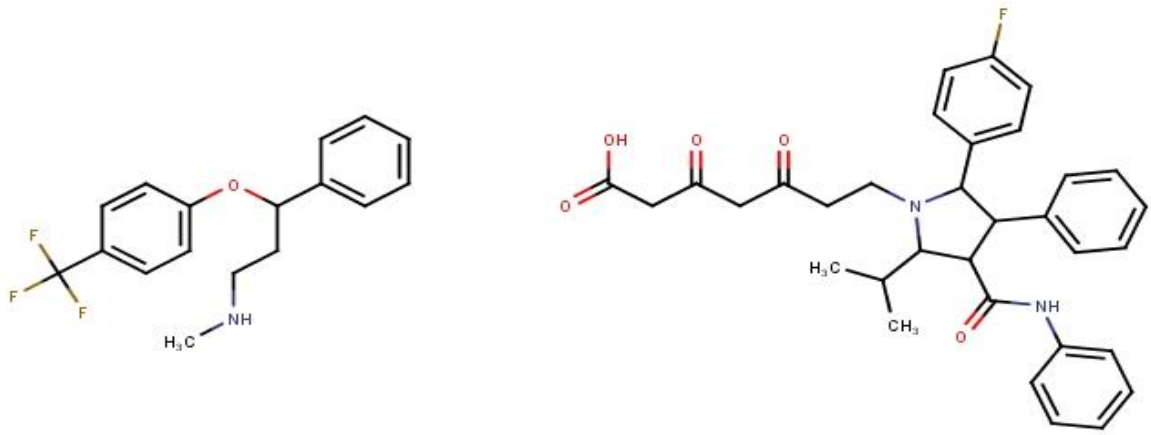


Figure 2. Fluorine containing pharmaceuticals. Left: Fluoexitine (Prozac), Right: Atorvastatin.

## 1.3 Fluorinated surfactants: synthesis, terminology, usage and their role in aqueous film forming foams

### 1.3.1 Fluorinated surfactants - synthesis

Fluorinated surfactants are an important class of molecules consisting primarily of a carbon chain backbone entirely (per) or partially (poly) saturated with fluorine atoms, with terminal functional group such as sulfate or carboxylic acid (Buck *et al.*, 2011). Fluorinated surfactants are synthesized using two different approaches: electrochemical fluorination and telomerization. Electrochemical fluorination is described in detail by Alsmeyer *et al.* (1994); briefly, a scaffold hydrocarbon such as octane sulfonyl fluoride is electrolyzed in the presence of hydrogen fluoride such that most, if not all, carbon-hydrogen bonds are replaced with carbon-fluorine bonds. Due to the nature of the reaction, it is difficult to control where and how many fluorines are incorporated into the organic scaffold. For example, Buck *et al.* (2011) reported that during electrochemical fluorination of octane sulfonyl fluoride, up to 80% of the final product was linear perfluorinated octane sulfonyl fluoride and 20 to 30% was branched perfluorinated carbon scaffold produced by carbon-carbon bond breakage. Consequently, the main purpose of electrochemical fluorination is the production of perfluorinated 6, 8 and 10 carbon sulfonyl fluorides which can be further derivatized (Lehmler, 2005 and Buck *et al.*, 2011). On the other hand, telomerization reactions offer a certain degree of control over the production of fluorinated surfactants by allowing the synthesis of surfactants containing carbon-fluorine and carbon-hydrogen bonds. Telomerization reactions first involve the synthesis of perfluoro iodide by reacting tetrafluoroethylene with pentafluoroethylene iodide. Subsequently, ethylene is radically inserted into the perfluoro iodide compound producing a polyfluoro iodide which can be further modified to an alcohol, amine, sulfate or sulfonyl functional group (Lehmler, 2005 and Kissa, 2001). Fluorotelomer synthesis generally produces an even number of perfluorocarbons with lengths varying from four to eight carbons since pentafluoroethylene iodide can react with more than one tetrafluoroethylene

molecule. Common end products of telomerization include 4:2, 6:2 and 8:2 fluorotelomer alcohols, carboxylic acids, amines and sulfonates (Lehmler, 2005).

### 1.3.2 Fluorinated surfactants - terminology

Fluorinated surfactants are numerous and diverse in structure in environmental matrices (Buck *et al.*, 2011), however, the naming paradigm for PFAS has been well established in the literature (Muller and Yingling, 2017). Organic compounds on the other hand, in particular pharmaceuticals, containing one fluorine atom are exempt. For the purposes of this study, compounds containing 1 to 3 fluorine atoms, with the exception of trifluoroacetate, will be referred to as fluorine containing organic compounds or organic fluorine compounds. PFAS are named following a consistent paradigm by first indicating the number of fluorinated carbons, followed by the number of non-fluorinated carbons that are present. For example, a polyfluorinated telomer sulfonate with six carbons saturated with fluorine and two hydrogenated carbons next to the terminal functional group would be named 6:2 fluorotelomer sulfonate (6:2 FTSA). Perfluorinated alkyl substances are named with the perfluoro prefix followed by the name of the organic scaffold. Well known perfluorinated surfactants include perfluorooctane sulfonate and perfluorooctanoic acid. Some fluorinated surfactants may be capped with more complex hydrocarbon moieties such as alkyl betaine or sulfonamides groups such as 6:2 fluorotelomer sulfonamidoalkyl betaine (6:2 FTAB) (Place *et al.*, 2012). Acronyms are typically used to shorten the names of fluorocarbons and acronym paradigms may vary; for the purpose of this study, acronyms are defined after they have been fully written and follow the paradigm outlined by Muller and Yingling (2017). Here, per- and poly-fluorinated alkyl substances as a group are abbreviated to PFAS.

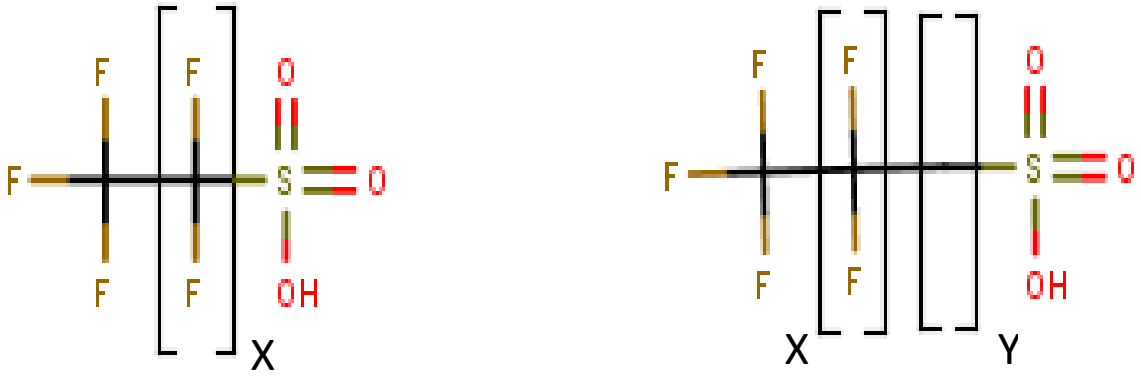


Figure 3. Naming convention of per and poly fluorinated surfactants. For the left structure, X = 7 would be perfluorooctane sulfonate; for the right structure, X = 5 and Y = 2 would be 6:2 fluorotelomer sulfonate.

### 1.3.3 Fluorinated surfactants - usage

Fluorinated surfactants are used by both the military and the public sector. Industries that use the most highly fluorinated surfactants include the textile, metal plating and aqueous film forming foam production industries. These industries are primarily interested in the chemical and heat resistance of fluorinated surfactants (Schroder and Meesters, 2005). Fluorinated surfactants exhibit both lipophobic and hydrophobic properties and therefore aggregate at gas-liquid interfaces, producing film-like barriers (Kissa, 2001). Chrome plating industries use the film-forming property of fluorinated surfactants as a mist suppressant to prevent the escape of carcinogenic $Cr^{6+}$ aerosols during non-decorative chromium plating (Poulson *et al.*, 2011). In the textile industry, fluorinated finishes on fabrics impart desirable water repellent properties, particularly important for medical personnel, chemical industry workers, and outdoor enthusiasts (Hill *et al.*, 2015, Schellenberger *et al.*, 2019, Ramaswamy *et al.*, 2004 and Mitchel *et al.*, 2015). Fluorinated surfactants are key components in the formulation of aqueous film forming foams (AFFF); foams which are used as hydrocarbon fire retardants. The first AFFF to hit the US market was produced by 3M, an electrochemical fluorination based AFFF (Place and Field, 2012).

Production of 3M AFFF was ceased in 2008 (Place and Field, 2012) due to the toxic and bioaccumulative nature of its active ingredient, PFOS; however, special exceptions have been made for military purposes in the United States.

### 1.3.4 Fluorinated surfactants - aqueous film forming foams

While formulations are proprietary, AFFF contain a mix of hydrocarbon and fluorocarbon-based surfactants (Kissa, 1994) and it has been found that the fluorinated surfactants typically used are polyfluoroalkyl sulfonates ranging from 4-10 perfluorinated carbons with various functional groups attached (Place *et al.*, 2012). AFFF currently on the US market are primarily telomerization based and are sold under brand names such as Ansul and National Foam. The primary fluorinated component in these foams are polyfluorinated 4:2 to 10:2 sulfonamide or thioether amido sulfonates (Place and Field, 2012). The most attractive quality of fluorinated surfactants in AFFF is their ability to form a barrier between a burning fuel source and oxygen. This arrests the combustion process and prevents fuel from re-igniting. It is critical that the active components of these foams are not destroyed during the combustion process and therefore, purely hydrocarbon based AFFF are much less effective than their fluorinated counterparts (Kissa, 1994). Fluorinated surfactants used in the textile industry include perfluorinated sulfonyl compounds ranging from 4 to 8 carbons long, with PFOS historically used in the chromium plating industry (Schellenberger *et al.*, 2019 and Poulsen *et al.*, 2011).

In some cases, non-fluorinated analogues have been found to have similar repellent properties as their fluorinated counterparts. For example, Schellenberger *et al.*, 2019 demonstrated that hydrocarbon-based durable wear repellents had similar water repellency properties as their fluorinated counterparts. Furthermore, non-PFOS fluorinated alternatives in the chrome plating industry have been shown to be effective (Poulsen *et al.*, 2011). These replacement efforts offer a potential solution to the unintended release of fluorinated hydrocarbons in the environment.

As part of military preparedness exercises, off grade fuels and combustible substitutes are ignited and quenched using AFFF. Moody and Field (2000) reported that these exercises were historically conducted on a regular basis at US Airforce bases, and on average 3000 liters of fuel would be extinguished with up 3200 liters of AFFF per week. Disposal of the AFFF – fuel mixtures often entailed release to local wastewater treatment plants or on site; for the latter this resulted in the contamination of an estimated 1621 groundwater wells across the United States with levels above 70 parts per trillion of PFOS or PFOA, the US Environmental Protection Agency level for safe lifetime consumption (US DoD, 2018). Efforts to phase out PFOS and PFOA based AFFF by the US Department of Defense are underway (US DoD, 2016). Current cleanup efforts include the use of tertiary wastewater treatment solutions such as activated carbon, nanofiltration and advanced oxidation processes (Schroder and Meesters, 2005, Eschauzier *et al.*, 2012 and Arvaniti and Stasinakis *et al.*, 2015).

## 1.4 Detection of PFAS in living organisms and the environment

In 1999, Moody *et al*. reported the detection of PFAS in groundwater where extensive military fire-fighting exercises had taken place. It was estimated that PFAS concentrations in groundwater near the Naval Air Station Fallon in Nevada and Tyndall Air Force Base in Florida ranged from 124 to 7090 µg/L, with the PFOA being found at the highest levels (Moody and Field, 1999). Between 4 and 110 µg/L of PFOS was detected in groundwater near decommissioned Wurtsmith Airforce Base in Michigan, five years after fire-fighting exercises ceased (Moody *et al.*, 2003 and Schultz *et al.*, 2004). PFAS have been detected at the nanogram per liter levels in various lakes in Canada including Lake Ontario, Huron and Superior (Scott *et al.*, 2006), and PFAS were also detected downstream of the John C. Munro International Airport in Hamilton, Ontario (de Solla *et al.*, 2012) with PFOS in microgram per gram quantities in turtle plasma and nanogram per gram quantities in homogenized amphipods (de Solla, *et al.* 2012). De Solla *et al.* (2012) suggested that the John C. Munro airport is likely the source PFOS contamination in the downstream rivers due to a combination of AFFF usage and release of AFFF-contaminated wastewater. These observations are further supported by a study that found ground and surface waters as well as soil and sediment near fire-fighting training areas to be contaminated with PFOS (City of Hamilton, 2011).

Globally, PFAS have been detected in many environmental matrices including rain, snow, marine and freshwater, air (Kim and Kannan, 2007, Muir *et al.* 2019, Wong *et al.*, 2018 and Yamashita *et al.*, 2005) and dust particles found in homes in Canada, the United States, China, Sweden and Japan (Yao *et al.*, 2018, Winkens *et al.*, 2018 and Awasum *et al.*, 2009). Furthermore, PFAS have been detected in living matrices including polar bear liver samples from both western and eastern Arctic borders of Canada (Smithwck *et al.*, 2006), plasma of Bottlenose Dolphins from the Gulf of Mexico and the Atlantic Ocean (Houde *et al.*, 2005), in plasma, liver and brains of Norwegian Gulls (Verrault *et al.*, 2005), in kidney, liver, blubber, muscle and spleen of seals off the coast of

the Netherlands (van de Vijver *et al.*, 2005), egg yolk of birds from Korea (Yoo *et al.*, 2008), and in serum samples from blood donors in the United States and China (Olsen *et al.*, 2003 and Yeung *et al.*, 2008).

## 1.5 Toxicology of PFOS

The health impacts of PFOS and other PFAS were not fully explored until the early 2000s. PFOS is considered to be bioaccumulative, to cause developmental problems (Liew *et al.*, 2018) disrupt the immune system of animals (Penden-Adams *et al.*, 2009), including humans, and can potentially cause reproductive dysfunction (Gao *et al.*, 2017). PFOS primarily targets the kidneys in humans and has been associated with chronic kidney disease in animal models (Shanker *et al.*, 2011). Although the classification of PFOS as a carcinogen is debated (Arrieta-Cortes *et al*. 2017), the 3M assessment of PFOS toxicological profile suggested that PFOS accumulates in the liver of model animals and is associated with an increase in benign tumor presence (3M Company, 2003 and Arrieta-Cortes, 2016). PFOS causes developmental abnormalities in chicken models; chickens exposed to PFOS *in ovo* had appendage abnormalities and brain asymmetries (Penden-Adams *et al.*, 2009). Although inconsistent across some studies, increasing prenatal exposure to PFOS has been negatively associated in some cases with low birth weights in humans (Apelberg *et al.*, 2007, Washino *et al.*, 2009 and Liew *et al.*, 2018). Immunotoxicity by PFOS is primarily caused by suppression of antibody response (DeWitt *et al.*, 2012). Gao *et al.* (2017) demonstrated that PFOS caused mis-localization of structural proteins in a blood-testis barrier model and could be one of the mechanisms by which PFOS causes reproductive dysfunction. The reported half-life of PFOS in humans is between 3-5 years (Li *et al.*, 2017, Olsen *et al.*, 2012 and USEPA, 2009) with shorter half-life values for women (Li *et al.*, 2017).

## 1.6 Global regulation of PFOS

In 2008, 3M voluntarily ceased PFOS production for civilian use (3M Company, 2003 and Place and Field, 2012) and in 2009, the Stockholm Convention on Persistent Organic Pollutants adopted decision SC-4/17 and placed PFOS and PFOS fluoride on Annex B of pollutants for worldwide elimination of production and usage. Canada, the United States and the European Union have placed restrictions on PFOS usage. Canada placed PFOS and its salts on the Virtual Elimination List, a list of substances whose use and release into the environment are restricted (Government of Canada, 2009) and the European Parliament banned PFOS usage in 2006 for consumer-end products, with some exemptions for industrial applications (European Parliament, 2006). Since their global phase out in the early 2000s, PFOS levels of blood donors in the United States decreased by 76% in geometric mean concentration from 2000-2010 (Olsen *et al.*, 2012). Olsen *et al.* (2012) suggested that the decrease in mean geometric concentration could be due to the decrease in environmental exposure to PFOS.

## 1.7 Environmental fate of PFAS

Early studies into the environmental fate of PFAS included the examination of municipal wastewater treatment system inflows and outflows in an effort to identify biodegradation products of PFOS and PFOA. Monitoring of PFOS and PFOA concentration in wastewater systems in New York (Sinclair and Kannan, 2006) and in Georgia and Kentucky (Loganathan *et al.*, 2007), revealed differences between inflow and outflow concentrations of PFOS and PFOA. Sinclair and Kannan (2006) compared two wastewater treatment plants that treated domestic and commercial wastewaters, with one plant additionally treating industrial wastewater and concluded that PFOS and PFOA were at higher concentrations in outflows at both plants following activated sludge treatment. Loganathan *et al.* (2007) reported similar findings where PFOS and PFOA concentration increased in wastewater treatment plant effluent. Taken together, these studies suggest that PFOS and PFOA precursors such as POSF or fluorotelomer alcohols (FTOH) are degraded during wastewater treatment.

Liu and Avendaño (2013) have extensively reviewed the degradation of PFAS by mixed and pure microbial cultures. For the purposes of this review, only the degradation of PFAS and precursors by mixed and pure microbial cultures will be discussed. Aside from being directly used in the formulation AFFF, some perfluoroalkyl sulfonates are degradation products of *N*-ethyl perfluoroalkane sulfonamides (Liu *et al.*, 2013). *N*-ethyl perfluorooctane sulfonamidoethanol is aerobically degraded in activated sludge to perfluorooctane sulfonate via dealkylation of the amidoethanol functional group (Lange, 2000 and Rhoads *et al.*, 2008). With regards to the polyfluoroalkyl sulfonates, 6:2 FTSA has been detected along with PFOS in groundwater affected by AFFF usage (Schultz *et al.*, 2004). Wang *et al.* (2011) and Zhang *et al.* (2016) assessed the biodegradation of 6:2 FTSA in activated sludge and aerobic and anaerobic sediment. Specifically, Wang *et al.* (2011) monitored the degradation of 6:2 FTSA in three different diluted activated sludge samples for 90 days and observed a relatively slow degradation rate; overall 63.7% of the

initially dosed 6:2 FTSA remained with degradation products accounting for 6.3% of the overall disappearance. Zhang *et al.* (2016) assessed the biodegradation of 6:2 FTSA in aerobic and anaerobic sediment and in contrast to Wang *et al.* (2011), nearly 80% of the applied 6:2 FTSA was degraded in aerobic sediment after 14 days. Zhang *et al*. (2016) found that 6:2 FTSA did not degrade in anaerobic sediment after 100 days of incubation, with no degradation products identified. Fluorotelomer thioether amido sulfonate (FtTAoS), also called Lodyne, is the active component of Ansul, a telomerization-based AFFF with a carbon-sulfur bond connecting a fluorotelomer chain to a hydrocarbon functional group. Harding-Marjanovic *et al.* (2015) found that 6:2 FtTAoS was the major degradation product of Ansul accounting for 8% of the added AFFF solution after 60 days in live soil microcosms. Contrary to these findings, D'Agostino and Mabury (2017) found that 6:2 fluorotelomer sulfonamide (FTSm) was the major product of sulfonamide based fluorotelomer alkylbetaine (FTAB) and alkylamine (FTAA) degradation, accounting for 0.9% and 6.9% of the degraded fluorotelomer, respectively. Regardless, these data clearly suggest a major rate-limiting step in the degradation of sulfonate, thiol and sulfonamide containing fluorotelomers is the desulfonation of FTSA or FTSm, a process hypothesized to be mediated by a monooxygenase. The degradation routes discussed are illustrated in Figure 4.

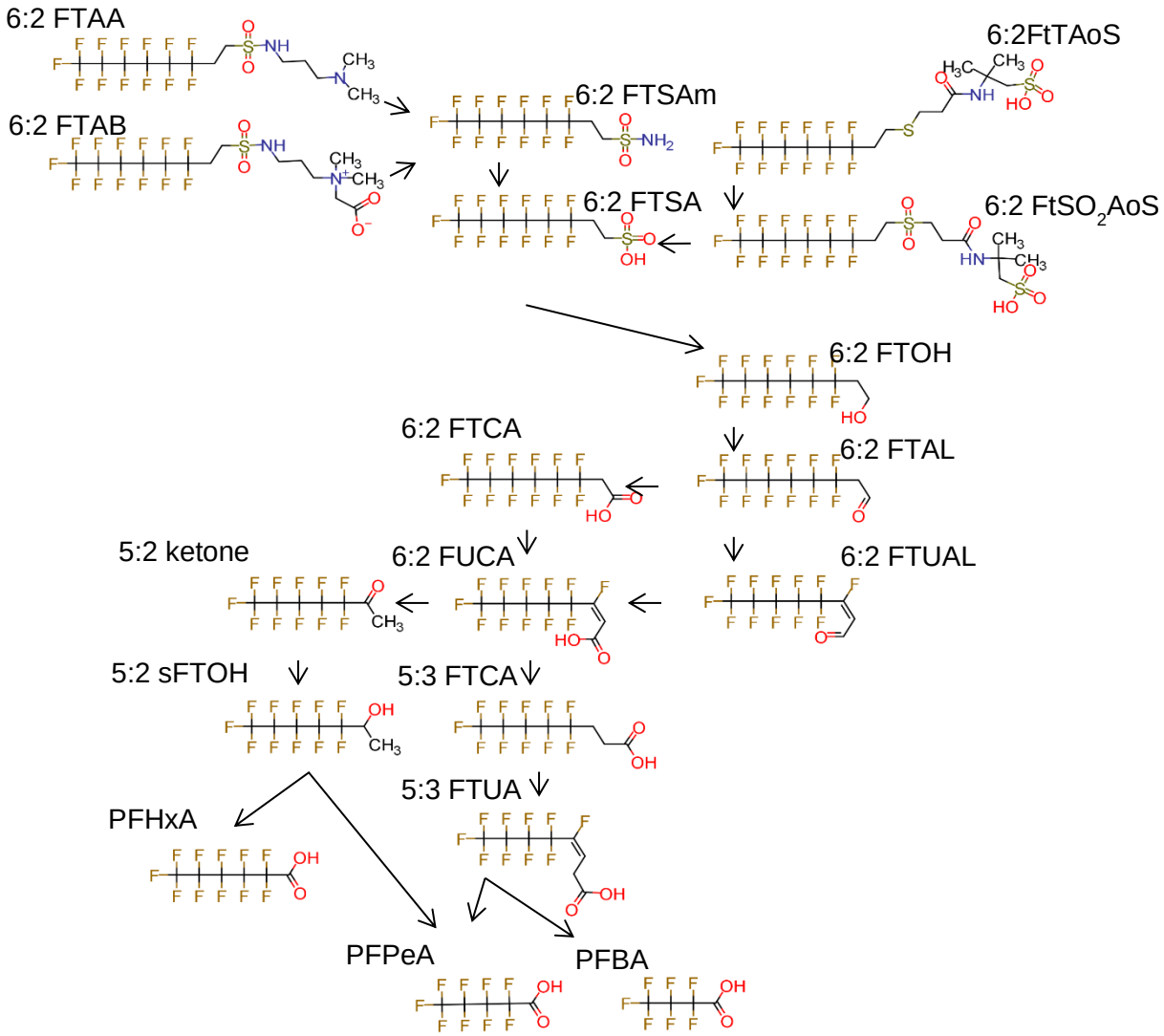


Figure 4. Degradation pathway of 6:2 FTSA and precursors by mixed microbial communities. Degradation routes adapted from Harding-Marjanovic *et al*. (2015) and adjusted to include data from D'Agostino and Mabury (2017). Acronyms are as follows: FTAA (fluorotelomer sulfonamide alkylamine), FTAB (fluorotelomer alkyl betaine), FTSAm (fluorotelomer sulfonamide), FTSA (fluorotelomer sulfonate), FtTAoS (fluorotelomer thioether amido sulfonate), $FtSO_2AoS$ (fluorotelomer sulfone amido sulfonate), FTOH (fluorotelomer alcohol), FTAL (fluorotelomer aldehyde), FTUAL (fluorotelomer unsaturated aldehyde), FTUCA (fluorotelomer unsaturated acid), FTCA (fluorotelomer carboxylic acid), ketone (fluorotelomer ketone), sFTOH (fluorotelomer isoalcohol), PFHxA (perfluorohexanoic acid), PFPeA (perfluoropentanoic acid), PFBA (Perfluorobutanoic acid). 6:2 FTAA, 6:2 FTAB, 6:2 FtTAoS and 6:2 FTSA were starting points for each degradation study. Proposed but not identified metabolites are 6:2 FTAL, 6:2 FTUCA and 5:3 FTUCA.

## 1.8.0 Bacterial two-component monooxygenases

### 1.8.1 Bacterial monooxygenases - Organic sulfonate cycling in the environment

Organic sulfonates, sulfonate esters, cysteine and methionine represent up to 95% of the available sulfur in soils, with inorganic sulfate representing the remainder (Kertesz, 1999). Examples of naturally occurring organic sulfonates include sulfoquinovose, cysteate and coenzyme M, and anthropogenic sulfonates include toluenesulfonate, dodecyl sulfate and octane sulfonate (Kertesz, 1999 and Könnecker *et al.*, 2011). Prokaryotes can use two different systems when acquiring sulfur, aryl or alkylsulfatases for aryl and alkyl sulfate esters, or mono and dioxygenase systems for aliphatic sulfonates (Toesch *et al.*, 2014, Huijbers *et al.*, 2013 and Eichhorn *et al.*, 1997). With regards to the former, sulfatases are conserved in sequence with Hanson *et al.* (2004) reporting 20-60% sequence homology of known prokaryotic and eukaryotic sulfatases at the time. The defining characteristic of a sulfatase is the post-translational α-formylglycine (FGly) residue in the active site of the enzyme (Hanson *et al.*, 2004 and Toesch *et al.*, 2014). The FGly motif has been proposed to act as an electrophile for an anionic sulfur-bound oxygen, or nucleophile for the sulfur center of a sulfate ester (Hanson *et al.*, 2004). Prokaryotic metabolism of aliphatic sulfonates is primarily carried out by two component flavin-dependent monooxygenase systems and in some cases, dioxygenases systems (Eichhorn *et al.*, 1999 and Eichhorn *et al.*, 1997).

### 1.8.2 Bacterial monooxygenases – Mechanism, distribution and substrate range

Monooxygenases in nature are found in Gram-negative and -positive bacteria and accomplish a variety of metabolic functions such as biodegradation or secondary metabolite modification (Eichhorn *et al.*, 1999, Thibaut *et al.*, 1995 and van Berkel *et al*., 2006). Monooxygenases are grouped by sequence similarity (van Berkel *et al.*, 2006) and the class C monooxygenases include degradative enzymes such as alkanesulfonate monooxygenase (SsuD), dibenzothiophene

monooxygenase (DszC), dibenzothiophene sulfone monooxygenases (DszA/B) and nitrilotriacetate monooxygenase (NtaA) (van Berkel *et al.*, 2006). Degradative monooxygenase systems are typically found in operon-like arrangements; one or more oxygenase is accompanied by a reductase and in some cases, an adenosine triphosphate (ATP) binding cassette (Ellis, 2010). For example, in *E. coli* the *ssu* operon regulates the uptake and degradation of aliphatic sulfonates and encodes *ssuD*, an alkanesulfonate monooxygenase, *ssuE,* a flavin reductase, *ssuA*, an aliphatic sulfonate-binding protein, *ssuB*, an aliphatic sulfonate import ATP-binding protein and *ssuC*, an aliphatic sulfonate permease (Eichhorn *et al.*, 1999 and 2000). The *dsz* operon responsible for dibenzothiophene metabolism in *Rhodococcus* and *Gordonia* have similar arrangements to the *ssu* operon however, encode three monooxygenases and a reductase; one monooxygenase oxygenating the sulfur center of dibenzothiophene (DszC), two oxygenolytically cleaving the carbon-sulfur bonds (DszA/B) and a reductase producing $FMNH_2$ (DszD) (Matsubara *et al.*, 2001, Ohshiro *et al.*, 1999). The substrate specificity of class C monooxygenases is diverse, but almost always restricted to the substrate-type of the archetypical monooxygenase system (Ellis *et al.*, 2010). For example, SsuD can catalyze the desulfonation of aliphatic sulfonates as well as substituted aliphatic sulfonates such as 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) and 3-(*N*-morpholino)propanesulfonic acid (MOPS) (Eichhorn *et al.*, 1999). Furthermore, although nitrilotriacetate (NTA) and ethylenediaminetetraacetate (EDTA) are similar in structure, two distinct monooxygenase systems oxygenate NTA and EDTA separately in *Aminobacter aminovorans* and bacterium BNC1 (Uetz *et al.*, 1992 and Payne *et al.*, 1998). The EDTA monooxygenase is the only monooxygenase that can degrade more than one substrate archetype (Jun *et al*. 2016).

In prokaryotes, the function of the reductase and monooxygenase components in desulfonation systems are typically coupled; the reductase oxidizes nicotinamide adenine dinucleotide (NADH to $NAD^+$) while reducing flavins (e.g. FMN to $FMNH_2$) and transfer the flavin to the

monooxygenase component which will use reduced flavins and oxygen as substrate to incorporate a single oxygen atom into a substrate (Ellis, 2010). Flavin reduction of SsuE is coupled to the oxygenation of aliphatic sulfonates through protein-protein interaction with a conserved alpha-helix on SsuD, however, SsuE is not required for the oxygenation activity of SsuD (Dayal *et al.*, 2015 and Koch *et al.*, 2005). Dayal *et al.* (2015) demonstrated that in the absence of structural interaction between SsuE and SsuD, SsuD activity remained. Furthermore, when the *ssuE* homolog, *ssuI*, from *Corynebacterium glutamicum* ATCC 13032 was deleted, growth was still observed when sulfur was limited to aliphatic sulfonates, although at slower rates (Koch *et al.*, 2005).

The catalytic mechanism of the prototypical alkanesulfonate monooxygenase, SsuD, has been extensively studied (Ellis, 2011 and Armacost *et al.*, 2014), however, due to the nature of the substrate involved, no substrate-bound structure has been proposed. In short, reduced flavin can diffuse from the active site of SsuE or cell cytoplasm (Dayal *et al.*, 2015) and react with oxygen within the active site generating a c4a-peroxyflavin (Ellis, 2010). This intermediate has been proposed to nucleophilically attack a sulfate group generating a peroxyflavin-organosulfate adduct and, following a Baeyer-Villiger rearrangement, sulfite is released (Ellis, 2011). Abstraction of the alpha hydrogen of the aliphatic aldehyde-peroxyflavin adduct results in the release of the aldehyde product and has been proposed to be one of the rate limiting steps in catalysis (Robbins and Ellis, 2012 and 2013).

### 1.8.3 Bacterial monooxygenases – Expression and purification methods

Historically, bacterial flavin dependent monooxygenases have been overexpressed from *lac*-controlled pET vectors or purified from whole cell extracts with a combination of crude extract precipitation, anion, cation, size exclusion and affinity chromatography (Eichhorn *et al.*, 1999, Oshiro *et al.*, 1999 and Uetz *et al.*, 1992). Recently, monooxygenase purifications have been

achieved by producing proteins with a histidine tag and purifying by nickel-affinity, with optional size exclusion, chromatography to remove soluble aggregates and imidazole (Adak and Begley, 2016, Carpenter *et al.*, 2011). Monooxygenases are primarily found in the cytoplasm of bacterial cells and expression with a histidine tag can lead to high levels of cytoplasmic proteins, however, this approach offers little to no help in promoting natural protein folding (Novagen, 2003). With regards to recombinant protein production, soluble tags can be used to increase yields or aid in correctly folding a protein (Bedouelle and Duplay, 1988). For example, the maltose binding protein tag (MBP) is soluble in *E. coli* and can facilitate solubilization of otherwise insoluble proteins (Rondard *et al.*, 1997). However, MBP tags can result in soluble protein aggregates (Raran-Kurussi and Waugh, 2012).

## 1.9 Fluorinated surfactant degradation by pure bacterial cultures

To date, five studies have identified seven bacterial species capable of PFAS biodegradation in pure culture; five species of *Pseudomonas* (Key *et al.*, 1998, Kim *et al.*, 2012, Liu *et al.*, 2007), one species of *Acidimicrobium* (Huang and Jaffe, 2019) and one species of *Gordonia* (Van Hamme *et al.*, 2013). With respect to fluorotelomer alcohols, *Pseudomonas* OCY4 and OCW4 were reported to aerobically co-metabolize 8:2 FTOH when grown on octanol to similar end products found in soil microcosms (Liu *et al.*, 2007). *Pseudomonas butanovora* and *Pseudomonas oleovorans* were found to degrade 4:2, 6:2 and 8:2 fluorotelomer alcohols to shorter chained PFAS (Kim *et al.*, 2012). The only species of *Pseudomonas* reported to degrade perfluorinated molecules is *Pseudomonas* sp. strain D2, reported to completely defluorinate difluoromethane sulfonate (DFMS), trifluoroethane sulfonate (TES) and partially defluorinate 6:2 FTSA; all of the aforementioned compounds could be used by *Pseudomonas* D2 as sole sulfur sources for growth (Ket *et al.*, 1998). In 2013, Van Hamme *et al.* reported that *Gordonia* sp. NB4-1Y, a vermicompost isolate, is capable of aerobically degrading 6:2 FTSA as a sole sulfur source. Shaw *et al.*, 2019 later followed up with a mass-balance study and reported that NB4-1Y can also use 6:2 FTAB as a sole sulfur source for growth. NB4-1Y was reported to use two degradation routes for the degradation of both 6:2 FTSA and FTAB, a major one terminating in the production of perfluorohexanoic acid (PFHxA), perfluoropentanoic acid (PFPeA), 5:2 fluorotelomer ketone (5:2 ketone) and 5:2 fluorotelomer isoalcohol (5:2 sFTOH), and a minor one terminating in perfluorobutanoic acid (PFBA), 4:3 fluorotelomer carboxylic acid 4:3 (4:3 FTCA), 4:2 fluorotelomer unsaturated acid (4:2 FTUA) and 4:2 fluorotelomer carboxylic acid (4:2 FTCA) (Shaw *et al.*, 2019).

It has been proposed that *Acidimicrobium* strain A6 is capable of metabolizing PFOS and PFOA anoxically although the authors only analyzed the aqueous phase of their cultures without extraction, and without including biomass (Huang and Jaffe, 2019). Prior to 2019, anoxic PFAS degradation by pure bacterial culture had not been reported; however, it had been hypothesized.

Ochoa-Herrera *et al.* (2008) reported that, under anoxic conditions, Ti(III) citrate and vitamin B12 could reductively defluorinate PFOS at 70°C and pH 9.0. The researchers postulated that given vitamin B12-catalyzed reductive dechlorination of perchloroethylene likely occurs via a radical reaction (Schmitz *et al.*, 2007), the biological transformation of PFOS might be possible via a similar reaction. In 2009, Colosi *et al.* reported that a combination of horseradish peroxidase, hydrogen peroxide and 4-methoxyphenol could degrade PFOA through a proposed radical activation of 4-methoxyphenol.

## 1.10 Biodegradation of 6:2 FTSA by Gordonia NB4-1Y and candidate 6:2 FTSA degradation genes

Of the seven bacterial species reported to degrade PFAS, Van Hamme *et al.* (2013) published the only study to support the degradation of PFAS with genomic DNA sequencing and proteomic analysis. *Gordonia* NB4-1Y is closely related to the dibenzothiophene desulfurizing *Gordonia desulfuricans*, and, while NB4-1Y is unable to grow on dibenzothiophene or dibenzothiophene sulfone, it does grow on fluorinated of non-fluorinated aliphatic sulfonates, alkyl thiols and cyclic thiols such as octane sulfonate, 6:2 FTSA, octyl sulfide and tetrahydrothiophene. Van Hamme *et al.* (2013) did not carry out an exhaustive search for PFAS breakdown products during the initial characterization of 6:2 FTSA degradation; however, 6:2 FTSA, 6:2 FTCA, 6:2 FTUCA, 5:3 Uacid and 5:3 acid were detected. In a more exhaustive search, Shaw *et al.* (2019) proposed that degradation of 6:2 FTSA follows similar routes to those taken by microbial communities (Figure 5).

Using two-dimensional differential gel electrophoresis followed by mass spectrometry of a subset of proteins differentially produced when NB4-1Y was growing on 6:2 FTSA instead of $MgSO_4$, Van Hamme *et al.* (2013) identified two monooxygenases that were differentially produced along with a double bond reductase, sulfate ABC transporter and two hydroperoxide reductases. The two identified monooxygenases, ISGA 1218 and 1222 (re-annotated ISGA_RS09775 and ISGA_RS09755) were reported to be nitrilotriacetate monooxygenases (NtaA), enzymes reported to catalyze the biodegradation of nitrilotriacetate to iminodiacetate and glyoxylate (Uetz *et al.*, 1992). ISGA 1218 and 1222 were later re-annotated as luciferase-like monooxygenase (LLM)-class flavin-dependent oxidorecutases. Of the nearly 120 oxygenases annotated in the NB4-1Y genome, ISGA 1218 and 1222, were found to have fewer than 0.67% sulfur containing amino acids, a number between four- and 14-times lower than all other predicted proteins in the NB4-1Y genome. Gene expression of lower sulfur content proteins is a response associated with sulfur

starvation (Scott *et al.*, 2006) and it could be hypothesized that these two genes are involved in sulfur metabolism. Although NtaA has not been documented to carry out sulfur bond cleavage, sequence similarity between the NtaA of *Aminobacter aminovorans* and the DszA of *Rhodococcus erythropolis* has been noted (Knobel *et al*.,1996 and Xu *et al.*,1997) and could be suggestive of a mis-annotation of ISGA 1218 and 1222 as NtaAs in 2013.

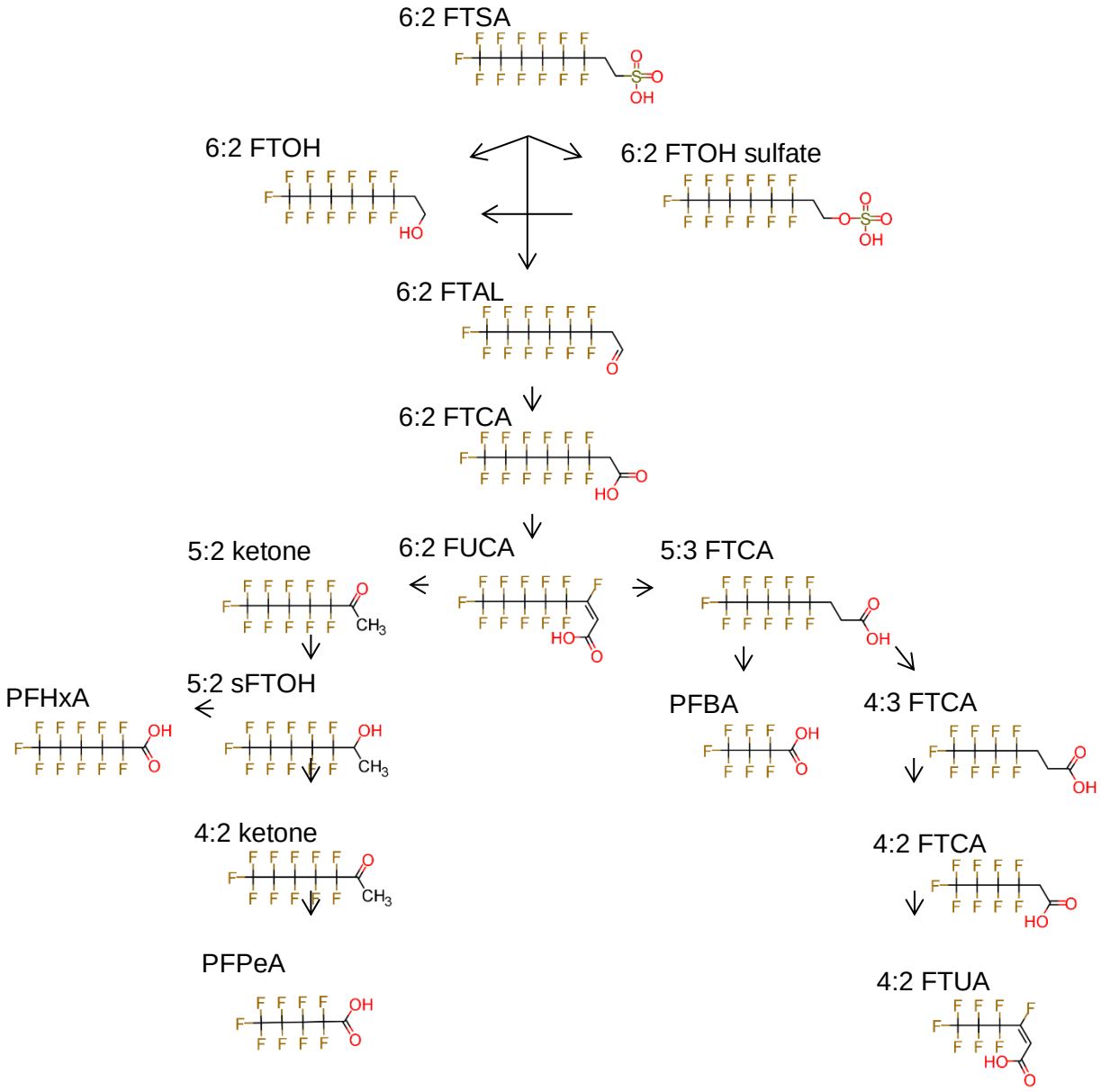


Figure 5. Degradation pathway of 6:2 FTSA by *Gordonia* NB4-1Y under sulfur limiting conditions; pathway adapted from Shaw *et al*. (2019). Acronyms not described in Figure 4 are FTOH sulfate (fluorotelomer sulfate ester).

## 1.11 Overview

To date, exact PFAS biodegradative pathways used by bacteria have been hypothesized primarily from pure culture studies and mixed microbial community studies. Here we report on the construction of seven expression and two mutagenesis vectors, biochemical characterization of the proteins produced by the expression vectors, a kinetic assessment of SsuD with Fre, and mutagenesis of *Gordonia* NB4-1Y ISGA 1218 and 1222. Gas chromatography (GC) mass spectrometry (MS) was used to identify octane sulfonate degradation products, flame ionization detection (FID) was used to quantify octanal production, and Ellman's reagent was used to spectrophotometrically quantify sulfite production from octane sulfonate and 6:2 FTSA; by ISGA 1218, 1222 and SsuD *in vitro*. In line with oxygen dependent degradation of fluorotelomer sulfonates by pure and mixed bacterial cultures and the structural similarity between octane sulfonate and 6:2 FTSA, SsuD is a likely catalyst for the degradation of 6:2 FTSA however, no activity was observed with ISGA 1218 and 1222.

# 2.0 Materials and methods:

## 2.1 Chemicals, buffers and microbiological media

Glycerol, 1-octanesulfonic acid sodium salt, hexanal, octanal, decanal, 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB), potassium phosphate dibasic, imidazole, β-nicotinamide adenine dinucleotide reduced disodium salt hydrate (NADH), riboflavin 5′-monophosphate sodium salt hydrate (FMN) and egg-white lysozyme were purchased from Sigma-Aldrich (St. Louis, MO) and were of molecular biological grade or higher. Tris(hydroxymethyl)aminomethane (Tris) base, acetic acid, ethylenediaminetetraacetic acid (EDTA), pentane, calcium chloride, ammonium chloride, sodium phosphate monobasic, magnesium sulfate heptahydrate, maltose, ethyl acetate, sodium chloride, ampicillin, isopropyl β-d-1 thiogalactopyranoside (IPTG) and Coomassie brilliant blue R-250 dye were purchased from ThermoFisher (Waltham, MA) and were of the highest purity available. Glucose was purchased from MP Biomedicals (Irvine, CA) and was molecular biological grade. Sodium dodecyl sulfate (SDS), bacteriological agar, kanamycin, hydrochloric acid (HCl) and glycine were purchased from VWR (Rando, PA) and were bacteriological grade or higher. Ammonium persulfate (APS) and acrylamide/bis (37.5:1) were purchased from Bio-Rad (Hercules, CA). Urea was purchased from EMD Millipore (Burlington, MA) and was ACS grade. Fluorinated compounds were purchased from SynQuest Laboratories (Alachua, FL). Unless otherwise noted, all solutions were dissolved in 15.6-18 Mega-Ohm distilled deionized water. Concentrations of solute in solution were reported in molarity or % wt/vol or vol/vol representing the amount, in grams or milliliters per 100 mL, of solid dissolved in solution or the volume of a given liquid with respects to total volume of an aqueous solution, respectively.

Lysogeny broth (LB) was prepared by dissolving yeast extract, sodium chloride (NaCl) and tryptone to a final concentration 0.5, 0.5 and 1.0% wt/vol, respectively, in water. Nutrient broth (NB) was purchased pre-mixed (Oxoid, Basingstoke, UK) and prepared to a final concentration

of 0.2% wt/vol yeast extract, 0.5% wt/vol peptone, 0.5% wt/vol NaCl, and 0.1% wt/vol ‘Lab-lemco’ powder. Nutrient agar contained the same concentrations as above save for the addition of bacteriological agar to a final concentration of 1.5% wt/vol. Tris-HCl was prepared by dissolving Tris base powder in water and adjusting to the desired pH with HCl. Concentrated (10X) M9 salts were prepared to 70% wt/vol sodium phosphate monobasic, 30% wt/vol potassium phosphate dibasic, 5% wt/vol NaCl and 1% wt/vol ammonium chloride. M9 minimal medium, pH 7.0, broth and solid media were prepared by diluting 10X M9 salts with water to 1X and adding glucose to 0.5% wt/vol. Unless otherwise noted, M9 minimal medium was supplemented with 1 mM magnesium sulfate ($MgSO_4$). All solid media were prepared by dissolving bacteriological agar to a final concentration of 1.5% wt/vol. All media were autoclaved at 121°C for 20 minutes at 827 kPa, and 0.22-0.45 µm filter sterilized carbon sources were added after cooling below 45°C as required. All culture media and chromatography buffers were prepared fresh or stored at 4°C until use.

NADH and FMN stocks were prepared to a final concentration of 50 mM, and 1 mM, respectively in water. Coomassie protein staining solution was prepared by dissolving 0.25% wt/vol Coomassie Brilliant blue in 45% vol/vol ethanol and 10% vol/vol acetic acid. Base binding buffer for protein chromatography was prepared by filtering 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 10% vol/vol glycerol with a 0.45 µm filter. Binding buffer for nickel affinity chromatography contained 20 mM imidazole and elution buffer contained 500 mM imidazole. Elution buffer for amylose affinity chromatography contained 10 mM maltose. Hexanal, octanal, octanol and decanal working stocks were prepared to a final concentration of 8.30, 6.40, 6.37 and 5.40 mM, respectively, by dissolving 1 µL of 98-100% purity stock in 999 µL of ethyl acetate. Working stocks were further diluted in ethyl acetate for analytical standard purposes. Working solutions of 1-octanesulfonic acid sodium salt, hereafter referred to as sodium octanesulfonate or octane sulfonate, and ammonium 1H, 1H, 2H, 2H-perfluorooctane-1-sulfonate (6:2 FTSA) were prepared

in 50% vol/vol ethanol to a final concentration of 3 mM. Octane sulfonate and 6:2 FTSA for growth assay were prepared to 10 and 1 mM, respectively, in water. Sodium sulfite working stocks were prepared to a final concentration of 50 mM in $N_2$-purged water and DTNB working stocks were prepared to a final concentration of 10 mM in $N_2$-purged 50 mM Tris-HCl pH 8.0, aliquoted and frozen at -20°C until use. NADH stocks solutions were prepared to a final concentration of 50 mM and stored at -20°C until use. FMN stock solutions were prepared to a final concentration of 1 mM and stored at -4°C until use. Ampicillin, Kanamycin and IPTF stock solutions were prepared at 100, 50 mg/mL and 1 M, respectively, and stored at -20°C until use. EDTA stock solutions were prepared by adding EDTA to a final concentration of 0.5 M and adjusting to pH 8.0 with sodium hydroxide until the EDTA dissolved.

## 2.2 Primers for polymerase chain reaction (PCR) used in this study

All primers in this study were purchased from Alpha DNA (Montreal, QC) and were delivered desalted and lyophilized. Primers stocks were reconstituted with water to 100 µM and further diluted to 5 and 10 µM for routine use.

Table 1. Primers used in this study.

| Name | Sequence [1] | Source | Purpose | $T_m$ (°C) [2] |
|---|---|---|---|---|
| 1218F | GTA**GAATTC**ATGAACGTAAACGTTG | Milton-Wood (2016) | Cloning | 50 |
| 1218F(3) | GTA**CCATGG**TCATGAACGTAAACGTTGTTGG | This Study | Cloning | 59 |
| 1218R -S (2) | TAC**AAGCTT**CGCCGCCCCCAC | This Study | Cloning | 58 |
| 1222F | GAC**GAATTC**ATGGCTGATCGAGAG | Milton-Wood (2016) | Cloning | 50 |
| 1222 F(3) | GAC**CCATGG**TTATGGCTGATCGAGAGCTCC | This Study | Cloning | 61 |
| 1222R -S (3) | TAC**AAGCTT**ACCGGTCCGGCG | This Study | Cloning | 56 |
| EcoliSsuD_F(2) | TAC**GAATTC**ATGAGTCTGAATATGTTCTGGTTTTTACC | This Study | Cloning/ Sequencing | 62 |
| SsuD F(3) | TAC**CCATGG**CCATGAGTCTGAATATGTTCTGGTTTTTACC | This Study | Cloning | 64 |
| EcoliSsuD_R (2) | TAC**AAGCTT**TTAGCTTTGCGCGACTTTACG | This Study | Cloning/ Sequencing | 61 |
| SsuDR -S | TAC**AAGCTT**GCTTTGCGCGACTTTACG | This Study | Cloning | 59 |
| FreF (3) | TAC**CCATGG**CCATGACAACCTTAAGCTGTAAA | This Study | Cloning | 59 |
| FreR -s (3) | TAC**AAGCTT**GATAAATGCAAACGCATC | This Study | Cloning | 52 |
| 18AF - HindIII | ACG**AAGCTT**TCGACCTTCTTCTCCGAGT | This Study | Cloning | 59 |

| | | | | |
|---|---|---|---|---|
| 18AR - XbaI | ACG**TCTAGA**TCGTCCTCGATCAACTGACC | This Study | Cloning | 61 |
| 18BF - XbaI | ACG**TCTAGA**CGAGATCTCTCCGTTCGTT | This Study | Cloning | 58 |
| 18BR - BamHI | ACG**GGATCC**CGCGACCCTGCCCAA | This Study | Cloning | 62 |
| 18-250 | CCAATCCTGCGCCGAG | This Study | Sequencing | 60 |
| 18-750 | CCGGCCAACCTCGG | This Study | Sequencing | 58 |
| 22AF – SacI | ACG**GAGCTC**GTCGCGATCAGCCG | This Study | Cloning | 56 |
| 22AR – XbaI | ACG**TCTAGA**GTCGCTCGTCCGGA | This Study | Cloning | 57 |
| 22BF – XbaI | ACG**TCTAGA**TGTTCGCCGCTGTC | This Study | Cloning | 55 |
| 22BR – EcoRI | ACG**GAATTC**CACCGGCGATTTCTTC | This Study | Cloning | 54 |
| 22-250 | CGACGTTGCCCGCC | This Study | Sequencing | 60 |
| 22-750 | CTCTGAAAGTTGCTTATCTCGAGT | This Study | Sequencing | 60 |
| T7F | TAATACGACTCACTATAGGG | UBC SBC | Sequencing | 52 |
| T7TR | GCTAGTTATTGCTCAGCGG | UBC SBC | Sequencing | 58 |
| T7Pol_F | ACTCTGGCTTGCCTAACCAGT | This Study | Sequencing | 64 |
| T7Pol_R | CCTTGCGGTACACAGCA | This Study | Sequencing | 59 |
| MBP-F | GATGAAGCCCTGAAAGACGCGCAG | Milton-Wood (2016) | Sequencing | 68 |
| 21M13 | TGTAAAACGACGGCCAGT | UBC SBC | Sequencing | 59 |
| M13R | CAGGAAACAGCTATGAC | UBC SBC | Sequencing | 51 |

[1] Bolded characters represent restriction enzyme cut sites.
[2] $T_m$ denotes the predicted annealing temperatues of the primer to plasmid or genomic DNA.

## 2.3.0 DNA visualization, manipulation *in vitro* and *in vivo* and sequencing conditions

### 2.3.1 Agarose gel electrophoresis

Agarose gels for DNA separation based on size were prepared by heat dissolving high-purity agarose in 1X TAE (40 mM Tris base, 20 mM acetic acid and 1 mM EDTA) to 0.8-1.2% wt/vol. Before pouring, GreenView Plus (ABP Bioscience, Beltsville, MD) or Gel Red (Biotium, Hayward, CA) was added to 0.5X of stock concentration for visualization using a Bio-Rad Gel Doc XR+ UV-Vis transilluminator (Bio-Rad, Hercules, CA). The 1 Kb plus DNA ladder (Invitrogen, Carlsbad, CA) was separated alongside samples to estimate DNA fragment size. Gels photos were taken using the Gel Doc XR imaging software (Bio-Rad, Hercules, CA).

### 2.3.2 Genomic DNA extractions

Genomic DNA extractions from Gram-positive and -negative bacteria were carried out using the PureLink genomic DNA Mini Kit (Invitrogen, Carlsbad, CA) following the manufacturer instructions. In brief, a single, isogenic colony was inoculated in 5-10 mL of LB or NB and grown to saturation at 30-37°C. Following incubation, cells were harvested by centrifugation and lysed with PureLink Genomic Digestion buffer (Invitrogen, Carlsbad, CA). For Gram-positive bacteria, Lysozyme Digestion buffer (Invitrogen, Carlsbad, CA) was used instead. Genomic DNA was then applied to a PureLink Spin Column (Invitrogen, Carlsbad, CA), washed and eluted with PureLink Genomic Elution buffer (Invitrogen, Carlsbad, CA).

### 2.3.3 Plasmid extractions

All plasmid extractions were carried out following the E.Z.N.A. Plasmid DNA Mini Kit 1 (Omega Bio-Tek, Norcross, GA) protocol. In brief, a single, isogenic colony was inoculated in 5-10 mL of LB or NB with antibiotics (50 µg/mL of kanamycin or 100 µg/mL of ampicillin) and grown overnight

at 30-37°C. The following day, cells were harvested by centrifugation, re-suspended in Solution I (Omega Bio-Tek, Norcross, GA) with RNase and lysed with Solution II (Omega Bio-Tek, Norcross, GA). Protein and genomic DNA was precipitated with Solution III (Omega Bio-Tek, Norcross, GA) and spun at 13,000 *g* for 10 minutes. Plasmid containing solutions were applied to a HiBind DNA Mini Column (Omega Bio-Tek, Norcross, GA) and washed twice with HBC (Omega Bio-Tek, Norcross, GA) and DNA Wash buffer (Omega Bio-Tek, Norcross, GA) and eluted in Elution buffer (Omega Bio-Tek, Norcross, GA).

### 2.3.4 PCR conditions

All PCRs were carried out on an Applied Biosystems SimpliAmp thermocycler (ThermoFisher, Waltham, MA). For gene cloning and mutagenesis related PCRs, Q5 2X Master Mix (New England BioLabs, Ipswich, MA) or Phusion polymerase with Phusion HF or GC buffer based reactions (New England BioLabs, Ipswich, MA) were used. A typical PCR contained the following: 1 unit of polymerase, 0.5-1 µM of primers, 200 µM of nucleotides, 1X HF or GC buffer, and 0.5 to 50 ng of template DNA. For screening related PCRs, the GoTaq Green PCR Master mix (Promega, Madison, WN) was used. Primers used in screening reactions were the T7 and T7term, T7pol_F and T7_polR or 18-250/750 and 22-250/750. Reactions were carried out with 1X GoTaq, 0.1-1 µM of primer and 0.5-50 ng of template DNA or a single isolated colony. Typical cycling conditions are given in Table 2.

Table 2. PCR cycling conditions.

| Temperature (°C) | Time (seconds) | Number of Cycles |
|---|---|---|
| 95 | 60-300 | 1 |
| 95 | 30 | 30-35 |
| Variable[1] | 15 | |
| 72 | 30-60 | |
| 72 | 180 | 1 |

[1]Annealing temperatures were 5°C lower than the lowest predicted annealing temperature for any primer pair.

### 2.3.5 Gel extractions

All gel extractions were carried out following the E.Z.N.A Gel Extraction Kit (Omega Bio-Tek, Norcross, GA) protocol. In brief, DNA in agarose gels were visualized under blue light transillumination with GreenView Plus dye (ABP Bioscience, Beltsville, MD) and excised depending on the desired size. Gel fragments were solubilized with XP2 Binding buffer (Omega Bio-Tek, Norcross, GA), with shaking. Soluble gel fragments were applied to a HiBind DNA Mini Column, washed with SPW buffer (Omega Bio-Tek, Norcross, GA), and eluted in Elution buffer.

### 2.3.6 PCR and enzymatic digestion reaction DNA clean-up

All DNA clean-up protocols were carried out following the E.Z.N.A. Cycle Pure Kit (Omega Bio-Tek, Norcross, GA) protocol. DNA containing solutions to be cleaned were diluted 1:3 or 1:4 in CP buffer (Omega Bio-Tek, Norcross, GA), and applied to a HiBind DNA Mini Column. The column was then washed with DNA Wash buffer and eluted in Elution buffer.

### 2.3.7 Restriction digestion conditions

DNA fragments and plasmid with restriction enzyme cut sites were treated with appropriate restriction enzymes to allow for sticky-end ligation or to confirm the size of an insert within a multiple cloning site. FastDigest restriction enzymes were purchased from ThermoFisher (Waltham, MA) and all other restriction enzymes were purchased from Invitrogen (Carlsbad, CA) or New England BioLabs (Ipswich, MA). All digestion reactions were carried out at 37°C with 1 unit of enzyme per 20 µL in 1X FastDigest Green buffer (ThermoFisher, Waltham, MA). A unit of enzyme is defined by ThermoFisher or New England BioLabs and did not surpass 1 µL of enzyme per 20 µL reaction volume. The amount of DNA per reaction varied depending on the end goal. For the purposes of this study, the term restriction digestion and digestion were used interchangeably.

### 2.3.8 Ligation reactions

Ligation reactions were carried out on DNA fragments treated with restriction enzymes to ligate paired sticky ends. Reactions were carried out at a 3:1 insert to vector ratio in 1X T4 DNA ligase buffer (Invitrogen, Carlsbad, CA) using 1 unit of T4 DNA ligase (Invitrogen, Carlsbad, CA) and incubated overnight at 4°C or 1 hour at room temperature. A unit of T4 DNA ligase is defined by Invitrogen (Carlsbad, CA) and the total weight of DNA per reaction did not exceed 100 ng.

### 2.3.9 Design of protein production and mutagenesis vectors

Vectors intended for protein purification were designed such that the gene, backbone encoded ribosomal binding site, **ATG** start codon and affinity tag were in frame. Genes destined for pET vectors included an NcoI (CC**ATG**G) cut site followed by two nucleotides on the primer binding the 5′ end of the gene and HindIII cut site on the primer binding to the 3′ end. When the gene was amplified from genomic DNA and ligated into a pET vector, these sites put the **ATG** start codon of the vector backbone in frame with the gene being expressed and the histidine tag. The stop

codon of each gene was omitted from the 3′ primer for readthrough and production of a histidine tag; a stop codon was encoded in the vector backbone after the histidine tag to terminate translation. These modifications produced a protein with an extra methionine and spacer amino acid at the N-terminal end and a histidine tag at the C-terminal end. Genes destined for pMAL-c2 vectors were amplified with primers only containing a restriction enzyme cut site. All restriction enzyme cut sites in pMAL-c2 are in frame with the start codon of maltose binding protein (MBP) and therefore, the primers binding to the 5′ and 3′ ends of the genes to be produced included a HindIII and EcoRI or XbaI restriction enzyme cut site respectively. Furthermore, stop codons were not omitted from the 3′ primer as to terminate translation with the genes native stop codon. This resulted in the production of MBP followed by a fusion linker and the target protein.

In order to produce mutants with gene deletions for the genes encoding ISGA 1218 and 1222 in *Gordonia* NB4-1Y, two suicide vectors were built by ligating 1000 base pairs of the genomic regions up- and down- stream of the genes encoding ISGA 1218 and 1222. The vector backbone used was pK18mobsacB which contains an origin of replication for *E. coli* and two selectable marker genes, *sacB* and *kanR*. The regions 1000 base pairs upstream and downstream of ISGA 1218 and 1222 were amplified generating the A (upstream) and B (downstream) fragments. Each fragment pair (A+B) was amplified with primers containing three unique restriction enzyme cut sites; one unique site for the A and B fragments and one shared between the two. The shared cut site was used to ligate the fragments together. For the gene encoding ISGA 1218, amplicons of the A fragment contained a 5′ HindIII and 3′ XbaI cut site and amplicons of the B fragment contained a 5′ XbaI and 3′ BamHI cut site. For the gene encoding ISGA 1222, amplicons of the A fragment encoded a 5′ SacI and 3′ XbaI cut site and amplicons of the B fragment encoded a 5′ XbaI and 3′ EcoRI cut site. The A and B fragments flanking ISGA 1222 were first cloned into pET23d and excised with SalI and EcoRI considering the pK18mobsacB plasmid encodes two SacI cut sites whereas pET23d contains one. To construct the mutagenesis vectors, the A and B

fragments were first amplified independently, digested with XbaI and ligated together. Ligated A and B fragments are hereafter referred to as AB fragment. Following, the AB fragment was digested with the HindIII and BamHI or SalI and EcoRI pair, cleaned and ligated into pK18mobsacB. Screening of transformants with AB fragment insertion was done with 18-250/750 or 22-250/750 primer pairs where one primer bound to the A fragment 250 base pairs from the XbaI cut site and the other 750 base pairs from the XbaI cut site on the B fragment. Colony PCR of positive transformants would produce a 1000 base pair product whereas AA or BB ligations would produce 500 or 1500 base pair products, respectively.

### 2.3.10 Preparation of electro- and chemically- competent cells

Competent cells were prepared by inoculating a single isogenic colony in 5-10 mL of LB or NB and grown overnight at 37°C. Subsequently, 0.4-1 mL of culture was added to 50 mL of LB or NB and grown to an optical density ($OD_{660}$) of 0.5 at 660 nm. The cells were then harvested at 4°C and re-suspended in 15 mL ice-cold, sterile, 15% vol/vol glycerol if preparing electrocompetent cells and 15% vol/vol glycerol with 150 mM calcium chloride if preparing chemically competent cells. The centrifugation-re-suspension step was repeated twice, suspending in 5 and 0.5 mL then aliquoted and stored at -80°C until use.

### 2.3.11 Transformation of *E. coli* by electroporation or heat shock

Plasmids were quantified using the Qubit™ dsDNA HS Assay Kit (ThermoFisher, Waltham, MA) or with a NanoDrop One (Thermo Fisher, Waltham, MA) at 280 nm. Each transformation was carried out using 50-100 µL of competent *E. coli* BL21(DE3), DH5a or S17.1 cells. For electroporation, ligation reactions were cleaned with an E.Z.N.A. Cycle Pure Kit and eluted to a final volume of 30 µL. Electroporation mixtures were prepared by mixing 10-20 µL of ligation reaction with 50-100 µL of electrocompetent cells and placed in a pre-chilled Gene Pulser Electroporation cuvette (Bio-Rad, Hercules, CA). Each electroporation was carried out on an

Eppendorf Electroporator 2510 (Hamburg, Germany) set to 2.2 kV for *E. coli*. For heat-shock, ligation reactions were added directly to chemically competent cells and incubated on ice for 30 minutes. The cell-DNA solution was placed in a 42°C water bath or thermocycler for exactly 30 seconds and then recovered on ice for 2 minutes. Transformed cells were further recovered in 1 mL of LB or NB without antibiotics at 37°C for 30-60 minutes. Varying amounts of transformed, recovered bacteria were plated on selective media and incubated at 37°C until isolated colonies appeared. Confirmed transformants were frozen at -80°C in a Microbank™ (Pro-Lab Diagnostics, Toronto, ON) or diluted 1:1 in sterile, 60% vol/vol glycerol for long term storage.

### 2.3.12 Sanger sequencing of plasmids

Plasmid samples to be Sanger sequenced were prepared at 150 ng/µL and sent to the University of British Columbia Sequencing and Bioinformatics Consortium (UBC SBC) sequencing service. Chromatograms were analyzed using FinchTV (Geospiza Inc., Seattle, WA) and multiple sequence alignments were performed using the MultAlin software (F. Corpet).

### 2.3.13 Basic local alignment search tool (BLAST) parameters

In ordered to search the *Gordonia* NB4-1Y genome for proteins similar to SsuD, the blastp suite was used (NCBI, Bethesda, MD). Query sequences were searched, in single letter amino acid format, against the *Gordonia* sp. NB4-1Y genome (taxid: 1241906). Alignments were scored with the BLOSUM62 matrix with an 11 existence and 1 extension gap cost. A maximum of 100 sequences were displayed, an expected threshold of 10 and a word size of 6 were set.

## 2.4.0 Protein production, release, visualization and purification conditions

### 2.4.1 Protein production assays

In order to determine optimal protein production conditions, an induction assay varying both IPTG concentration and temperature was carried for selected strains carrying protein production plasmids. A single colony of each strain was inoculated into 10 mL of LB or NB broth containing 50 µg/mL of kanamycin or 100 µg/mL ampicillin and grown overnight at 30-37°C with shaking. The bacterial concentration of the culture was estimated by measuring $OD_{660}$ and 30 mL of selective LB or NB was inoculated to a final $OD_{660}$ of 0.05. The resulting culture was incubated for 3 hours or until an $OD_{660}$ of 0.5 was reached. Cultures were then separated into three, 10 mL aliquots and IPTG was added to a final concentration of 0, 0.3 or 0.6 mM for pMAL vectors and 0, 0.5 or 1.0 mM for pET vectors. For a single strain, nine 10 mL cultures were prepared such that 18, 30 and 37°C production temperatures could be tested for all IPTG concentrations. Protein production with pMAL vectors were limited to a maximum of 2 hours and the pET overnight (~16 hours). Large scale protein production was carried out with the same method save for the addition of 2 g of glucose to the growth medium. Large volumes of cells were harvested at 4°C with 5,000 *g* using a Beckman (Brea, CA) J2-H5 centrifuge in 500-mL centrifuge bottles. Supernatant was discarded, cell paste collected with a clean spatula and frozen at -80°C until use.

### 2.4.2 Protein release from cells

For small scale protein productions (<10 mL), cell pastes were re-suspended in 1 mL of binding buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 10% vol/vol glycerol and sonicated with a Misonix Microson Ultrasonic (Misonix, NY) cell disruptor on ice with 15-30 seconds on time and 30-60 seconds off time. For large scale protein production (1-3 L), cell pastes were typically frozen at -80°C and re-suspended in 10-30 mL of 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10% vol/vol glycerol, imidazole as needed and 5-10 mg of lysozyme, incubated for 1 hour at room

temperature and sonicated with 1 minute on and 1-2 minutes off time. To pellet insoluble proteins, cell lysates were spun at 10-15,000 *g* for 10-20 minutes at 4°C. Efficient lysis was seen by a clear strong yellow-brown hue which was filtered with a 0.45 µm filter prior to chromatography. Supernatant was treated as the soluble fraction whereas the pellet was treated as the insoluble fraction and the supernatants stored on ice or at 4°C until use.

### 2.4.3 Sodium dodecyl sulfate (SDS) poly-acrylamide gel electrophoresis (PAGE) preparation

In order to visualize proteins by size and to estimate relative abundance, SDS-PAGE was used. SDS-PAGE gels were first prepared by making a resolving gel (Table 3) in a total volume of 10 mL and polymerized by the addition of tetramethylethylenediamine (TEMED). The resolving gel solution was quickly added between two SDS-PAGE glass plates with an internal spacing of 1.0 mm. A layer of 70% isopropanol was added on top of the resolving solution to ensure level polymerization. Once the resolving solution polymerized, the isopropanol was poured off and traces removed with Kimwipes (Irving, TX). A stacking gel (Table 3) was prepared in 5 mL and added onto the resolving gel followed by insertion of a 10-well comb. Alternatively, Mini-PROTEAN TGX stain-free precast 12% wt/vol acrylamide gels were purchased from Bio-Rad (Hercules, CA).

Table 3. SDS-PAGE resolving and stacking gel concentrations.

| Component | Final concentration in resolving gel | Final concentration in stacking gel |
|---|---|---|
| Tris-HCl pH 8.8 | N/A | 375 mM |
| Tris-HCl pH 6.8 | 125 mM | N/A |
| 40% wt/vol acrylamide/Bis (37.5:1) | 8% wt/vol | 10% wt/vol |
| 20% SDS wt/vol | 0.1% wt/vol | 0.1% wt/vol |
| 10% wt/vol ammonium persulfate | 0.05% wt/vol | 0.05% wt/vol |

### 2.4.4 Estimation of protein concentration

In order to estimate overall protein content in a solution, a NanoDrop One was used to read absorbance at 280 nm; 1 absorbance unit equated to approximately 1 mg/mL of protein. For more sensitive estimates, a Qubit Protein Assay Kit (Thermo Fisher, Waltham MA) was used.

### 2.4.5 Sample preparation, separation conditions and visualization techniques of SDS-PAGE gels

Samples to be separated on SDS-PAGE were first diluted 1:1 in 2X Lammeli sample buffer (Bio-Rad, Hercules, CA) with β-mercaptoethanol (βME) and heated to 95°C for 5-10 minutes to denature proteins. The SDS-PAGE gels were prepared and placed into an appropriate electrophoresis apparatus. A buffer dam was used when required to make a liquid-tight seal

between the inside of the cassette and the buffer reservoir. Running buffer containing 24.76 mM Tris base, 191 mM glycine and 3.467 mM SDS was placed inside the cassette such that the wells of the gels were covered. The buffer reservoir was filled to labeled marks on the container depending on the number of gels. A voltage between 170-220 volts was applied to the gels with a power pack (Bio-Rad, Hercules, CA) for 45-60 minutes or until the dye front reached the bottom of the gel. SDS-PAGE gels to be stained were submerged in Coomassie protein staining solution and incubated at room temperature for 15 minutes with light vertical orbital shaking at 19 rpm. The staining solution was recovered, and the gel washed with water twice before being de-stained in 45% vol/vol ethanol and 10% vol/vol acetic acid. Gels were visualized on a Bio-Rad Gel Doc XR+ UV-Vis transilluminator (Bio-Rad, Hercules, CA).

### 2.4.6 Amylose- and nickel- affinity chromatography

Lysates containing recombinant protein were diluted with 3-4 volumes of binding buffer and applied to 1-4, sequential, MBPTrap™ HP columns (GE Healthcare, Chicago, Il) or a 5 mL HisTrap™ HP column (GE Healthcare, Chicago, Il) with a P960 sample pump (GE Healthcare, Chicago, Il) at a flow rate of 0.8-1.5 mL/min. UV absorbance was monitored with a UV-900 ultraviolet lamp detector (GE Healthcare, Chicago, Il) set at 280 nm absorbance in line with an ÄKTA 10 Purifier (GE Healthcare, Chicago, Il) chromatography system to monitor chromatography progression. Instances hereafter referring to UV chromatogram are referring to the UV chromatogram of the subject chromatography run. The column(s) were washed with binding buffer until the UV absorbance values returned to baseline. Samples bound to a HisTrap™ HP column were washed with 1-3 column volumes of 20-40 mM imidazole followed by 1-3 column volumes 60-80 mM imidazole. Protein bound to MBPTrap™ HP columns were eluted with an isocratic gradient of 10 mM maltose and protein bound to HisTrap™ HP were eluted with an isocratic gradient of 500 mM imidazole. All fractions were collected with a Frac-950 fraction collector (GE Healthcare, Chicago, Il). For large (>1 L) scale amylose affinity purifications, an XK-

16 column (GE Healthcare, Chicago, Il) was packed with amylose resin (New England Biolabs, Ipswich, MA) to a final volume of 9 mL. The same column application and elution procedure was followed; samples were applied using a P960 sample pump, washed until baseline values and eluted with 10 mM maltose. All chromatography buffers were maintained at 4°C or on ice during chromatography procedures and all fractions were collected and stored at 4°C.

### 2.4.7 Protein sequencing sample preparation

Samples for protein sequencing were prepared following the guidelines of the University of Guelph, Mass Spectrometry Facility (UoG MSF). In brief, semi-purified MBP1218 and MBP1222 were manipulated in clean, sterile environment to avoid contaminating protein and transported to the UoG MSF by carrier at room temperature. An in-solution trypsin digestion was performed by the UoG MSF facility and peptides analyzed by liquid chromatography – mass spectrometry (LC-MS). Identified peptides were searched against proteins in the genome of *Gordonia* NB4-1Y and *E. coli* by the UoG MSF.

### 2.4.8 Size exclusion chromatography

Proteins produced from the pMAL vectors were further purified with a size exclusion column to separate degradation products from desired proteins. All size exclusion procedures were carried out on a Sephacryl 16/600 HR column (GE Healthcare, Chicago, Il). The column was equilibrated with a half column volume of water followed by two column volumes of 50 mM Tris-HCl and 150 mM NaCl at a flow rate of 0.5 mL/min. Fractions to be purified were first concentrated with an Amicon Ultra-15 centrifugal 30,000 NMWL filter unit (EMD Millipore, Burlington, MA) to a volume at or lower than 1 mL. All of the protein was applied onto the column with a P960 sample pump at a flow rate of 0.5 mL/min and the protein eluted with 50 mM Tris-HCl pH 7.5 and 150 mM NaCl. Eluted fractions were concentrated, and buffer exchanged into 50 mM Tris-HCl, 150 mM NaCl

and 10% vol/vol glycerol with an Amicon Ultra-15 30,000 NMWL filter unit (EMD Millipore, MA). All buffers were maintained at 4°C or on ice during chromatography procedures.

## 2.5.0 Enzymatic assessment and analyte detection and quantification conditions

### 2.5.1 Reaction conditions

All enzymatic reactions were carried out at room temperature in buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% vol/vol glycerol, 1 mM NADH, 1 µM FMN, 400 µM octane sulfonate or 6:2 FTSA and 1 µM of catalyst. Reactions were started by the addition of 0.2 µM of FreH and allowed to proceed for 2 hours in open-top 5-mL glass vials, and then spiked with hexanal to a final concentration of 8.13 mM. For kinetic assessment, reactions were carried out at room temperature in the same buffer as above however, FreH was added to a final concentration of 0.8 µM, catalyst at 0.2 µM and NADH at 500 µM.

Typically, a 1 mL reaction would produce two or four samples for analysis; one HPLC and one sulfite quantification sample or three gas chromatography (GC) and one sulfite quantification sample. In brief, a 1 mL reaction would be split into two 500 µL aliquots, one would be diluted with 500 µL of acetonitrile, the other extracted with 500 µL of ethyl acetate or both would be extracted with 500 µL of ethyl acetate. Ethyl acetate extraction was carried out in microcentrifuge tubes, vortexed and centrifuged at 13,000 *g* for 1 min to separate protein, aqueous and organic phases. For GC samples, aliquots of 100 µL of organic phase were removed with a graduated Hamilton glass syringe (Hamilton Company, Reno, NV) and transferred to 2-mL amber glass vials (Canadian Life Science, Peterborough, ON) with a 150-µL low volume glass insert (Canadian Life Science, Peterborough, ON) and spiked with decanal to a final concentration of 5.3 mM. For sulfite quantification, the remaining organic phase was removed with a glass Pasteur pipette and dried before transferring 400 µL of aqueous phase to a plastic cuvette containing 400 µL of 1 mM DTNB.

The instrument was first zeroed with a solution containing the same volumes of reaction buffer without catalyst or reductase and 1 mM DTNB prepared that day. For kinetic assessments reactions were stopped with 1 volume of 4 M urea, 0.1 M Tris-HCl, pH 8.0 and 1 mM DTNB. Acetonitrile diluted HPLC samples were first diluted in a microcentrifuge tube then transferred to a clean, amber chromatography vial with the PFTE side of the silicon/PFTE cap facing away from the aqueous samples as to not leach PFAS. Samples were stored at 4°C until shipped by carrier, at room temperature, to the facilities of Dr. Jinxia Liu at McGill University. HPLC destined samples were shipped with 3 controls, 100% acetonitrile, 75:25 and 50:50 acetonitrile:reaction buffer to assess contaminating 6:2 FTSA.

### 2.5.2 Sulfite oxidation assay

Sulfite oxidation assays were carried out using the same conditions used during the enzymatic assessment reactions with some notable changes. Catalysts were omitted and octane sulfonate and sulfite were added to 400 µM. Reactions were stopped with 1 volume of 50 mM Tris-HCl, pH 8.0, 4 M urea with 1 mM DTNB.

### 2.5.3 Spectrophotometric conditions

Absorbance readings from 260-412 nm and optical density readings at 660 nm were taken on a Varian Cary 50 UV-Visible spectrophotometer (Agilent, Santa Clara, CA) using plastic cuvettes with a 1 cm light path or using a NanoDrop One. All readings were taken at room temperature.

### 2.5.4 Gas chromatography – flame ionization detection conditions

Reaction extracts were analyzed on a Varian 3800 Gas Chromatograph (Agilent, Santa Clara, CA) equipped with a flame ionization detector (FID) and CombiPal Autosampler (Agilent, Santa Clara, CA) equipped with a 10-µL glass syringe (Hamilton Company, Reno, NV). The syringe was rinsed 3 times with 10 µL of ethyl acetate followed by sample before injecting 1 µL of sample onto

the column. Samples were separated on a 30 meter, 0.25 mm I.D. DB-5 column (Agilent, Santa Clara, CA) with a film thickness of 1.0 µm. Initial column temperature was maintained at 50°C for 5 minutes, the temperature was then increased to 200°C at a rate of 10°C/min, held for 3 minutes, then increased to 275°C at a rate of 20°C/min and held for a final 5 minutes. Total run time was 31.75 minutes. Analytes were loaded onto the column with a moving split ratio starting at an initial split ratio of 1:20 for 0.01 minutes followed by a split ratio of 1:5 for 0.06 minutes after which a constant 1:100 split ratio was maintained. Chromatogram data was analyzed using the Varian Workstation software (Agilent, Santa Clara, CA). Blanks containing only ethyl acetate were run every six samples to monitor for analyte contamination between samples.

### 2.5.5 Gas chromatography – mass spectrometry conditions

All gas chromatography – mass spectrometry (GC-MS) conditions were carried out with the same analyte separation temperatures as above on an Agilent 7890B GC-system (Agilent, Santa Clara, CA) paired with a 5977 Mass Selective Detector (MSD). Samples were delivered to the GC-system with a CombiPal Autosampler (Agilent, Santa Clara, CA) equipped with a 10 µL glass syringe (Hamilton Company, Reno, NV). The same column was used during GC-FID analysis save for the film thickness was 0.25 µm and the carrier gas was maintained at a flow of 1 mL per minute. A constant split ratio of 1:40 was used and a solvent delay of 4.5 minutes was set.

### 2.5.6 Analytical standard preparation

Three analyte calibration curves were prepared in this study: sulfite, octanal and octanol and are presented in Figures S1-3. In addition, hexanal was added to confirm analyte retention order in GC-FID and -MS.

Sulfite and thiols react equimolarly with DTNB producing equal moles of 2-nitro-5-thiobenzoate ($TNB^{2-}$). The molar extinction coefficient of $TNB^{2-}$ is 14.1 $mM^{-1}cm^{-1}$ and has been used to independently calculate the same kinetic parameters of SsuD as studies that followed NADH

consumption (Collier, 1973, Zhan *et al.*, 2008 and Eichhorn *et al.*, 1999). Being such, the molar extinction coefficient $TNB^{2-}$ was used to quantify sulfite production in this study, however, a calibration curve was prepared by mixing excess DNTB with concentrations varying between 5-200 µM sulfite to support the use of the molar extinction coefficient (Figure S4). Solutions of sodium sulfite were reacted with DTNB at room temperature for 3-5 minutes and absorbance readings taken at 412 nm; the instrument was zeroed with 500 µM sulfite with no DTNB.

Due to the different columns and separation conditions for GC-FID and GC-MS, two retention times were identified for each organic analyte. Hereafter, "on GC-FID or GC-MS" refers to the retention times of the organic analyte in question using the conditions described in Section 2.5.4 or 2.5.5, respectively.

In order to account for variation between gas chromatography separations, decanal was used as an internal standard. Decanal was added to 5.30 mM prior to each run and had a retention time of 17.94 minutes on GC-FID and 13.25 minutes on GC-MS. The response ratio of analyte to decanal peak area was used to construct calibration curves and calculate analyte concentrations in reaction extracts.

In order to confirm the retention order of organic analytes in this reaction were consistent between GC-FID and -MS, hexanal was spiked to a concentration of 8.13 mM prior to transferring reactions to microcentrifuge tubes. Hexanal had a retention time of 10.62 minutes on GC-FID and 4.15 minutes of GC-MS.

The two enzymatically interesting organic analytes are octanal and octanol. If enzymatic reactions turned over 100% of the substrate in solution, 400 µM of either analyte would be returned. Therefore, a calibration curve for octanal and octanol were prepared between 50 to 400 µM. Octanal had a retention time of 14.75 minutes on GC-FID and 9.5 minutes of GC-MS; octanol had a retention time of 15.8 minutes on GC-FID and 10.9 minutes on GC-MS.

On GC-FID and -MS chromatograms, peaks with areas larger than three times the standard deviation of the chromatogram baseline were considered. The limit of quantification in this study is reported as the response ratio that equated to 0 µM or mM on the trendline of any given calibration curve.

## 2.6.0 Mutagenesis and growth assay conditions

### 2.6.1 Conjugation and electroporation of *Gordonia* NB4-1Y

In order to conjugate pK18mobsacB1218AB and pK18mobsacB1222AB into *Gordonia* NB4-1Y, starter cultures of *E. coli* S17.1 carrying either plasmid and *Gordonia* NB4-1Y were grown to saturation. *Gordonia* NB4-1Y was grown for 4 days in LB or NB and *E. coli* S17.1 strains were grown for 2 days in LB or NB with 50 µg/ml of kanamycin. Cells were pelleted and washed twice with water and resuspended in water. To 1 mL of LB or NB, 1 volume of *Gordonia* NB4-1Y was added for every 3 volumes of respective *E. coli* S17.1 strain and incubated at 30°C. Every 1, 2, 3, 4, 6 and 8 days, 100 µL of culture was removed and plated on LB or NB agar facing up producing a mating spot and incubated at room temperature. One day after each plating, the mating spot was harvested from the plate, suspended in 1 mL of sterile water and further diluted to 1:100 and 1:1000 in sterile water. Diluted mating spots were plated on M9 minimal medium agar with kanamycin and incubated at 30°C until *Gordonia* NB4-1Y colonies appeared.

Alternatively, electroporation of *Gordonia* NB4-1Y with purified pK18mobsacB1218AB or pK18mobsacB1222AB following the same conditions described in section 2.3.11 was used at 1.8, 2.2 or 2.5 kV and cells were recovered for 2 hours in LB or NB at 30°C. Candidate transformant colonies were streak purified on LB or NB agar with 50 µg/mL of kanamycin.

### 2.6.2 Sulfur limiting growth assay: *E. coli* BL21(DE3)

In this study, two growth assays were employed: a 96-well plate assay following $OD_{621}$ and a biomass yield assay following $OD_{660}$. The former was modeled after the growth assays described by Eichhorn *et al* (2000) with the exception that M63 media was replaced with M9 media. In brief, strains to be tested were first grown in M9 minimal medium with glucose and diluted 1:100 in M9 minimal medium with glucose and no sulfur source added. The cells were then added to M9 minimal medium with glucose and appropriate sulfur source at a ratio of 1:10 and 100 µL aliquoted in sextuplicate to a 96-well plate. Each plate contained three blanks with no cells added. Optical density readings were taken with a Thermo Multiskan Ascent (ThermoFisher, Waltham, MA) spectrophotometer equipped with a 96-well plate reader placed inside a Forma Scientific Model 3110 Series Water Jacket Incubator (ThermoFisher, Waltham, MA). Readings were taken every hour for 48 hours at $OD_{621}$ with or without a 96-well plate lid and background shaking of 60 rpm for 5 seconds every 30 seconds.

The biomass yield assay was employed to mimic the 96-well plate conditions in a larger volume. In brief, a single isogenic colony of *E. coli* BL21(DE3) was inoculated in 50-100 mL of M9 minimal medium with glucose and grown overnight at 37°C. The following day, the cells were collected and washed twice with water and resuspended in 1-2 mL of water. In four separate, clean and sterile 125-mL flasks, *E. coli* BL21(DE3) was added to a final $OD_{660}$ of 0.05 to 100 mL of M9 minimal medium with glucose and 400 µM of $MgSO_4$, octane sulfonate, 6:2 FTSA or no added sulfur. The flasks were mixed, and 5 mL aliquots were dispensed to either 15- or 50-mL culture tubes. Cultures were incubated at 37°C with shaking at 250 rpm; culture tubes were sacrificed at 0, 24 and 48 hours and $OD_{660}$ readings taken. The instrument was zeroed with M9 minimal medium with glucose and no sulfur source.

Two oxygen conditions were maintained in both assays. Oxygen permissive, where oxygen could be replenished in the tube headspace and oxygen restrictive, where oxygen was restricted to that in the tube headspace and dissolved in the media. For the 96-well plate assay, oxygen restrictive conditions were maintained by incubating the cultures with a 96-well plate lid. The lid was removed for oxygen permissive conditions. For the biomass yield assay, oxygen restrictive conditions were achieved by sealing a 50-mL Kimax culture tube (Kimble-Chase, TN) with an airtight screw cap and oxygen permissive conditions were achieved by sealing a 15 mL disposable culture tube with a KimKap cap (Kimble-Chase, TN). The protrusion on the inner surface of the KimKap cap produced a small gap allowing for oxygen diffusion in the culture tube while maintaining sterility.

## 2.7 Statistical analysis of raw data and means

In order to determine if a calculated mean is statistically significant with respects to another, four statistical analyses were performed: a Ryan-Joiner normality test, a Levene's two sample variance test, a two sample T-test and a Welch T-test. Raw data was output to and analyzed with Minitab 19 (Minitab, State College, PA) software. In brief, a data set was first analyzed with a Ryan-Joiner normality test and if normally distributed, the variance of normally distributed data sets were compared with a Levene's two sample variance test. If equal, a two sample T-test was used to compare the means and if data sets had unequal variance, a Welch T-test was used. For each test, the null hypothesis is the data points are normally distributed, variance is the same and the means are the same. Statistical significance was determined by p-values lower than 0.05. Values reported as N/D are not determined due to lack of comparison.

## 2.8 Phylogenetic analysis of Class C monooxygenases

### 2.8.1 Collection of amino acid sequences

Amino acid in this study were collected from the UniProt (UniProt Consortium) or the National Center for Biotechnology Information (NCBI, Bethesda, MD). Sequences from the *Gordonia* NB4-1Y genome were collected from APHK00000000.2 assembly. Sequences collected from UniProt were associated with experimental evidence at the protein or mutant level informing on the protein substrate.

### 2.8.2 Phylogenetic tree construction parameters

Phylogenetic analysis was carried out using the Molecular Evolutionary Genetics Analysis (MEGAX, Penn State University, PA, Kumar *et al*. 2018) software package. All sequences to be analyzed were imported in FASTA format and aligned with the alignment explorer function in MEGAX. Alignment parameters were followed using the ClustalW (Conway Institute, University College Dublin, Ireland) option with the following parameters: pairwise alignment gap opening penalty of 10.00 and a gap extension penalty of 0.10; multiple alignment gap opening penalty of 10.00 and a gap extension penalty of 0.20. A negative weight matrix was set to off and delay divergent cutoff percentage was set to 30.

Once aligned, a phylogenetic tree was constructed using the neighbor-joining method (Saitou and Nei, 1987) and bootstrap support was calculated over 2000 replicates (Feksenstein, 1985). Evolutionary distances were calculated using the Poisson correction method (Zuckerland and Pauling, 1956) and ambiguous positions removed for each sequence pair.

# 3.0 Results

## 3.1 Comparison of the operon-like regions surrounding the genes encoding ISGA 1218, 1222 and *ssuD*-like genes in *Gordonia* NB4-1Y

In order to support the hypothesis that ISGA 1218 and 1222 are involved in the degradation of 6:2 FTSA, it was necessary to compare the genomic context of each monooxygenase to the genomic context of the *ssu* operon. The *ssu* operon of *E. coli* K-12, which is identical in *E. coli* BL21(DE3), is responsible for the partial metabolism of octane sulfonate, a structurally similar compound to 6:2 FTSA. The *ssu* operon includes five genes: a monooxygenase (*ssuD*), a reductase (*ssuE*) and three ABC type transporter proteins (*ssuA*, *B*, and *C*). The transport proteins include an aliphatic ATP-binding protein (*ssuB*), permease (*ssuC*) and substrate-binding protein (*ssuA*). For the purpose of this study, genes in the *Gordonia* NB4-1Y genome were referend to by their initial annotation with the ISGA prefix. When compared to the *ssu* operon from *E. coli*, the genomic context of the genes encoding ISGA 1218 and 1222 in *Gordonia* NB4-1Y is similar and includes an ATP-binding protein (ISGA 08710), a permease (ISGA 1221) and a substrate-binding protein (ISGA 1220) but no corresponding *ssuE*-like reductase (Figure 6).

A search of the *Gordonia* NB4-1Y genome to identify potential *ssuD*-like genes revealed seven other putative monooxygenases and are as follows: the genes encoding ISGA 205, 1666 and 1835 which are annotated as alkanesulfonate monooxygenases and the genes encoding ISGA 3420, 4167, 08960 and 08770 which are annotated as luciferase-like monooxygenases. The latter four were the top four hits of a blastp search against the *Gordonia* NB4-1Y genome using SsuD as a query. ISGA 205, 1666 and 1835 are 36.30%, 45.97% and 22.07% similar in amino acid sequence to SsuD and, on average, 71.33 amino acids longer. ISGA 3420, 4167, 08960 and 08770 are 33.43, 33.07, 46.63 and 29.76% similar to SsuD and, on average, 19.33 amino acids longer.

Of the seven identified monooxygenases, the genes encoding ISGA 205, 1835 and 08960 have genetic contexts most similar to s*suD* due to the presence of an ABC transporter permease, ATP-binding protein and substrate binding protein. The genes encoding ISGA 1666 and 08770 are not encoded nearby a substrate binding protein or an ATP binding protein while the genes encoding ISGA 3420 and 4167 are not encoded nearby any putative transporter proteins. None of the reported genes are encoding nearby an *ssuE*-like reductase.

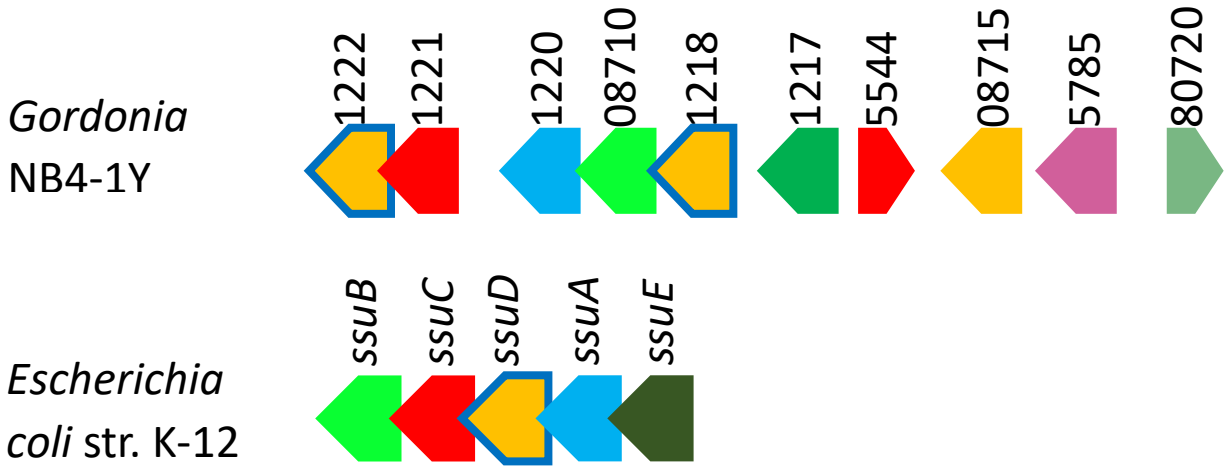


Figure 6. Genomic context surrounding the genes encoding ISGA 1218 and 1222 (top) in *Gordonia* NB4-1Y and *ssuD* (bottom) in E. coli K-12. The numbers above each gene represent their respective ISGA number. In the region surrounding the genes encoding ISGA 1218 and 1222 there is predicted to be three ABC type transporter proteins (**1221**, **1220** and **08710**), a tyrosine phosphatase (**1217**), a *tetR* family transcriptional regulator (**5544**), a *dsbA* oxidoreductase (**08715**), a *lysR* family transcriptional regulator (**5785**) and a FAD linked oxidase (**80720**). The region surrounding *ssuD* in *E. coli* encodes three aliphatic sulfonate transporter proteins (*ssuB, C*, *A*) and an FMN reductase (*ssuE*). Blue trimmed genes are reference monooxygenases.

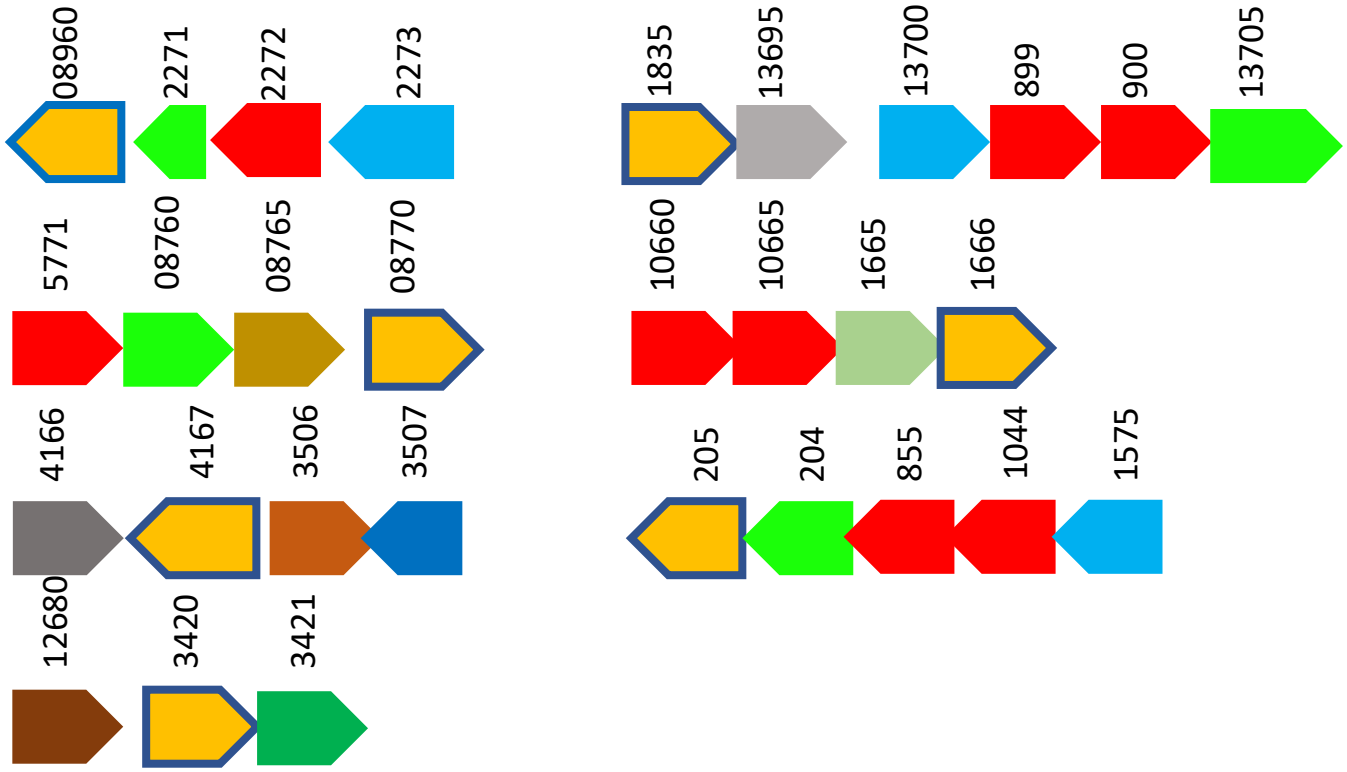


Figure 7. Genomic context of *ssuD*-like monooxygenases (left) and annotated alkanesulfonate monooxygenase (right) in the genome of *Gordonia* NB4-1Y. The numbers above each gene correspond to their respective ISGA number and are annotated as follows: **08960**: LLM class flavin-dependent oxidoreductase; **2271**: ABC transporter ATP-binding protein; **2272**: ABC transporter permease; **2273**: aliphatic sulfonate ABC transporter substrate-binding protein; **5771**: iron ABC transporter permease; **08760**: ABC transporter ATP-binding protein; **08765**: hypothetical protein; **08770**: LLM class flavin-dependent oxidoreductase; **4166**: endonuclease/exonuclease/phosphatase family protein; **4167**: LLM class flavin-dependent oxidoreductase; **3506**: acetyltransferase; **3507**: 4-hydroxbenzoate 3- monooxygenase; **12680**: ketopantoate reductase family protein; **3420**: LLM class flavin-dependent oxidoreductase; **3421**: DUF1684 domain-containing protein; **1835**: alkanesulfonate monooxygenase; **13695**: oxidoreductase; **13700**: ABC transporter substrate-binding protein; **899**: ABC transporter permease; **900**: ABC transporter permease; **13705**: ABC transporter ATP-binding protein; **10660**: ABC transporter permease; **10665**: ABC transporter permease; **1665**: acyl-CoA dehydrogenase; **1666**: alkanesulfonate monooxygenase: **205**: alkane sulfonate monooxygenase; **204**: ABC transporter ATP-binding protein; **855**: ABC transporter permease; **1044**: ABC transporter permease; **1575**: peptide ABC transporter substrate-binding protein.

## 3.2 Phylogenetic analysis of Class C monooxygenases and subject *Gordonia* NB4-1Y enzymes

### 3.2.1 Collection of Class C monooxygenases and subject *Gordonia* NB4-1Y enzymes

Hujibers *et al*. (2014) described 12 monooxygenase archetypes in the Class C monooxygenases group. In order to understand the phylogenetic placement ISGA 1218, 1222, 205, 1835, 1666 and 08960 among these archetypes, 14 representatives were collected from UniProt or NCBI. Experimental evidence for each representative is found in the following: Eichhorn *et al*. (1999), van der Ploeg *et al*. (1998), Kahnert *et al*. (2000), Fisher *et al*. (1996), Li *et al*. (2007), Feng *et al*. (2006), Denome *et al*. (1994), Uetz *et al*. (1992), Thibaut *et al*. (1995), Boden *et al*. (2011), Iwaki *et al*. (2013), Jun *et al*. (2016) and Mukherjee *et al*. (2010). A listing of all archetypes, their protein names and originating organism are given in Table 6.

Table 4. Class C monooxygenases archetypes as described by Hujibers *et al*. (2014).

| Full name | Protein name(s) | Organism(s) | Experimental evidence[1] |
|---|---|---|---|
| Alkanal monooxygenase | LuxA | *Vibrio harveyi* | Yes |
| Alkanesulfonate monooxygenase | SsuD_1 | *Escherichia coli,* | Yes[2] |
| | SsuD_2 | *Bacillus subtilis* | Yes |
| | SsuD_3 | *Pseudomonas putida* | Yes |
| Dimethylsulfide monooxygenase | DmoA | *Hyphomicrobium sulfonivorans* | Yes |
| 3,6-diketocamphane monooxygenase | CamE36 | *Pseudomonas putida* | Yes |
| 2,5-diketocamphane monooxygenase | CamP | *Pseudomonas putida* | Yes |
| Long-chain alkane monooxygenase | LadA | *Geobacillus thermodenitrificans* NG80-2 | Yes |
| Nitrilotriacetate monooxygenase | NtaA | *Aminobacter aminovorans* | Yes |
| Dibenzothiophene monooxygenase | DszA, SoxA | *Rhodoccocus* sp. IGTS8 | Yes |
| Dibenzothiophene sulfonate monooxygenase | DszB, SoxB | *Rhodoccocus* sp. IGTS8 | Yes |
| Pristinamycin II synthase | SnaA | *Streptomyces pristinaespiralis* | Yes |

| | | | |
|---|---|---|---|
| Ethylenendiaminetetraacetate monooxygenase | EmoA | *Chelativorans* sp. BNC1 | Yes |
| Pyrimidine oxygenase | RutA | *Escherichia coli* | Yes |

[1] Experimental evidence is considered *in vitro* chemical assessment of substrate transformation or *in vivo* mutations demonstrating substrate specificity of the enzyme in question.

[2] No direct experimental evidence has been recorded for the SsuD of *B. subtilis* in the manner described in [1], however, gene disruption studies described in van der Ploeg *et al*. (1998) are consistent with the *ssu* operon. Therefore, the SsuD of *B. subtilis* was considered.

Table 5. *Gordonia* NB4-1Y enzymes considered for phylogenetic placement among Class C monooxygenases.

| Protein | Accession[1] | Contig | New locus tag [2] |
|---|---|---|---|
| ISGA 1218 | EMP10004.2 | 54 | ISGA_RS09775 |
| ISGA 1222 | EMP10005.1 | 54 | ISGA_RS09755 |
| ISGA 205 | EMP14314.1 | 72 | ISGA_RS15565 |
| ISGA 1666 | EMP12617.1 | 74 | ISGA_RS14690 |
| ISGA 1835 | EMP12962.1 | 143 | ISGA_RS24605 |
| ISGA 08960 | KOY49635.1 | 58 | ISGA_RS10415 |

[1] Accession numbers given by the APHK00000000.2 assembly.

[2] New locus tags were given upon re-submission of the *Gordonia* NB4-1Y genome in 2015

### 3.2.2 Phylogenetic placement of *Gordonia* NB4-1Y enzymes among Class C monooxygenases

In order to assess the phylogenetic relationship of ISGA 1218, 1222, 205, 1666, 1835 and 08960 among Class C monooxygenase, a neighbor-joining phylogenetic tree was constructed with the mentioned proteins and known Class C monooxygenases (Table 6 and 7). The resulting tree produced four apparent clades (Figure 7). ISGA 1218 and 1222 formed a clade with DszA, NtaA, SnaA and EmoA; ISGA 1218 was found next to DszA and ISGA 1222 next to EmoA. ISGA 205, 1666 and 1835 formed a clade with LadA and DmoA. None of the subject *Gordonia* NB4-1Y enzymes were found next to the clade formed by LuxA, CamE36 and CamP. The final clade was formed with the SsuD from *E. coli*, *B. subtilis* and *P.* putida and ISGA 08960. DszB acted as an outgroup to the phylogenetic tree and RutA did not cluster in any apparent clade. Bootstrap values for all nodes 57% or higher.

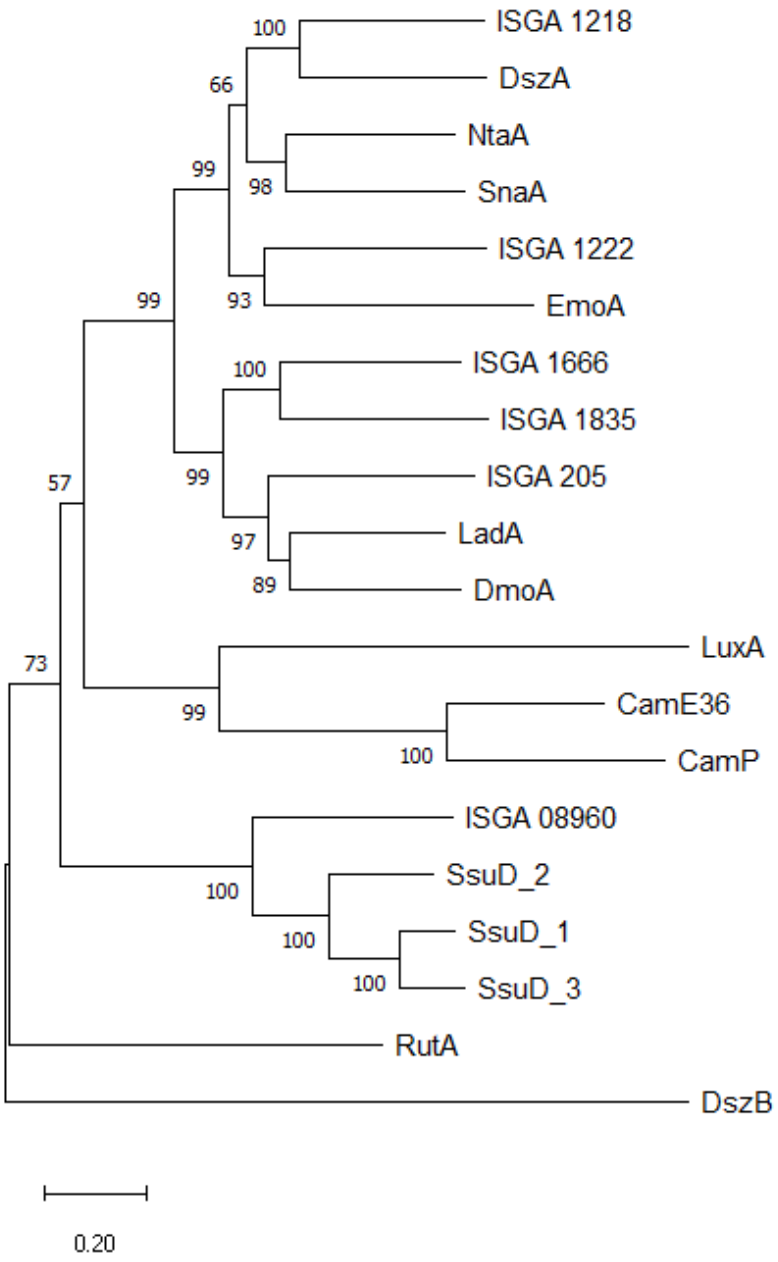


Figure 8. Phylogenetic grouping of Class C monooxygenases (Hujiber *et al*. 2014) and of *Gordonia* NB4-1Y monooxygenases. See Section 2.8.2 for full alignment and phylogenetic tree construction methods. A total of 20 amino acid sequences were involved for a total of 549 position; all ambiguous positions were removed for each sequence pair.

## 3.3 Construction of protein production and mutagenesis vectors

In order to characterize genes potentially involved in 6:2 FTSA metabolism, the genes encoding ISGA 1218 and 1222, SsuD were cloned into pET28b and pMAL-c2 vectors, and the known $FMNH_2$ producing NADH:FMN oxidoreductase, Fre, in the pET28b vector, resulting in pET28b1218, pET28b1222, pET28bSsuD, pMAL1218, pMAL1222, pMALSsuD and pET28bFre vectors. Fre has no known structural interaction with ISGA 1218, 1222 or SsuD. Respectively, these vectors were used to produced, with a hexhistidine (H) or maltose binding protein (MBP) tag, recombinant enzymes in *E. coli* BL21(DE3); specifically, 1218H, 1222H, SsuDH, MBP1218, MBP1222, MBPSsuD, and FreH. The genes encoding ISGA 205, 1666 and 1835, which are putative alkanesulfonate monooxygenases, were cloned into the pMAL-c2 vector generating the pMAL205, pMAL1666 and pMAL1835 vectors. These vectors were used to respectively produce the MBP205, MBP1666 and MBP1835. Vector maps of protein production plasmids are given in Figure S19-21.

In order to construct the pET28b1218, pET28b1222, pET28bSsuD and pET28bFre vectors, primers were designed to bind the 5′ and 3′ end of each gene and encoded an NcoI cut site or HindIII cut site. A list of primers can be found in Table 1. Amplicons of the genes encoding ISGA 1218 and 1222, SsuD and Fre were 1518, 1346, 1191 and 747 base pairs (bp) respectively. Each amplicon was treated with NcoI and HindIII, ligated into pET28b, transformed into *E. coli* BL21(DE3) and confirmed by restriction digest from an *E. coli* BL21(DE3) isogenic culture (Figure 8 and 9) carrying one of the aforementioned vectors. The resulting 1218H, 1222H, SsuDH and FreH proteins are predicted to be 55.21, 50.67, 43.25 and 27.96 kDa respectively.

The pMAL1218 and 1222 plasmids were previously prepared (Milton-Wood, 2016) and the genes encoding ISGA 1218 and 1222 were excisable from each plasmid (Figure 8). In order to produce pMALSsuD, pMAL205, pMAL1666 and pMAL1835, EcoliSsuD_F(2), EcoliSsuD_R(2) and the

primers designed by McAmmond (2017) were used to amplify 1191, 1431, 1386 and 1398 bp amplicons for *ssuD* and the genes encoding ISGA 205, 1666 and 1835, respectively. Amplicons were treated with HindIII and EcoRI or XbaI, transformed and confirmed following the same method used for pET28b vectors (Figure 8 and 11). The resulting MBP1218, MBP1222, MBPSsuD, MBP205, MBP1666 and MBP1835 are predicted to be 96.40, 91.86, 84.67, 95.67, 93.87 and 95.11 kDa, respectively.

*Gordonia* NB4-1Y mutagenesis was attempted with the pK18mobsacB mutagenesis vector. In brief, 1000 bp upstream (A fragment) and downstream (B fragment) of ISGA 1218 and 1222 were amplified, ligated together and ligated into pK18mobscaB. Mutagenesis occurs through recombination with one fragment followed by second recombination event releasing the gene and vector leaving a restriction enzyme cut site scar in the genome where the A and B fragments were ligated *in vitro*. In order to produce pK18mobsacB1218AB and pK18mobsacB1222AB, the A and B fragment flanking each gene were amplified independently. The 1218 A fragment was amplified with 18AF - HindIII and 18AR - XbaI primer pair and the B fragment was amplified with 18BF - XbaI and 18BR - BamHI primer pair. The 1222 A fragment was amplified with 22AF - SacI and 22AR - XbaI primer pair and the B fragment with 22BF - XbaI and 22BR - EcoRI primer pair. The ligated AB fragment of 1222 was first cloned into the pET23d vector, excised with SalI and EcoRI and ligated into pK18mobsacB. Each plasmid was transformed into *E. coli* S17.1 and confirmed by re-isolating the plasmid from an isogenic *E. coli* S17.1 culture (Figure 10). Vector maps of mutagenesis plasmids are given in Figure S22.

In order to confirm the nucleotide sequence of the gene or fragment within each vector, Sanger sequencing was performed using sequencing primers outlined in Table 1. Sanger sequencing data was aligned with *in silico* constructs in order to confirm gene or fragment identity and similarity is reported as the percentage match, with respects to total gene or fragment length, between the Sanger sequence alignment and *in silico* constructs. Sanger sequencing revealed a

100% nucleotide similarity of each gene in pET28b when compared to *in silico* constructs. Sanger sequencing revealed an upwards of 99% similarity of the genes encoding ISGA 1218 and 1222 in pMAL1218 and pMAL1222 when compared to *in silico* constructs and upwards of 90% similarity of the genes encoding ISGA 205, 1666 and 1835 and *ssuD* in pMAL205, 1666, 1835 and SsuD. Sanger sequencing alignment and excision of amplicons corresponding to the sizes of the genes encoding ISGA 205, 1666, 1835 and *ssuD* from pMAL205, 1666, 1835 and SsuD was considered sufficient for these vectors only. Sanger sequencing revealed a 99.75% similarity of the ligated AB fragments within pK18mobsacB1222AB when compared to the *in silico* construct and 99.25% similarity of the ligated AB fragment in pK18mobsacB1218AB when compared to the *in silico* construct. The 0.25% dissimilarity for pK18mobsacB1222AB is found adjacent to the EcoRI cut site and considered inconsequential due to the 100% similarity between the pET23d1222AB T7 and T7T Sanger sequencing and the *in silico* pET23d1222AB. The 0.75% dissimilarity of pK18mobsacB1218AB is divided between 0.35% missing near the HindIII cut site and 0.40% near the connection between the A and B fragment. The 0.35% dissimilarity was considered inconsequential due the restriction digestibility of the A fragment with HindIII indicating that the HindIII cut site is intact. The final 0.40% missing similarity is a CGTCTAGA insertion near the XbaI cut site linking the A and B fragments and is not found in the *in silico* construct of pK18mobsacB1218AB; if mutagenesis proceeds as predicted, a TCTAGACGTCTAGA instead of a TCTAGA scar would remain.

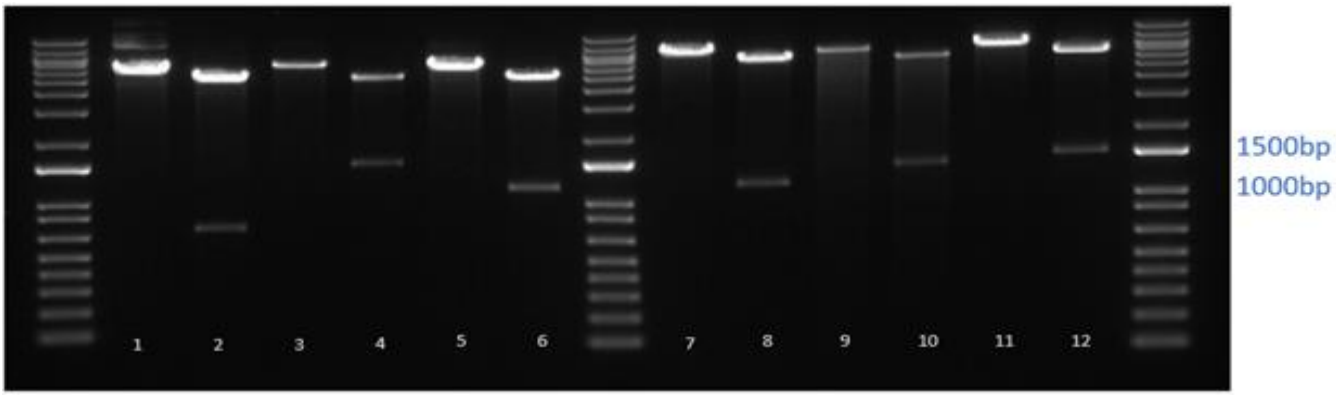


Figure 9. Restriction digestion analysis of pET28bFre, pET28b1218, pET28bSsuD, pMALSsuD, pMAL1222 and pMAL1218. Each restriction digestion for pET28b based vectors was done with NcoI and HindIII and pMAL-c2 based vectors with EcoRI and HindIII. All single digestions were done with HindIII. **1-2** single and double digestion of pET28bFre; **3-4** Single and double digestion of pET28b1218; **5-6** single and double digestion of pET28bSsuD; **7-8** single and double digestion of pMALSsuD; **9-10** single and double digestion of pMAL1222; **11-12** single and double digestion of pMAL1218. All DNA ladders are the 1 Kb Plus DNA Ladder (Invitrogen, Carlsbad, CA). Amplicons of *fre* are 747 bp, amplicons of the gene encoding ISGA 1218 are 1518 bp, amplicons of *ssuD* are 1191 bp and amplicons of the gene encoding ISGA 1222 are 1346 bp.

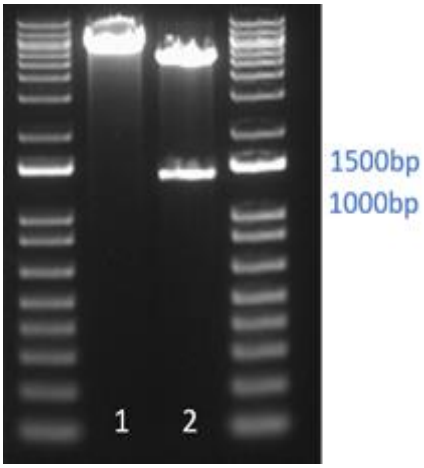


Figure 10. Restriction digestion analysis of pET281222. Restriction digestion was done with NcoI and HindIII; **1** digestion with HindII **2** double digest. All DNA ladders are the 1 Kb Plus DNA Ladder (Invitrogen, Carlsbad, CA). Amplicons of the gene encoding ISGA 1222 are 1346 bp.

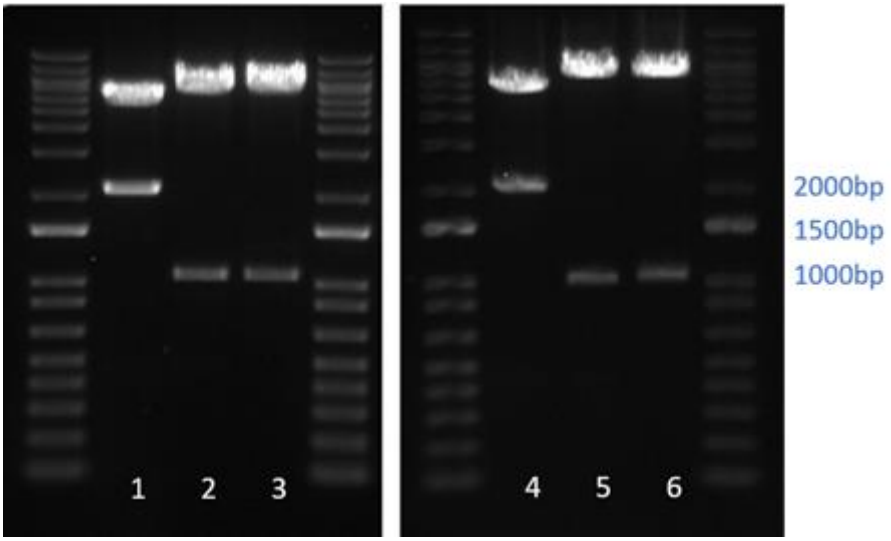


Figure 11. Restriction digestion analysis of pK18mobsacB1218AB and pK18mobsacB1222AB. Restriction digestions were done with a combination of BamHI, HindIII, XbaI, EcoRI, SalI or SacI. **1**-**3** double digestions of pK18mobsacB1218AB with BamHI and HindIII (**1**), HindIII and XbaI (**2**) and BamHI and XbaI (**3**); **4**-**6** double digestion of pK18mobsacB1222AB with SalI and EcoRI (**4**), SmaI and XbaI (**5**) and EcoRI and XbaI (**6**). All DNA ladders are the 1 Kb Plus DNA Ladder (Invitrogen, Carlsbad, CA). AB fragments are 2000 bp and A and B fragments are 1000 bp.

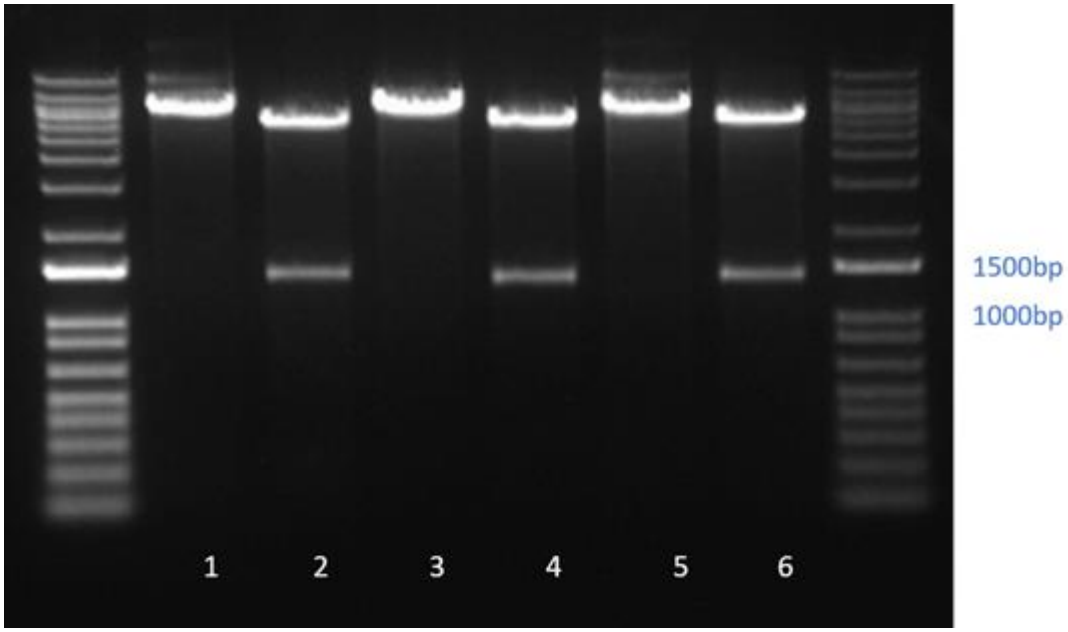


Figure 12. Restriction digestion analysis of pMAL205, pMAL1666 and pMAL1835. Each restriction digestion was carried out with EcoRI and HindIII (**1**-**4**) or XbaI and HindIII (**5**-**6**). All single digestions were done with HindIII. **1**-**2** single and double digestion of pMAL205; **3**-**4** single and double digest of pMA1666; **5**-**6** single and double digest of pMAL1835. All DNA ladders are the 1 Kb Plus DNA Ladder (Invitrogen, Carlsbad, CA). Amplicons for ISGA 205 are 1431 bp, amplicons for ISGA 1666 are 1386 bp and amplicons for ISGA 1666 are 1398 bp.

## 3.4 Protein production from pMAL and pET vectors and purification by amylose and nickel affinity chromatography

In order to determine the best protein production conditions for pMAL based vectors, small-scale protein productions assays were carried out in *E. coli* BL21(DE3), carrying pMAL1218, by inducing 10 mL cultures at mid log phase ($OD_{660}$ of 0.5) with IPTG at 0, 0.3 and 0.6 mM and incubating 18, 30 and 37°C. Protein production was carried out for two hours and was chosen based on recommendations (Nelson 2017, personal communication) and preliminary pilot assays demonstrating little increase in protein production past two hours (Figure S11). In brief, induction with IPTG concentrations greater than 0.3 mM did not increase MBP1218 yields, however, induction temperatures below 37°C lowered MBP1218 yields as visualized by SDS-PAGE (Figure S12). MBP1218 yields were qualitatively estimated by comparing band intensity on protein normalized SDS-PAGE gels and sample small-scale protein production assays are given in the Appendix (Figure S12-14). The optimal protein production conditions for MBP1222 and MBPSsuD followed the same trend (Figure S13 and S14), and therefore, the optimal production conditions for MBP1218 were used to produce MBP1222, MBPSsuD, MBP205, MBP1666 and MBP1835.

Following, partial purification of MBP1218 was used as a model to assess MBP tagged protein purification from whole cell lysates. When lysate containing MBP1218 was applied to amylose resin, washed with binding buffer to baseline UV signal and eluted, a single peak on the UV chromatogram was seen (Figure S17). The size of the eluted proteins in that peak were visualized by SDS-PAGE and found to be the expected size of MBP1218 (100 kDa) with a number of smaller proteins between 43 and 100 kDa observed (Figure 12). A number of proteins on SDS-PAGE between 43 and 100 kDa were also observed following partial purification of MBP1222, MBPSsuD, MBP205, MBP1666 and MBP1835 (Figure 12, 1222 **AC** and SsuD **AC** and Figure 13, **1**-**3**). Of the MBP tagged enzymes purified, MBPSsuD and MBP205 qualitatively had the highest

ratio of MBP tagged target to smaller proteins between 43 and 100 kDa (Figure 12, SsuD **AC** and 13, **3**).

The smaller contaminating proteins between 43 and 100 kDa on SDS-PAGE were assumed to be proteolytic degradation products of recombinant protein due to MBP being a native *E. coli* BL21(DE3) protein, the size of MBP being ~42.5 kDa and the lack of proteins smaller than 43 kDa. To enrich higher molecular weight proteins, size exclusion chromatography was employed and MBP1218 used as a model. When separated on a Sephacryl 16/600 HR column (GE Healthcare, Chicago, Il), two peaks were observed on the UV chromatogram; the first around 80 minutes and the second around 130 minutes (Figure S18). Using SDS-PAGE, proteins eluting in the first peak were found to be larger than 43 kDa, and the proteins in the second peak were smaller than 43 kDa (Figure S15). Protein yields of fractions containing desired protein post amylose affinity and size exclusion chromatography were estimated by NanoDrop One and reported with respect to grams of cell paste (Table 4). In summary, MBP205, 1666 and 1835 produced the highest protein yield per gram of cell paste post amylose affinity chromatography. MBP1666 and 1835, however, contained the largest amount of degradation products as seen by SDS-PAGE (Figure 13, **1** and **2**) and all contained an unknown protein near 11 kDa. Following, protein yields per gram of cell paste after amylose affinity and size exclusion chromatography are ranked highest to lowest as follows: MBPSsuD, MBP1222 and MBP1218. No unknown proteins near 11 kDa were observe by SDS-PAGE following MBPSsuD, MBP1222 and MBP1218 amylose affinity or size exclusion chromatography.

Table 6. Protein yields MBP tagged proteins post amylose affinity and size exclusion chromatography.

| Protein produced | Post amylose affinity chromatography (mg of protein per gram of cell paste) | Post size exclusion chromatography (mg of protein per gram of cell paste |
|---|---|---|
| MBP1218 | 1.7 - 2.3 | 0.2 |
| MBP1222 | 2.9 – 3.1 | 1.0 |
| MBPSsuD | 4.3 - 9.0 | 1.7 |
| MBP205 | 6.6 | N/D |
| MBP1666 | 7.6 | N/D |
| MBP1835 | 7.1 | N/D |

In this study, SsuDH and FreH were produced by inducing *E. coli* BL21(DE3) at $OD_{660}$ of 0.5 with 0.5 mM IPTG and incubating for 16 hours. In previous studies (Gao *et al.*, 2005 and Eichhorn *et al.*, 1999), SsuD had been induced for 5-6 hours; here 16 hours was chosen in order to produce more protein and allow for overnight induction. When lysate containing SsuDH or FreH were applied to nickel resin, washed to baseline UV signal and eluted, a single peak was seen on the UV chromatogram (Figure S19). The eluted proteins were visualized by SDS-PAGE and found to be the size of the desired recombinant protein (Figure 13). It was, however, found that binding buffer containing 20 mM imidazole and sequential washes with 40 followed by 80 mM of imidazole significantly decreased the number of non-recombinant proteins eluted from the column (Figure S16). Purification of 1218H and 1222H were attempted by with the same methods as above, and like the elution of SsuDH and FreH, proteins eluted in a single peak on the UV chromatogram. When visualized by SDS-PAGE, these peaks did containe proteins ranging from 11 to 100 kDa and none were disproportionally produced over any other. Furthermore, the only chromatographic step to contain what is thought to be 1218H or 1222H was the pellet of insoluble fragments following lysis (Figure 14, **P** and Figure 15, **P**).

Table 7. Protein yields following nickel affinity chromatography of SsuDH, FreH, 1218H and 1222H.

| Protein produced | Post nickel affinity chromatography (mg of protein per gram of cell paste) |
|---|---|
| SsuDH | 6.5 |
| FreH | 4.0 |
| 1218H | N/D |
| 1222H | N/D |

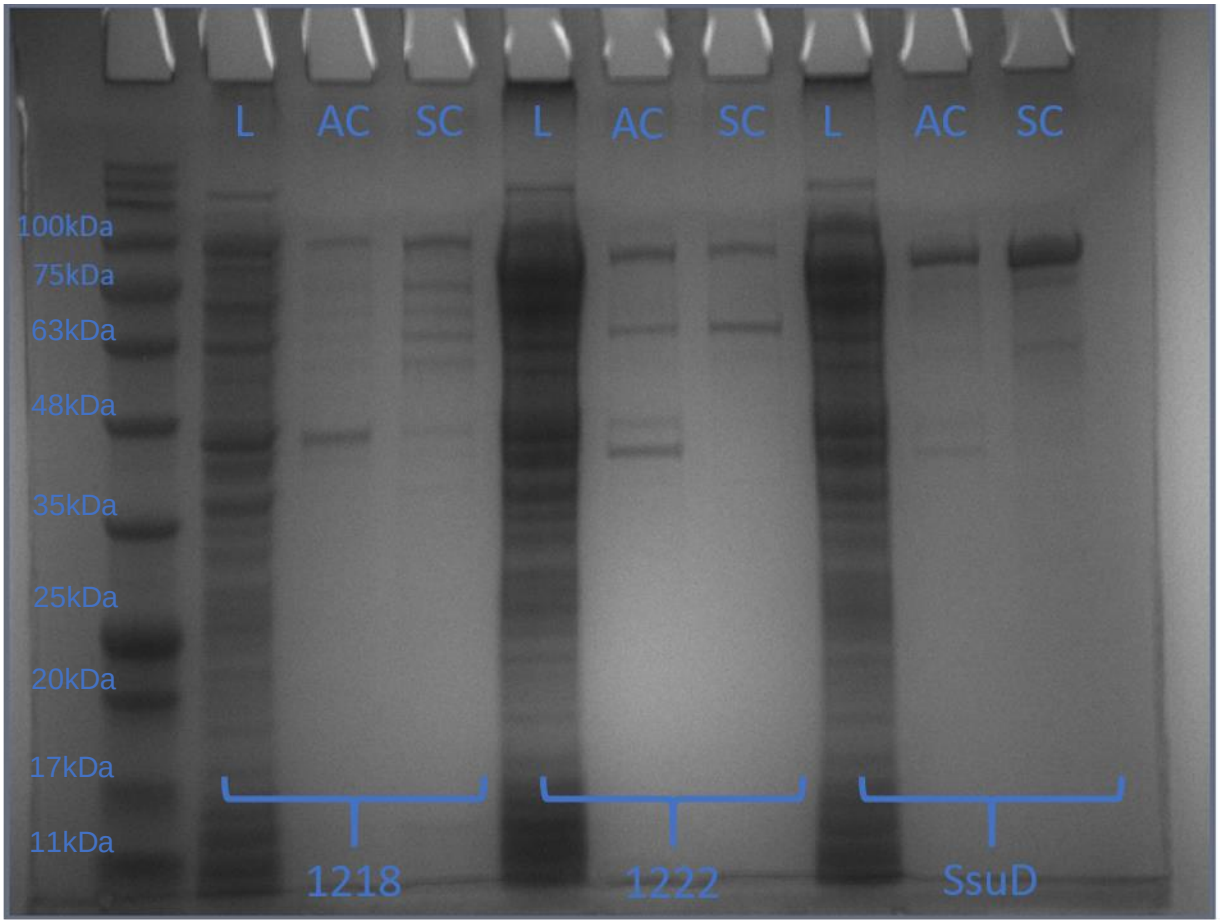


Figure 13. Partial purification and high molecular weight enrichment of MBP1218, MBP1222 and MBPSsuD. **L** denotes lysate of cultures induced to produce one of the MBP tagged monooxygenases, **AC** denotes concentrated amylose affinity semi-purified fractions and **SC** denotes concentrated size exclusion high molecular weight enriched fractions. The protein ladder is the BLUelf Prestained Protein ladder (FroggaBio, Toronto, ON). MBP1218 is predicted to be 96.40 kDa; MBP1222, 91.86 kDa; and MBPSsuD, 84.67 kDa.

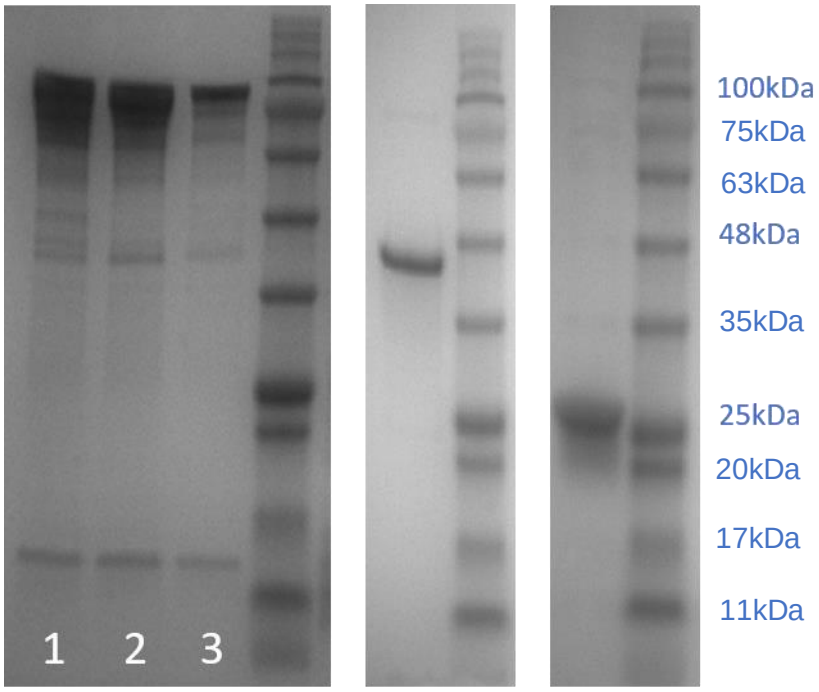


Figure 14. Partial purification of MBP1835 (**1**), MBP1666 (**2**) and MBP205 (**3**), purification of SsuDH (middle) and FreH (Right). The protein ladder is the BLUelf Prestained Protein ladder (FroggaBio, Toronto, ON). MBP1835 is predicted to be 95.11 kDa; MBP1666, 93.82 kDa; MBP205, 95.12 kDa; SsuDH, 43.45 kDa; and FreH, 29.96 kDa.

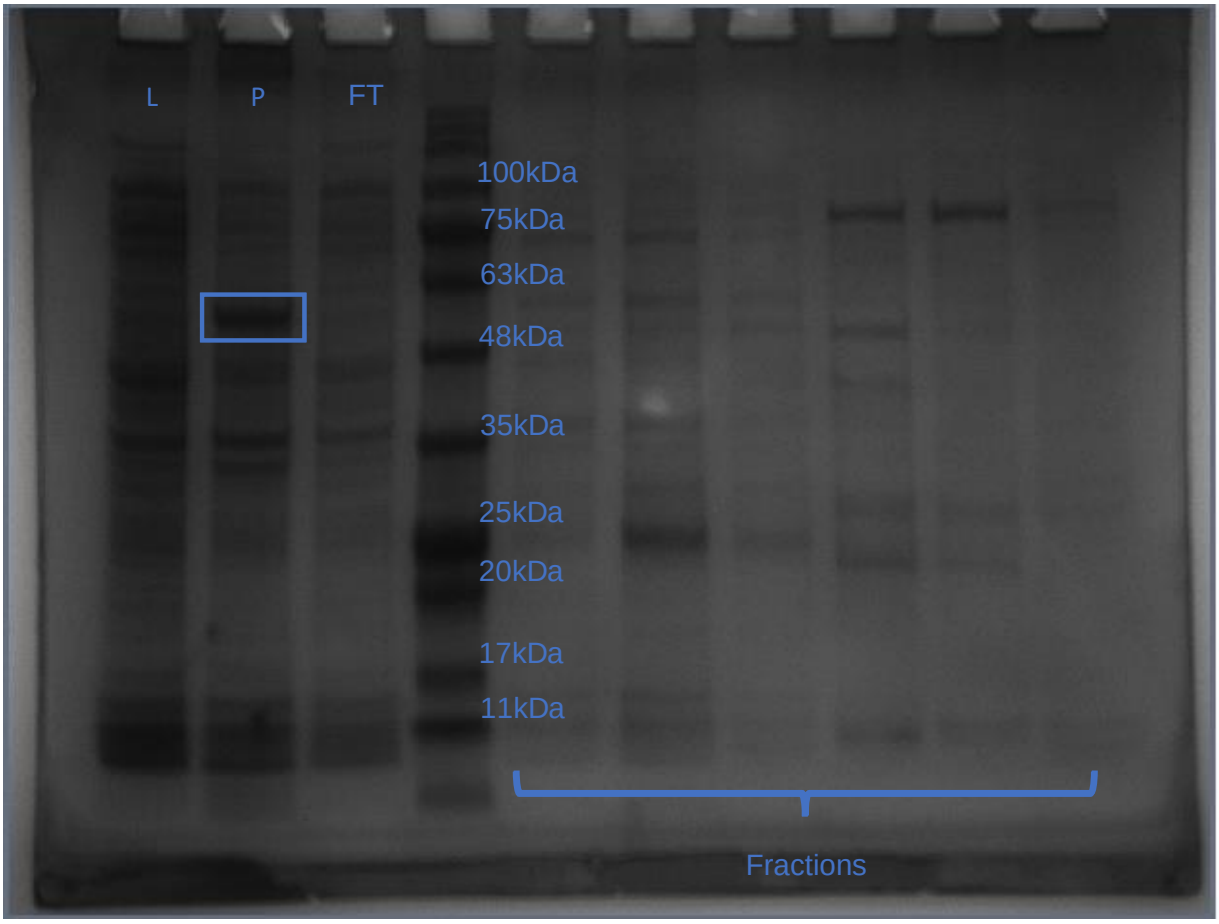


Figure 15. Attempted purification of 1218H from the pET28b1218 by nickel affinity chromatography. The band highlighted by the blue rectangle represents what is thought to be 1218H. **L** denotes lysate pre-column application; **P** denotes insoluble fraction separated by centrifugation and **FT** denotes flow through fractions collected after applying lysate to the column. Protein bound to the column was eluted in a single peak on the UV chromatogram on an isocratic gradient of 500 mM imidazole. The protein ladder is the BLUelf Prestained Protein ladder (FroggaBio, Toronto, ON). 1218H is predicted to be 55.21 kDa.

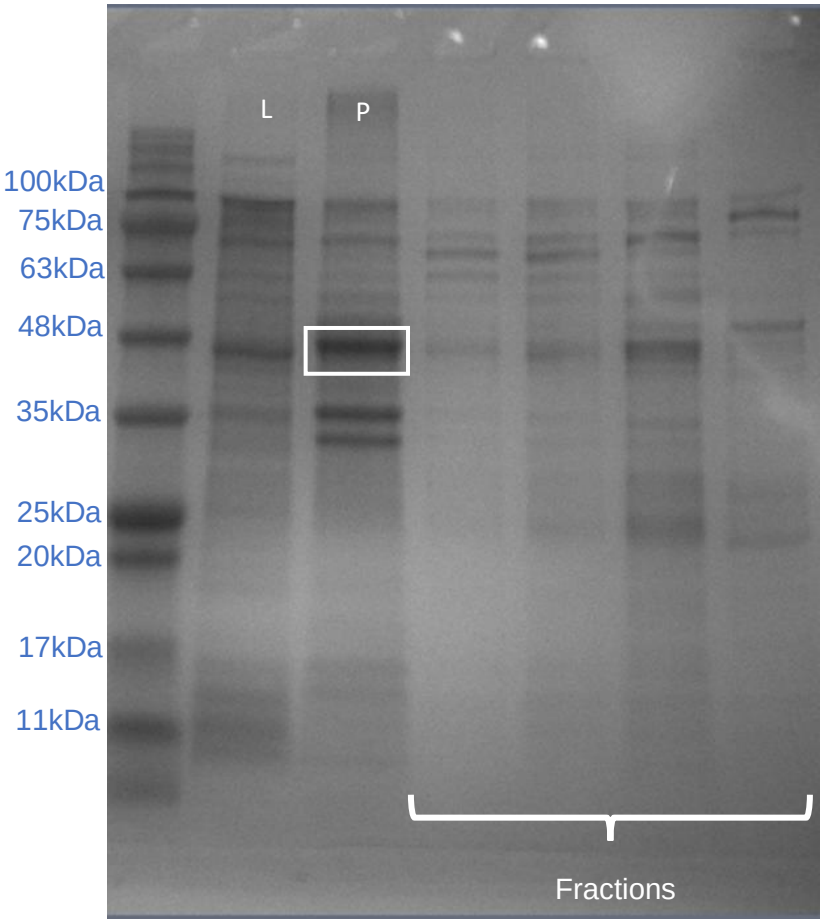


Figure 16. Attempted purification of 1222H from pET28b1222 by nickel affinity chromatography. The band highlighted by the white rectangle represents what is thought to be 1222H. **L** denotes lysate pre-column application; **P** denotes insoluble fraction separated by centrifugation. Protein bound to the column was eluted in a single peak on the UV chromatogram on an isocratic gradient of 500 mM imidazole. The protein ladder is the BLUelf Prestained Protein ladder (FroggaBio, Toronto, ON). 1222H is predicted to be 50.67 kDa.

## 3.5 Enzymatic assessment of ISGA 1218, 1222 and SsuD

SsuD is known to catalyze the conversion of octane sulfonate to octanal and sulfite and was used as a positive control for evaluating *in vitro* enzymatic desulfonation reactions by monitoring the production of octanal and sulfite by gas chromatography and $TNB^{2-}$ absorption, respectively. Considering 6:2 FTSA is a structurally similar compound to octane sulfonate, the enzymatic conditions developed to evaluate SsuD against octane sulfonate were extended to evaluate the *in vitro* desulfonation of 6:2 FTSA and HPLC was used to monitor 6:2 FTSA disappearance. Specifically, SsuDH, MBPSsuD, MBP1218 and MBP1222 purified here were challenged with 400 µM of octane sulfonate or 6:2 FTSA and MBP205, MBP1666 and MBP1835 were challenged with 400 µM of octane sulfonate.

### 3.5.1 Enzymatic assessment of ISGA 1218, 1222 and SsuD against octane sulfonate

In order to evaluate the *in vitro* desulfonation of SsuDH, MBPSsuD, MBP1218 and MBP1222 against octane sulfonate, GC-FID was employed to quantify octanal production, GC-MS to identify organic compounds in reaction extracts and $TNB^{2-}$ absorption to quantify sulfite production. When challenged with 400 µM of octane sulfonate, SsuDH produced 124.51 +/- 10.97 µM of octanal, and MBPSsuD produced 80.71 +/- 24.36 µM of octanal (see Figure S1-2 for calibration curves) accounting for 31.13 and 20.18 molar percent of added octane sulfonate. MBP1218 and MBP1222 reactions produced undetectable levels of octanal.

Peaks corresponding to octanol were observed in all GC-FID chromatograms with the exception of full SsuDH reaction. Octanol is not an expected product and therefore, GC-MS was employed to identify peaks in reaction extracts. Peaks found by GC-MS were searched against the National Institute of Standards and Technology library and in addition to the analytical standards, GC-MS identified peaks at 4, 12.8 and 15.6 minutes to be hexanoic, octanoic and decanoic acid respectively; all aldehyde analytical standards contained their carboxylic acid variant (Figure S5).

Octanoic acid was identified in all reaction extracts with the exception of SsuDH full reactions. The source of octanoic acid in reaction extracts is unknown and unlikely to be due to the oxidation of octanal due to the absence of octanal in reaction extracts containing octanoic acid. MBP205, 1666 and 1835 produced no quantifiable sulfite or octanal when challenged with 400 μM of octane sulfonate (Figure S10).

In order to quantify sulfite, enzymatic reactions were first extracted with ethyl acetate to precipitate the protein in solution. Protein precipitation was readily seen by the formation of a third phase between the aqueous and organic which consisted of the formerly aqueous protein. DTNB reacts with sulfite and thiols on cysteine or methionine residues, therefore, ethyl acetate extracted aqueous phases were analyzed to ensure that protein in solution were not confounding results. Protein, substrate and cofactors have absorbance readings below 0.05 when reacted with DTNB. When challenged with 400 μM of octane sulfonate and reacted with DTNB, SsuDH produced 51.56 +/- 7.49 μM of sulfite and MBPSsuD produced 34.52 +/- 1.28 μM of sulfite. Respectively these account for 12.89 and 8.63 molar percent of added octane sulfonate. These data are not in agreement with the octanal quantification above where the sulfite produced by SsuDH accounts for 41.41 molar percent of the detected octanal and the sulfite produced by MBPSsuD accounts for 42.77 molar percent of the detected octanal. MBP1218 produced 2.12 μM and MBP1222 produced 2.97 μM of sulfite under the reaction conditions used and account for less 1 molar percent of added octane sulfonate.

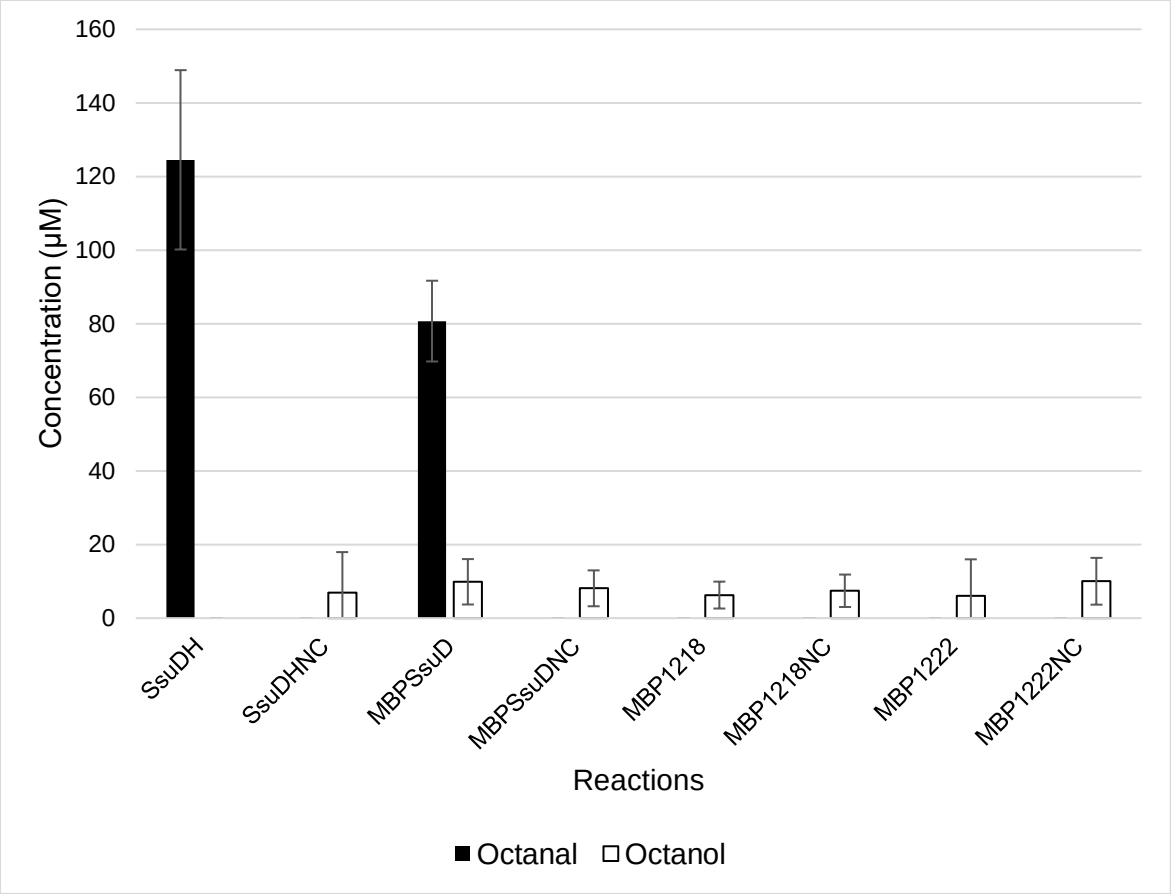


Figure 17. Concentration of octanal and octanol in reactions challenged with octane sulfonate. Enzyme names represent the catalyst used in the reaction and **NC** denotes no catalyst added. Ethyl acetate blanks were regularly run every six samples to identify contamination between samples. No hexanal, octanal, octanol or decanal was observed in blanks. Error bars represent standard deviation (n=3).

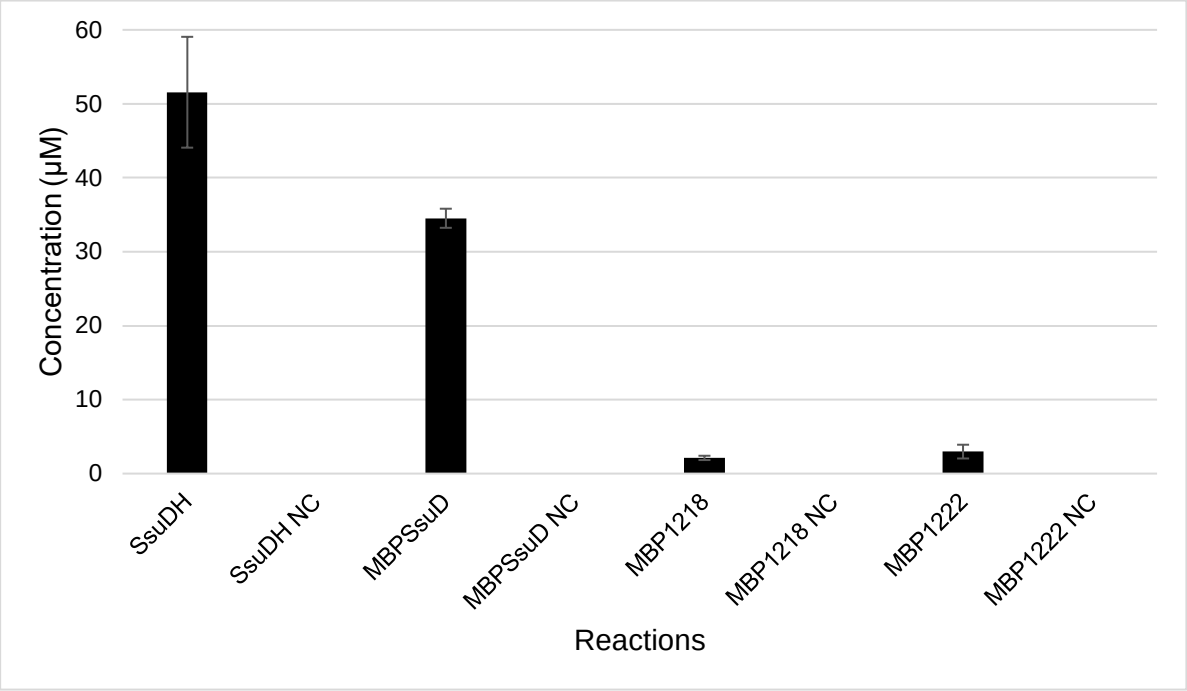


Figure 18. Concentration of sulfite in ethyl acetate extracted reactions challenged with octane sulfonate. Reactions were incubated with DTNB for 5 minutes at room temperature prior to measurement. Enzyme names represent the catalyst used in the reaction and **NC** denotes no catalyst added. Error bars represent standard deviations (n=3).

### 3.5.2 Enzymatic assessment of ISGA 1218, 1222 and SsuD against 6:2 FTSA

In order to evaluate the *in vitro* desulfonation of SsuDH, MBPSsud, MBP1218 and MBP1222 against 6:2 FTSA, HPLC was employed to quantify the depletion of 6:2 FTSA and $TNB^{2-}$ absorption to quantify sulfite production. When challenged against 400 µM 6:2 FTSA, no catalyst reactions did not return 100 molar percent of the added 6:2 FTSA. The molar recovery of no catalyst reactions averaged 49 molar percent of added 6:2 FTSA. As such, the difference between no catalyst and catalyst reactions were used to assess the loss of 6:2 FTSA. SsuDH full reaction decreased 103.08 +/- 37.93 µM, MBPSsuD full reaction decreased 130.28 +/- 91.30 µM, MBP1218 full reaction decreased 12.82 +/- 143.37 µM and MBP1222 full reaction increased 13.82 +/- 106.68 µM of 6:2 FTSA with respect to their no catalyst reactions. If only the decreases are considered and the difference equates to 6:2 FTSA conversion, 6:2 FTSA depletion by SsuDH accounts for 25.77, by MBPSsuD accounts for 32.57 and by MBP1218 accounts for 2.45 molar percent of added 6:2 FTSA.

Again, sulfite was quantified using DTNB on ethyl acetate extracted reactions challenged with 6:2 FTSA. SsuDH produced 36.24 +/- 5.21 µM and MBPSsuD produced 29.47 +/- 9.61 µM of sulfite. This represents a 9.06 and 7.36 molar percent of added 6:2 FTSA for SsuDH and MBPSsuD reactions respectively. MBP1218 produced 2.63 +/- 0.28 µM and MBP1222 produced 3.96 +/- 1.54 µM of sulfite which accounts for less than 1 molar percent of added 6:2 FTSA. Again, the sulfite quantification is not in agreement with the 6:2 FTSA depletion and sulfite quantification accounts for 35.15 and 22.62 molar percent of the depleted 6:2 FTSA for SsuDH and MBPSsuD reactions, respectively.

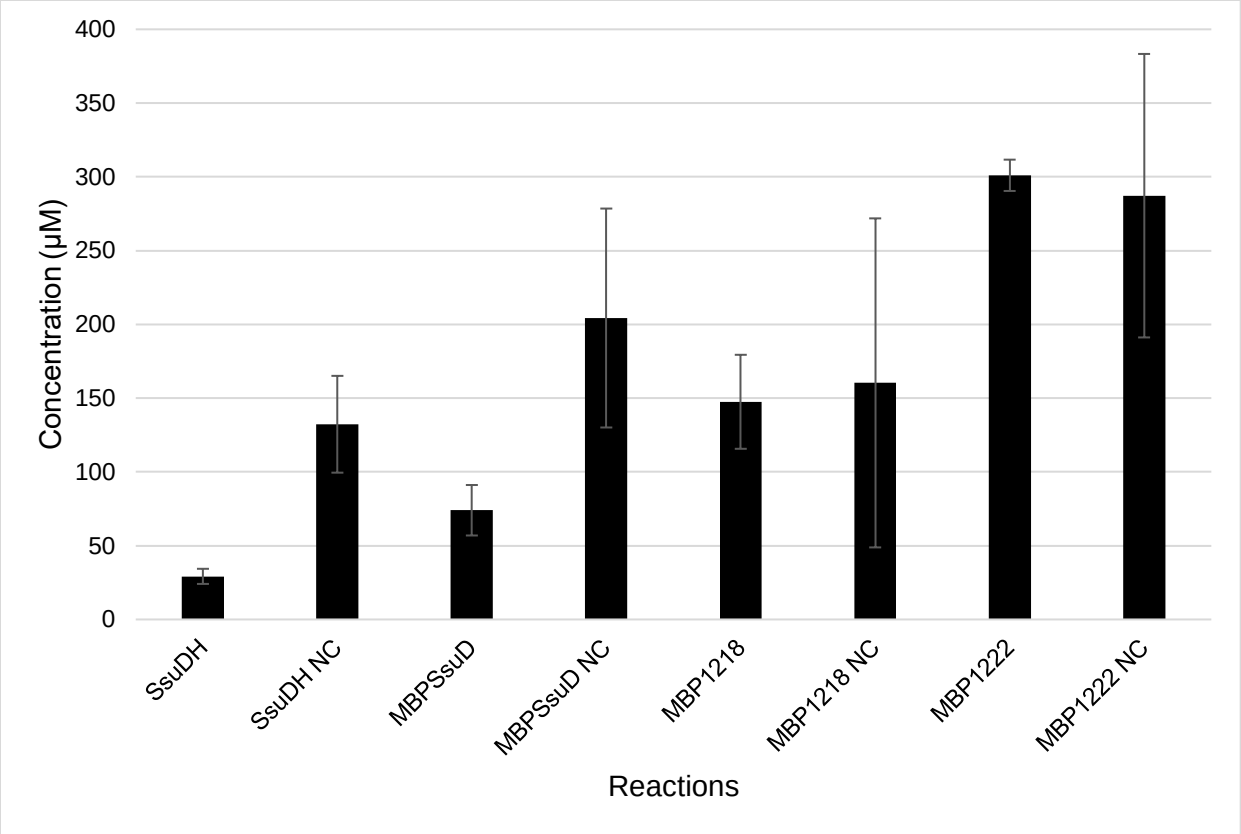


Figure 19. Concentration of 6:2 FTSA in reactions challenged with 6:2 FTSA. Enzyme names represent the catalyst used in the reaction and **NC** denotes no catalyst added. SsuDH and SsuDH NC were the only pair found to be statistically different. Error bars represent standard deviation (n=3).

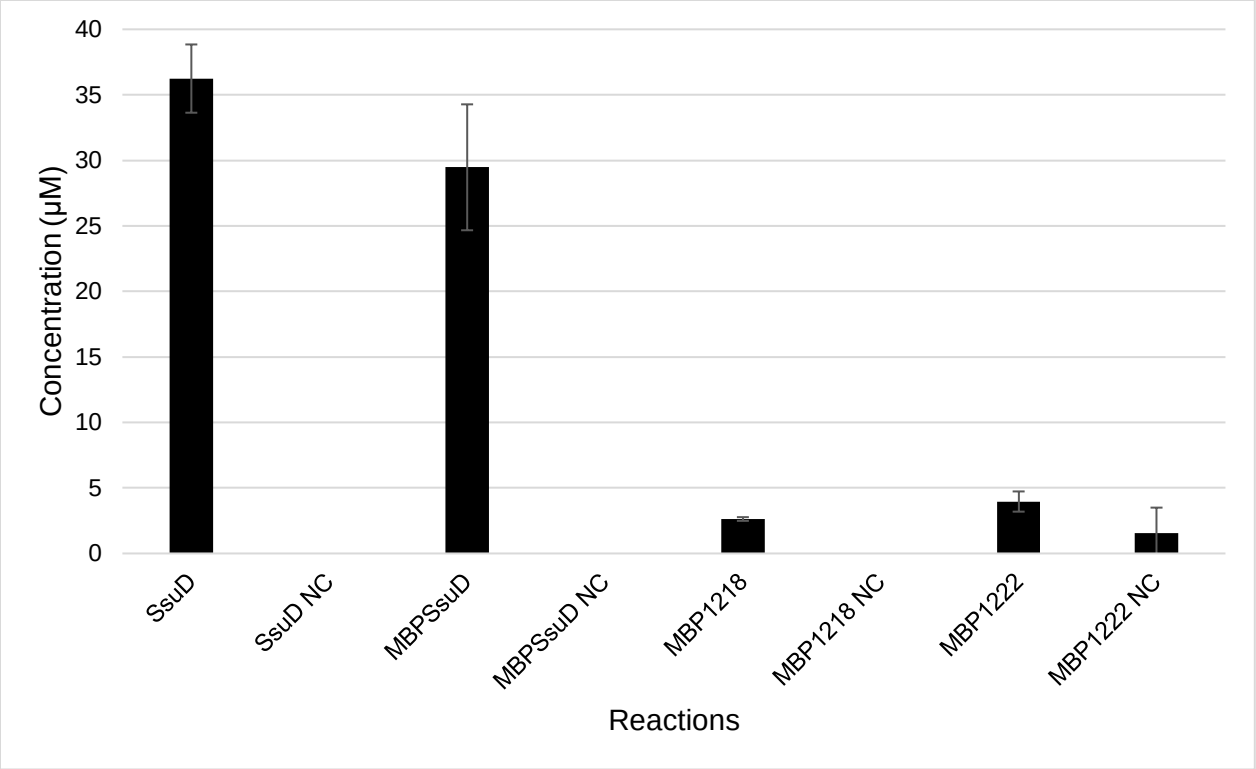


Figure 20. Concentration of sulfite in ethyl acetate extracted reactions challenged with 6:2 FTSA. Reactions were incubated with DTNB for 5 minutes at room temperature prior to measurement. Enzyme names represent the catalyst used in the reaction and **NC** denotes no catalyst added. Error bars represent standard deviations (n=3). MBP1222 and MBP1222NC are not statistically different.

### 3.5.3 Assessment of sulfite oxidation by Fre produced $FMNH_2$ or dissolved oxygen

Sulfite quantification consistently accounted for less than 43 molar percent of product formation or substrate depletion in all reaction conditions with all substrates. The oxidation of sulfite to sulfate in solution is thought to be a main contributor to the overall low molar percent accountability among all reactions. Sulfate does not react with DTNB and $FMNH_2$ produced by Fre could generate soluble c4a-peroxyflavin intermediates which could oxidize sulfite. In order to evaluate the potentially oxidizing activity of Fre, through $FMNH_2$, on sulfite, sodium sulfite was added to a final concentration of 400 µM to a no catalyst reaction with (Fre+) or without (Fre-) Fre and sulfite monitored with DTNB every 30 minutes for 2 hours. After 2 hours of incubation, the Fre+ reaction returned 210.97 +/- 15.41 µM of sulfite and the Fre- reaction returned 362.55 +/- 40.45 µM of sulfite. Respectively, the Fre+ reaction accounted for 52.74 molar percent and the Fre- 90.63 molar percent of the initially added sulfite. Approximately 47.26 and 9.37 molar percent of the initially added sulfite was undetectable following a two hour incubation in Fre+ or Fre- reactions, respectively.

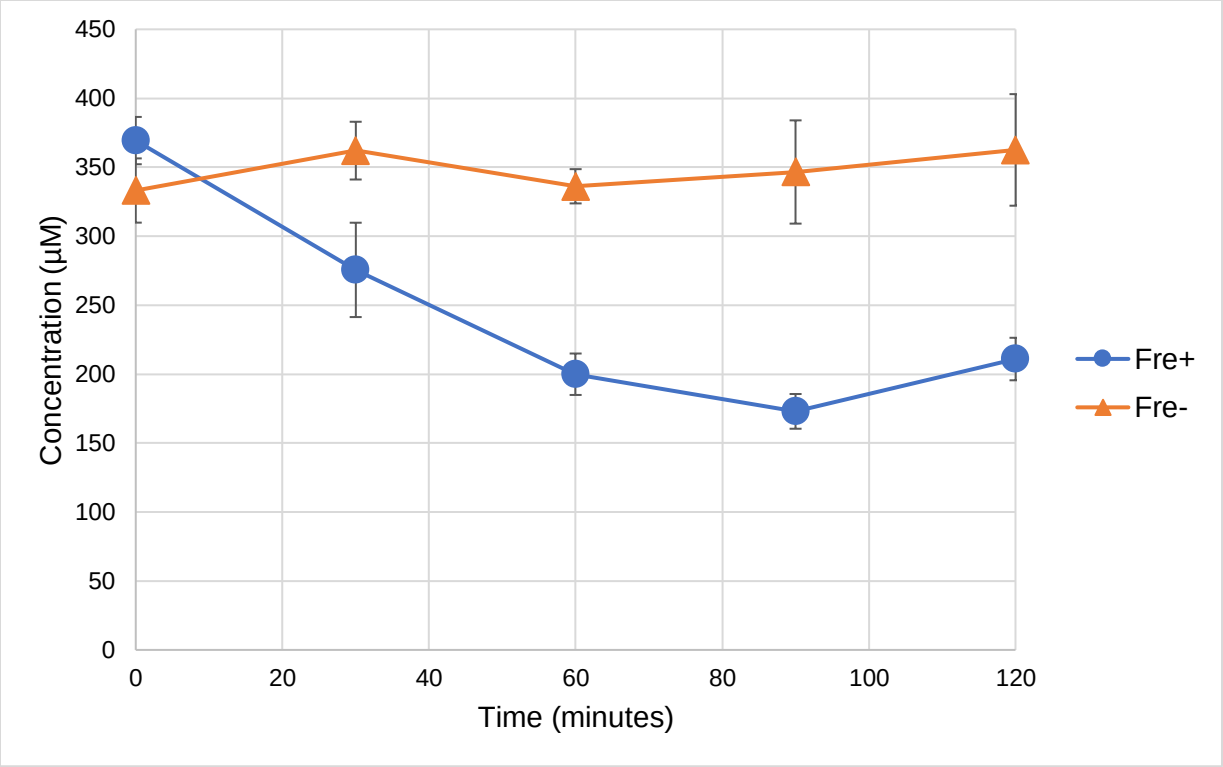


Figure 21. Concentration of sulfite in reactions with and without FreH. Fre+ indicates the presence and Fre- indicates the absence of FreH. Assays were carried out in triplicate and error bars represent standard deviations (n=3).

## 3.6 Kinetic assessment of octane sulfonate and 6:2 FTSA by SsuD under non-coupled $FMNH_2$ generating conditions

SsuE is the reductase of the *ssu* operon in *E. coli* and couples the oxidation of NADH to the reduction of FMN and supplies $FMNH_2$ to SsuD through protein-protein interaction. The genomic context of the genes encoding ISGA 1218 and 1222 do not encode an SsuE-like reductase and as such, in order to evaluate the kinetic properties SsuD against octane sulfonate and 6:2 FTSA, Fre was used to generate $FMNH_2$. The hypothesis here is Fre could produce soluble $FMNH_2$ in excess for SsuD and kinetic parameters for SsuD against octane sulfonate and 6:2 FTSA could be evaluated in a manner resembling *Gordonia* NB4-1Y *in vivo*. Here, purified SsuDH and FreH were used.

In order to produce a steady-state environment for SsuDH, the same reaction conditions used to enzymatically assess MBP1218, 1222 and SsuDH were used with some modifications. SsuDH concentrations were lowered from 1 µM to 0.2 µM and FreH concentrations were increased from 0.2 to 0.8 µM resembling the conditions used by Zhan *et al.* (2008). Octane sulfonate and 6:2 FTSA were tested at concentrations between 25 µM and 300 µM prepared in 50% vol/vol ethanol and in a parallel experiment, 200 µM PFOS used as potential competitive inhibitor. Kinetic parameters of interest are described in Table 8 and calculated following the analytical procedure of Zhan *et al.* (2008); Lineweaver-Burk plots are given in Figure 21. In brief, the $K_m$ of octane sulfonate for SsuDH is 1.24 times higher in the presence of 200 µM of PFOS; $V_{max}$ for octane sulfonate was 1.45 times lower in the presence of PFOS. The $K_m$ of 6:2 FTSA is 2.88 times higher in the presence of 200 µM of PFOS; $V_{max}$ values were similar in the presence or absence of PFOS.

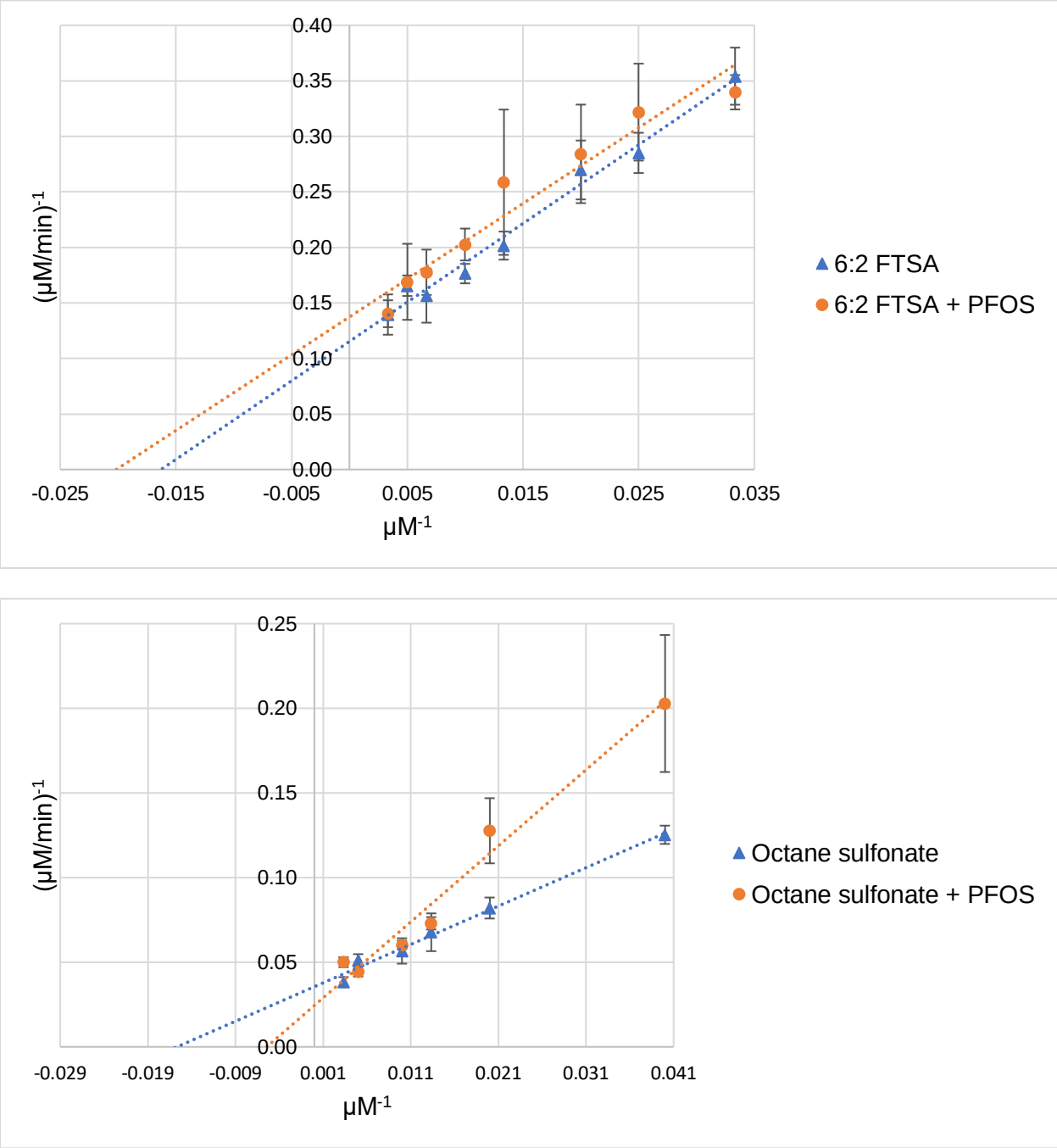


Figure 22. Lineweaver-Burk double reciprocal plot of SsuDH challenged with 6:2 FTSA (top) or octane sulfonate (bottom) in the presence (circle) or absence (triangle) of 200 µM of PFOS. SsuDH was provided $FMNH_2$ by FreH. Error bars represent standard deviations (n=3).

Table 8. Kinetic parameters for octane sulfonate and 6:2 FTSA conversion to octanal, an unidentified fluorotelomer and sulfite.

| Substrate | $K_m$ (μM) | $V_{max}$ (μM/min) | $k_{cat}$ ($min^{-1}$) |
|---|---|---|---|
| Octane Sulfonate | 63.87 | 28.17 | 140.85 |
| Octane Sulfonate + 200 μM PFOS | 184.05 | 40.98 | 204.92 |
| 6:2 FTSA | 61.42 | 8.67 | 43.36 |
| 6:2 FTSA + 200 μM PFOS | 49.64 | 7.28 | 36.42 |
| Octane Sulfonate[1] | 44 | 1.6[2] | 268.4[2] |

[1]Kinetic parameters reported by Eichhorn *et al.*, 1999 and Zhan *et al.*, 2008.

[2]Vmax reported as units/mg by Eichhorn *et al.,* 1999.

## 3.7 *E. coli* BL21(DE3) growth assays in no sulfur added mineral media supplemented with $MgSO_4$, octane sulfonate or 6:2 FTSA.

*E. coli* BL21(DE3) harbors the well described *ssu* operon which imports and partially degrades aliphatic sulfonates such as octane sulfonate. Growth assays have been reported by Eichhorn *et al.* (2000) using the M63 growth medium with sulfur limited to added aliphatic sulfur sources. *E. coli* growth is inhibited by the ethanol (Basu *et al.* 1994) and therefore, octane sulfonate and 6:2 FTSA were prepared in 0.2 µm filter sterilized water. In order to assess the ability of *E. coli* BL21(DE3) to grow when sulfur is limited to 6:2 FTSA, two approaches were taken.

In the first approach, in a 96-well plate, washed *E. coli* BL21(DE3) was added to M9 minimal medium supplemented with 200 or 400 µM of $MgSO_4$, octane sulfonate or 6:2 FTSA. Under oxygen restrictive conditions, $OD_{621}$ readings inconsistently changed depending on the position of the well in the 96-well plate. Growth curves were unproducible under oxygen permissive conditions on a 96-well plate. Due to the low volume, the wells dried after approximately 14 hours of incubation.

In the second approach, and in parallel with the first, *E. coli* BL21(DE3) biomass yield when sulfur was limited to 400 µM of $MgSO_4$, octane sulfonate or 6:2 FTSA was followed by measuring $OD_{660}$ at 24 and 48 hours. Here, washed *E. coli* BL21(DE3) grew to $OD_{660}$ of 0.59 +/- 0.24 when limited to octane sulfonate and 0.32 +/- 0.06 when limited to 6:2 FTSA under oxygen permissive conditions after 48 hours of incubation. Together, these equate to 3.86 and 2.12 times the $OD_{660}$ when no sulfur was added. Under oxygen restrictive conditions, *E. coli* BL21(DE3) grew to $OD_{660}$ 0.178 +/- 0.016 and 0.18 +/- 0.022 when limited to octane sulfonate and 6:2 FTSA, respectively which equates to 1.62 and 1.69 times the $OD_{660}$ when no sulfur was added.

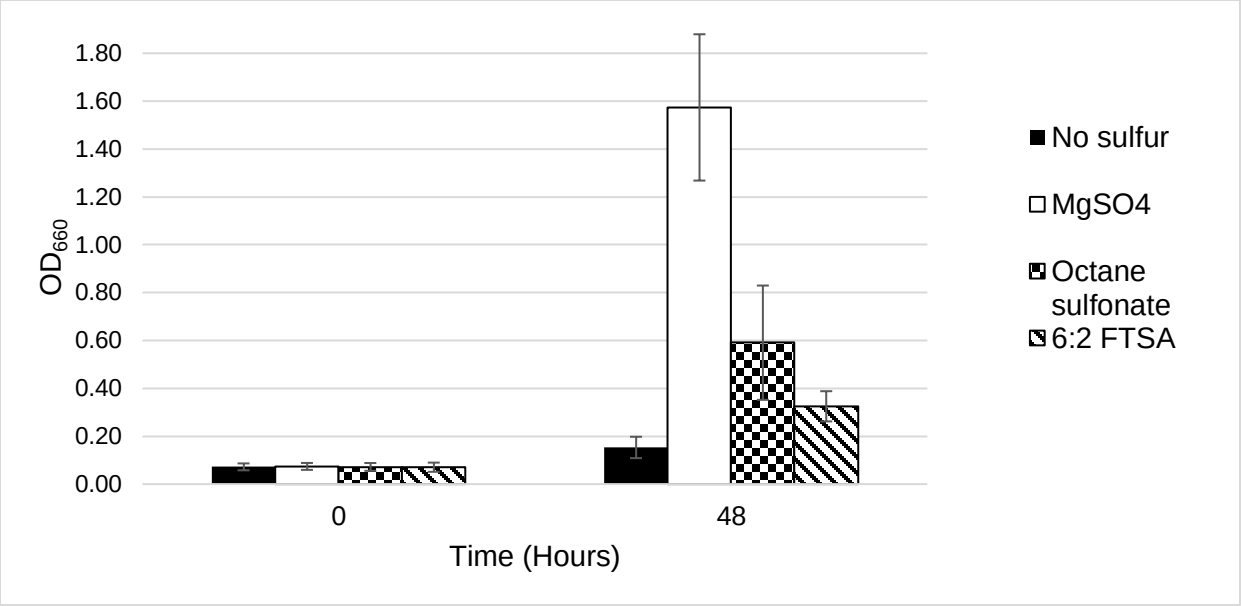


Figure 23. *E. coli* BL21(DE3) biomass yield under sulfur limited to no sulfur, $MgSO_4$, octane sulfonate and 6:2 FTSA in oxygen permissive conditions. Error bars represent standard deviation (n=6). At 48 hours, both 6:2 FTSA and octane sulfonate were statistically different from the No sulfur treatment.

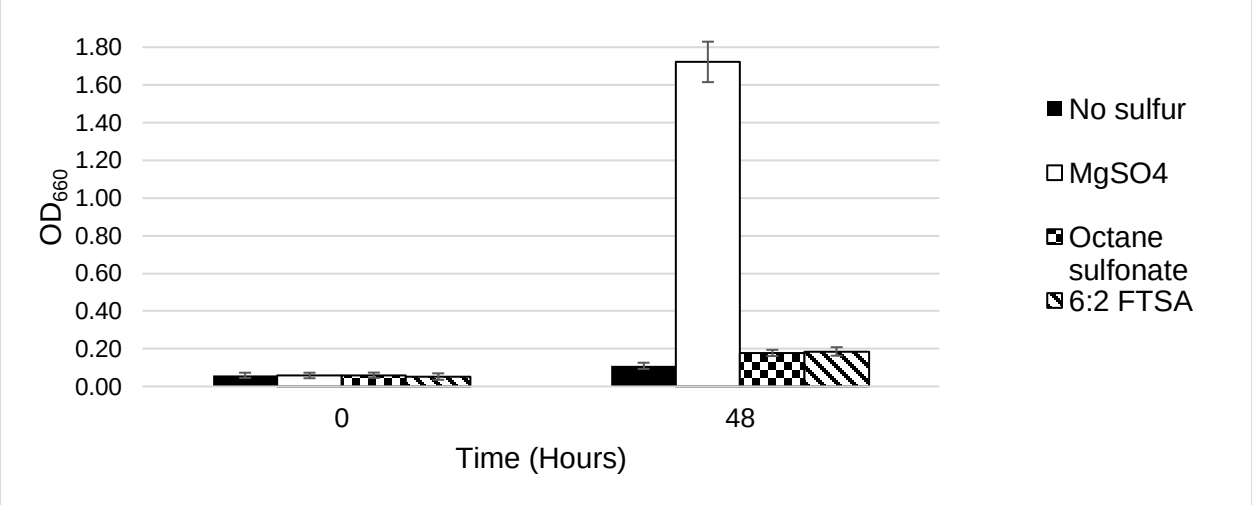


Figure 24. *E. coli* BL21(DE3) biomass yield under sulfur limited to no sulfur, $MgSO_4$, octane sulfonate and 6:2 FTSA in oxygen restrictive conditions. Error bars represent standard deviation (n=6). At 48 hours, both 6:2 FTSA and octane sulfonate were not statistically different with respect to one another, however, both were statistically different with respect to the No sulfur control.

## 3.8 Conjugation and transformation of *Gordonia* NB4-1Y with pK18mobsacB1218AB and pK18mobsacB1222AB

Gene knockout of the genes encoding ISGA 1218 and 1222 would inform, at the genomic level, what role these genes play in 6:2 FTSA degradation. In order to knockout the genes encoding ISGA 1218 and 1222 in the *Gordonia* NB4-1Y genome, knockout vector transfer was attempted by conjugation and electroporation. In the former, after 8 days of conjugation with *E. coli* S17.1 carrying pK18mobsacB1218AB or pK18mobsacB1222AB, no *Gordonia* NB4-1Y single recombinants were found growing on M9 minimal medium agar with glucose and kanamycin. When pK18mobsacB1218AB was electroporated into *Gordonia* NB4-1Y at 1.8, 2.2 and 2.5 kV and grown for 7 days, isolated, opaque, *E. coli*-like colonies were seen, however, colony PCR with FreF (3) and FreR -s (3) returned no product (Figure 24). Two candidate colonies were streak purified, grown in LB or NB with kanamycin and their plasmids extracted. Plasmid extracts separated on agarose gel revealed that unknowns 17 and 18 contained extractable plasmids suggesting non-genome integrated plasmid (Figure 25).

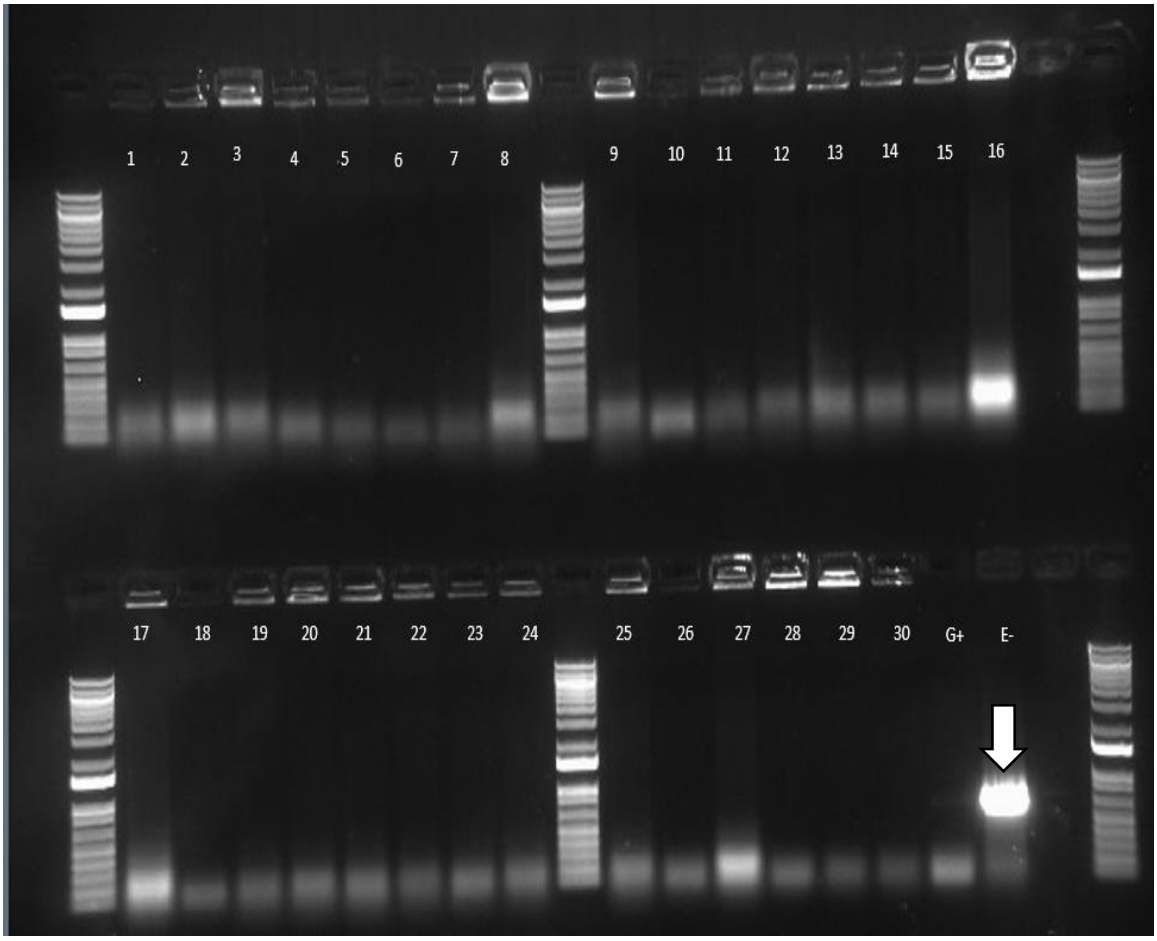


Figure 25. Colony PCR with FreF(3) and FreR-s (3) of candidate *Gordonia* NB4-1Y single recombinants transformed with pK18mobsacB1218AB. Templates for PCR are as follows: candidate colonies (**1-27**); wild-type *Gordonia* NB4-1Y (**28-30**); *Gordonia* NB4-1Y genomic DNA (**G+**) and *E. coli* BL21(DE3) genomic DNA (**E-**). The white arrow denotes *fre*, amplified from *E. coli* genomic DNA.

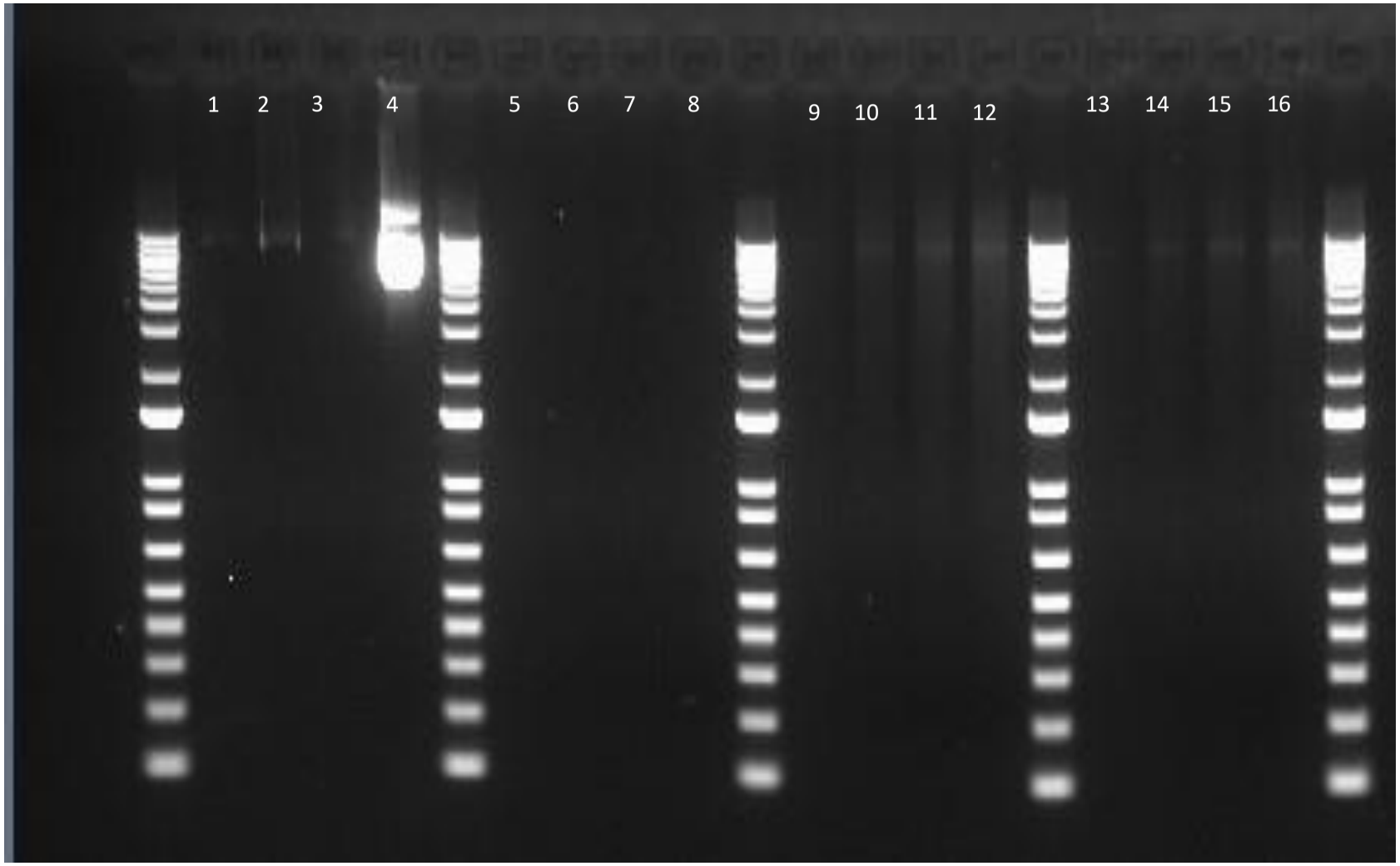


Figure 26. Plasmid extraction of *E. coli* S17.1 carrying pK18mobsacB1218AB (**1**-**4**), wild-type *Gordonia* NB4-1Y (**5**-**8**), unknown 17 (**9**-**12**) and unknown 18 (**13**-**16**). Turbid cultures had their plasmid extracted and diluted 1:10; 1 µL (**1**, **5**, **9**, **13**), 2 µL (**2**, **6**, **10**, **14**), 3 µL (**3**, **7**, **11**, **15**) and 5 µL (**4**, **8**, **12**, **16**) were separated.

# 4.0 Discussion

## 4.1 Recap of current literature

The *Gordonia* genus is metabolically diverse and sister to the *Rhodococcus* and *Mycobacterium* genera, which contain species capable of application as both powerful bioremediation and biotechnological tools as well as deadly pathogens (Arenskötter *et al.*, 2004). Species of *Gordonia* have been known to degrade both aliphatic (Van Hamme *et al.*, 2013) and heterocyclic (Kim *et al.*, 1999 and 2000) sulfur-containing compounds and, as of 2013, *Gordonia* NB4-1Y is known to metabolize 6:2 FTSA. Fluorotelomer sulfonates and sulfonamides are common degradation intermediates in the breakdown of complex fluorotelomers by aerobic microbial communities (Harding-Marjanovic *et al.*, 2015 and D'Agostino and Mabury, 2017) and the rate limiting step of their degradation is the oxygen-dependent, carbon-sulfur bond cleavage (Zhang *et al.*, 2016).

Sulfur acquisition from aliphatic sulfonate sources such as octane sulfonate or taurine has been well studied in *E. coli*. The alkanesulfonate monooxygenase (SsuD) is responsible for the carbon-sulfur bond cleavage in aliphatic sulfonates and the taurine dioxygenase (TauD) can convert taurine with 2-oxoglutarate to aminoacetaldehyde and sulfite (Eichhorn *et al.*, 1997 and 1999). Prokaryotic two-component monooxygenases co-ordinate $FMNH_2$ with $O_2$ in their active site to produce a c4a-peroxyflavine intermediate. This intermediate undergoes a Baeyer-Villiger rearrangement with an aliphatic sulfonate releasing sulfite and an aldehyde (Ellis, 2010).

Although several studies have been conducted on the degradation of PFAS in mixed and pure culture experiments (as reviewed by Liu and Avendaño, 2013), no study to date has demonstrated the enzymatic transformation, by bacterial enzyme, of fluorinated molecules with more than 3 fluorines (Murphy *et al.*, 2009). As such, there is a gap in our understanding of the exact mechanism of PFAS degradation within a bacterial cell. While this gap persists, bona fide assignment of PFAS with more than 3 fluorines, such as 6:2 FTSA, to known microbial

degradation pathways and effective biological-based remedies for PFAS contamination are held back.

The goal of this study was to produce the first evidence of PFAS degradation by bacterial enzyme. Here, the *Gordonia* NB4-1Y nitrilotriacetate monooxygenases ISGA 1218 and 1222 and the *E. coli* BL21(DE3) alkanesulfonate monooxygenase (SsuD) were purified, given soluble $FMNH_2$ by the *E. coli* NADH:FMN oxidoreductase (Fre) and challenged *in vitro* with 6:2 FTSA or octane sulfonate. Reaction products were determined by GC and spectrophotometric absorbance and substrate disappearance monitored by HPLC. Identification of the first step in the biological transformation of 6:2 FTSA would expand our effective knowledge of the fate of fluorinated pollutants in the environment. These findings could help develop bioremediation solutions to PFAS contamination and inform if an environmental matrix can effectively transform 6:2 FTSA.

## 4.2 Significance

### 4.2.1 ISGA 1218 and 1222

ISGA 1218 and 1222 were characterized, by DNA sequence, as nitrilotriacetate monooxygenases (NtaA) (Van Hamme *et al.,* 2013) and are a part of the class C of flavin-dependent prokaryotic two-component monooxygenases (van Berkel *et al.*, 2006). However, upon re-submission of a refined *Gordonia* NB4-1Y genome in 2015, ISGA 1218 and 1222 were re-annotated as LLM-class flavin-dependent monooxygenases. These monooxygenases use soluble or reductase provided $FMNH_2$ and $O_2$ to oxygenate a variety of cyclic and aliphatic compounds. Consequently, some monooxygenases in this class can cleave carbon-sulfur bonds via a proposed Bayer-Villiger rearrangement (Ellis, 2010 and Dayal *et al*, 2015).

When ISGA 1218 and 1222 were given soluble $FMNH_2$ by FreH and challenged with octane sulfonate and 6:2 FTSA, sulfite production accounted for less than 1 molar percent of added substrate. Furthermore, depletion of 6:2 FTSA was not found to be statistically different from the no catalyst added reactions (p=0.850-0.883). No octanal was quantified by either enzyme and no statistical difference in sulfite concentrations was found between 6:2 FTSA challenged MBP1222 based reactions and their no catalyst reaction (p=0.325). Therefore, sulfite produced by MBP1218 or 1222 challenged with 6:2 FTSA or octane sulfonate is likely insignificant.

Van Hamme *et al.* (2013) hypothesized that ISGA 1218 and 1222 mediated the degradation of 6:2 FTSA due to their differential production in conditions where sulfur is limited to 6:2 FTSA, their low overall sulfur containing amino acid content and close alignment to the alkanesulfonate monooxygenases and taurine dioxygenases. Differential production of proteins is indicative of a potential function, however, can sometimes be misleading. Under sulfur-limiting conditions for example, carbon-sulfur bond breaking genes are not uniquely upregulated in *Pseudomonas* (Tralau *et al.*, 2007 and Scott *et al.*, 2006). Knobel *et al.* (1996) reported a 40.1% amino acid

sequence similarity between the NtaA of *Aminobacter aminovorans* and the DszA of *Rhodococcus erythropolis* and suggested NtaA and DszA may share a common ancestor. If that is the case, it is possible that NtaA production is under a similar genetic regulation as sulfur assimilating enzymes and ISGA 1218 and 1222 were produced as a by-product of the sulfur starvation response in *Gordonia* NB4-1Y.

A perhaps misleading inference by Van Hamme *et al.* (2013) could be the conclusions based on the phylogenetic placement of the genes encoding ISGA 1218 and 1222 between *ssuD*, *ssuD*-like genes and *tauD*. NtaA belongs to the Class C of flavin-dependent monooxygenases (van Berkel, 2006 and Huijbers *et al*. 2014) along with enzymes such as SsuD, dibenzothiophene monooxygenase (DszA), alkanal monooxygenase (luciferase), long-chain alkane monooxygenase (LadA) and diketocamphane monooxygenase (Huijbers *et al.* 2014). These enzymes are marked by their TIM-barrel protein fold and use of $FMNH_2$ and $O_2$ as substrate (Huijbers *et al.* 2014). TauD, on the other hand, belongs to the Group II of α-ketoglutarate and Fe(II) dependent dioxygenases more closely related to 2, 4-dichlorophenoxyacetic acid dioxygenase and clavaminate synthase (Hogan *et al.*, 2000). These enzymes are marked by their use of Fe(II), α-ketoglutarate and $O_2$ as substrate (Eichhorn *et al.* 1997). Being both oxygenolytic enzymes, SsuD and TauD potentially share a common ancestor, however, their associated transporter systems are not hybridizable (Eichhorn *et al.* 2000). This suggests that Group II dioxygenases and Class C monooxygenases independently acquired desulfonation capabilities and that the phylogenetic alignment of the genes encoding ISGA 1218 and 1222 were adjacent to *ssuD* and other desulfinases with *tauD* acting as a more distant outgroup.

Upon further investigation of the phylogenetic placement of ISGA 1218 and 1222 among Class C monooxygenases, it was found that ISGA 1218 was aligned closest to DszA and ISGA 1222 with EmoA in a clade comprising of DszA, SnaA, NtaA and EmoA. This is not surprising, as mentioned earlier, Knobel *et al*. (1996) found a 40.1% amino acid sequence similarity between NtaA and

DszA whereas Xu *et al*. (1997) found a 49.2% amino acid sequence similarity with SnaA. Given the initial annotation of ISGA 1218 and 1222 as NtaAs, it could be expected that either enzyme would closely align with one of the enzymes within that grouping. In conjunction with the re-annotation of ISGA 1218 and 1222, these findings further suggest that ISGA 1218 and 1222 may not be NtaA, however, do partially explain the upregulation of ISGA 1218. In *C. glutamicum* ATCC 13032, *ssu* and *seu* genes are both upregulated under sulfur starvation conditions (Koch *et al*. 2005). Genes of the *seu* operon are analogues to *dsz* genes, liberating sulfur from sulfate esters (Koch *et al*. 2005). Should the true identity of ISGA 1218 be a DszA-like enzyme, then it can be expected that when sulfur is limited, this enzyme will be upregulated.

### 4.2.2 Alkanesulfonate monooxygenase

SsuD is known to convert octane sulfonate to octanal and sulfite if soluble $FMNH_2$ is present or provided by SsuE (Eichhorn *et al.* 1999 and Dayal *et al.* 2015). When SsuD was given soluble $FMNH_2$ by Fre and challenged with octane sulfonate or 6:2 FTSA, 6:2 FTSA disappearance and octanal and sulfite production accounted for up to 34 molar percent of added octane sulfonate or 6:2 FTSA. Octanal production was expected and MBPSsuD and SsuDH produced statistically significant concentrations of octanal and sulfite when challenged with octane sulfonate. Furthermore, MBPSsuD and SsuDH produced statistically significant concentrations of sulfite when challenged with 6:2 FTSA. Depletion of 6:2 FTSA was, however, statistically insignificant in some cases. Depletion of 6:2 FTSA by MBPSsuD was not found to be statistically different from its no catalyst control ($p=0.073$), however, depletion of 6:2 FTSA by SsuDH was ($p=0.012$). The reason for the statistical insignificance between MBPSsuD and its no catalyst reaction is due to the large standard deviation associated with the no catalyst reaction. The standard deviation for no catalyst reactions accounted for 24.30-36.32% of their calculated mean. The source of this large standard deviation is unknown, however, potential sources of error could include improper sealing of the inverted septa, absorption of 6:2 FTSA to glass reaction vials or microcentrifuge

tubes, absorption to protein or inaccurate preparation of 6:2 FTSA stock (3 mM 6:2 FTSA is 66.9 mg in 50 mL). Repeating the enzymatic assessment of 6:2 FTSA degradation with MBPSsuD and SsuDH with more than 3 biological replicates will provide further statistical strength to the claim that SsuD can degrade 6:2 FTSA.

When SsuD was first described by Eichhorn *et al.* (1999), the authors found that SsuD could metabolize a variety of substituted and unsubstituted aliphatic sulfonates so long as there was an unsubstituted alpha and beta carbon; none had carbon-fluorine bonds. Understanding the substrate specificity of SsuD is difficult due to the lack of substrate-bound crystal structures; to date, only 2 SsuD crystal structures have been reported, all without substrate. The catalytic mechanism, however, has been extensively studied by a combination of biochemical assessments and molecular modeling studies (Armacost *et al.*, 2016 and Ferrario *et al.*, 2012). In brief, the proposed mechanism is as follows: a c4a-peroxyflavin ($FMNOO^-$) nucleophilically attacks the sulfur center of the aliphatic sulfonate generating a peroxyflavin sulfonate adduct which undergoes a sulfite releasing (Bayer-Villiger) rearrangement to produce a peroxyflavin aldehyde adduct. Abstraction of a proton from the alpha carbon of the aldehyde adduct releases the aldehyde product and donation of a proton to the hemi-peroxyflavin ($FMNO^-$) regenerates flavin. No 3D conformation of SsuD has been proposed to anchor aliphatic sulfonates, however, arginine 226 (Arg226) has been found to closely interact with the peroxide group of $FMNOO^-$ when octane sulfonate is bound (Armacost *et al.*, 2016). Armacost *et al.* (2016) hypothesized that that Arg226 protects the $FMNOO^-$ from bulk solvent, however, when $FMNH_2$ was bound to the SsuD, the $NH_2$ group of Arg226 closely interacts with octane sulfonate. This suggests that Arg226 might play a role in anchoring aliphatic sulfonates through interaction with the sulfite group prior to $FMNH_2$ reacting with $O_2$. Together, it can be hypothesized that the important factors in fluorotelomer sulfonate binding and conversion are the electronic properties of the alpha carbon and sulfonate groups of fluorotelomer sulfonates.

### 4.2.3 *Escherichia coli* growth on 6:2 FTSA

Following the results of the SsuD characterization experiments, it was hypothesized that *E. coli* BL21(DE3) could grow on 6:2 FTSA as a sole sulfur source. Two conditions were used to test if an oxygen dependent mechanism allows *E. coli* BL21(DE3) to grow on 6:2 FTSA: an oxygen permissive or restrictive environment. In an oxygen permissive environment, *E. coli* BL21(DE3) grew to $OD_{660}$ of 0.32 when limited to 6:2 FTSA and 0.59 when limited to octane sulfonate; these readings were both found to be statistically different from the no sulfur treatment (p=0.01-0.001). Further, under oxygen restrictive conditions, growth conditions with sulfur limited to 6:2 FTSA and octane sulfonate were marginally, yet significantly (p=<0.05), different over the no sulfur growth condition and is likely due to the limited oxygen already present in the test tubes used. These findings are indicative that an oxygen-dependent mechanism is responsible for 6:2 FTSA conversion in *E. coli* BL21(DE3). With the biochemical data produced, SsuD may be the catalyst involved in that mechanism. Alternatively, however, it is possible that the stock of 6:2 FTSA was contaminated with a potential alternate sulfur source; in previous experiments, stocks of PFOS were found to contain alternate sulfur sources (Van Hamme and Bottos personal communication, 2020) leading to false positive growth assays. This source would have to be an aliphatic sulfonate to fit the observed production of sulfite by SsuD, the similar $K_m$ values obtained during the kinetic assessment and the apparent oxygen dependent mechanism of *E. coli* growth.

Activated aerobic sludge is typically dominated by gammaproteobacterial (Zhang *et al.*, 2015), of which *E. coli* is a part, and poorly degrades 6:2 FTSA (Wang *et al.*, 2011). Zhang *et al.* (2016), suggested the poor degradation of 6:2 FTSA may be due to low “monooxygenases levels” in aerobic activated sludge compared to aerobic sediment. Whether Zhang *et al*. (2016) refer to monooxygenase levels as the number of monooxygenase gene copies in the microbial consortium or the overall production of monooxygenases is unclear. Zhang *et al*. (2016) based their hypothesis on the similar levels of sulfate in both aerobic sediment and activated sludge

(Tchobanoglous and Burton, 1991), however, reported no meta-genomic data (Zhang *et al.*, 2016). In light of this study, perhaps a lack of monooxygenase production is a more meaningful explanation over an entire lack thereof.

Two sulfur metabolism archetypes have been well described in the literature, one for *E. coli*, which seems to serve as a model for Gram-negative bacteria (van der Ploeg *et al*. 2001) and one for *Corynebacterium glutamicum*, which seems to serve as a model from Gram-positive bacteria (Rey *et al*, 2005).

In the first, expression of cysteine biosynthesis and sulfur acquisition genes is globally regulated by CysB (Kreidich, 2008) and its co-inducer, *N*-acetylserine in *E. coli* (Ianicka-Nowicka and Hryniewicz, 1995). *N*-acetylserine production is inhibited by the presence of cysteine (Kreidich, 2008) and therefore, when cysteine levels are low, *E. coli* will respond by expressing cysteine biosynthesis genes and producing Cbl (CysB-like) (van der Ploeg *et al*, 1997 and 1999). In turn, Cbl will activate genes involved in the acquisition of sulfur from alternative sources such as the *ssuD* and *tauD* (van der Ploeg *et al*, 1997 and 1999). Cbl is unable to activate genes involved in the acquisition of sulfur from alternative sources when adenosine 5'-phosphosulfate (APS) is present, an intermediate in inorganic sulfur assimilation (Stec *et al*. 2006, Bykowski *et al*. 2002 and Mueller and Shafqat, 2013). With this archetype, *E. coli* downregulates genes involved in sulfur acquisition in presence of cysteine and inorganic sulfur. Therefore, the *ssu* operon is not activated when cysteine is provided as a sulfur source (van der Ploeg *et al*. 1999).

In the second, the McbR (methioneine and cysteine biosynthesis) regulatory protein primarily governs the McbR regulon in *C. glutamicum* by maintaining repression of cysteine biosynthesis genes and *cysR* (Rey *et al*. 2005). The McbR regulon is released from repression in the presence of *S*-adenosylhomosyteine (SAH) (Rey *et al*. 2005). SAH is thought to be involved in the synthesis of nascent DNA (Thomas and Surdin-Kerjan, 1997) and its levels allow *C. glutamicum* to sense

the growth stage of the cell; when SAH levels are high, McbR repression is released and the McbR regulon expressed (Koch *et al*. 2005 and Rey *et al*. 2005). CysR in turn activates expression of SsuR (Ruckter *et al*. 2008) which will activate the *ssu* and *seu* operons (Koch *et al*. 2005). SsuR is unable to activate its target genes in the presence of APS and sulfate (Koch *et al*. 2005). With this archetype, *C. glutamicum* only regulates the expression of the *ssu* operon in the presence of sulfate. As a consequence, the *ssu* operon is expressed when cysteine is given as the sole source of sulfur (Koch *et al*. 2005).

It is unclear if Gram-negative and -positive bacteria globally share the sulfur acquisition archetypes described for *E. coli* and *C. glutamicum*, respectively. However, some Gram-negative bacteria do share similar archetypes as that described in *E. coli*. *Salmonella typhimuirum* and *Burkholderia cenocepacia* have CysB and Cbl equivalents (Kredich, 1996 and Iwanicka-Nowicka, 2007). In addition, a recent study found a TetR family transcription factor to activate expression of *dsz* genes in *Gordonia* sp. IITR100 (Murarka *et al*. 2019). Although this classification is not analogous to the SsuR of *C. glutamicum* (Koch *et al*. 2005), the sulfur acquisition archetype of *Gordonia* has not been entirely elucidated. It is possible another transcription factor in *Gordonia* fills the role of SsuR in the *C. glutamicum* archetype or, the identified transcription factor acts as a functional analog.

If Gram-negative bacteria do share similar sulfur acquisition archetypes, it can be reasoned that the poor transformation of 6:2 FTSA by activated sludge is due to the poor expression of *ssu* genes. Degradation of proteins in activated sludge wastewater treatment occurs (Westgate and Park, 2010) and could act as a significant source of cysteine to microbial communities. If that is the case, cysteine would further repress the sulfur assimilation pathway in activated sludge microbial communities compared to aerobic sediment.

### 4.2.4 *Gordonia* NB4-1Y genomic DNA search and phylogenetic assessement

On the genomic level, identifying the monooxygenase responsible for octane sulfonate desulfonation in *Gordonia* NB4-1Y is paramount as it is also likely responsible for the desulfonation of 6:2 FTSA. Several alkanesulfonate monooxygenases (ISGA 205, 1666 and 1835) have been annotated in the genome and their cloning attempted by McAmmond (2017). Of these annotated monooxygenases, none demonstrated desulfonation activity against octane sulfonate or 6:2 FTSA when produced with an MBP tag. The identity of these genes as SsuDs is interesting since the protein BLAST search against the *Gordonia* NB4-1Y genome with the *E. coli* SsuD revealed four luciferase-like monooxygenase class flavin-dependent oxidoreductase but none of the annotated alkanesulfonate monooxygenases. A potential explanation for this may be the larger size of these monooxygenases when compared to SsuD; however, this does not explain the initial annotation.

When the genomic context of these enzymes was compared to the genomic context of *ssuD* in *E. coli,* the genomic context of the genes encoding ISGA 205, 1835 and 08960 were most similar. These genes were found nearby a substrate and ATP- binding protein and a permease much like *ssuD* is. Furthermore, the gene encoding ISGA 08960 was found near an aliphatic sulfonate substrate binding protein. It is important to note that none of the operon-like regions identified encoded an *ssuE*-like or NADH:FMN oxidoreductase. The lack of a reductase in the operon-like regions does not necessarily suggest these regions are vestigial genomic structures or pseudogenes. When the *ssuE* homolog (*ssuI*) from *Corynebacterium glutamicum* ATCC 13032 was deleted, growth was retarded but not prevented from reaching sulfate-grown densities after prolonged incubation (Koch *et al.* 2005) when sulfur was limited to aliphatic sulfonates. Furthermore, the dibenzothiophene monooxygenase (DszA) can maintain its activity with its native (DszD) and non-native (Fre) NADH:FMN oxidoreductase (Adak *et al.*, 2016).

Although the genetic context of ISGA 205 and 1835 are indicative of potentially being alkanesulfonate monooxygenases, the phylogenetic placement of these enzymes among Class C monooxygenases is contradictory. ISGA 205 and 1835 were found closely related to LadA and DmoA, enzymes not associated with aliphatic sulfonate degradation. DmoA is capable of breaking the carbon-sulfur bond in dimethyl sulfoxide (DMSO), however, the substrate range of DmoA is restricted to DMSO only (Boden *et al*. 2011). Furthermore, transcriptomics data of *Gordonia* NB4-1Y grown on 6:2 FTSA as a sole sulfur source indicate that ISGA 205 and 1835 are not upregulated (Van Hamme personal communications, 2020). These findings strongly indicate that ISGA 205 and 1835 are not candidates for 6:2 FTSA degradation. ISGA 08960, on the other hand, is a curious case. The aforementioned transcriptomics data also show that ISGA 08960 is not upregulated (Van Hamme personal communications, 2020), however, when placed among Class C monooxygenases, ISGA 08960 aligned closely to the SsuD of *E. coli*, *B. subtilis* and *P.* putida.

### 4.2.5 *Gordonia* NB4-1Y mutagenesis

Finally, developing an efficient genetic manipulation tool for *Gordonia* NB4-1Y is critical to rescue any pitfalls or dead-ends biochemical characterizations might lead to. In this study, two methods were attempted to transfer a modified pK18mobsacB vector into *Gordonia* NB4-1Y for quasi-scarless deletion of the genes encoding ISGA 1218 and 1222. The first method was electroporation of glycerol competent cells following similar electroporation conditions in Veiga-Crespo *et al.* (2006). Two isolates were identified growing on NB agar with kanamycin. These two isolates were found and grew much slower than wild-type *Gordonia* NB4-1Y and had *E. coli* colony features but *fre*, an *E. coli* gene, could not be amplified via colony PCR. Both isolates appeared to contain DNA, likely plasmid, of an apparent size similar to pK18mobsacB1218AB. The pK18mobsacB vector can only be maintained in *E. coli* and closely related species outside of the genome, therefore, propagation of pK18mobsacB mediated kanamycin resistances requires integration into the host genome (Schafer *et al.*, 1997). These two isolates were, therefore,

considered contaminants and their identity remains a mystery. The second method was *E. coli* S17.1 mediated transfer by conjugation. After up to 8 days of conjugation, no *Gordonia* single recombinants were identified. Although these experiments yielded inconclusive results, the two slow growing *E. coli*-like isolates should be further investigated.

## 4.3 Limitations

### 4.3.1 Analyte quantification discrepancy

A major discrepancy in the analytical data in this study is the difference between octanal, 6:2 FTSA and sulfite quantification. GC-FID quantified octanal and HPLC quantified depletion of 6:2 FTSA was more than twice that of the sulfite quantified. Although this would suggest that one method is flawed, the likely culprit for the discrepancy is the oxidizable nature of sulfite. In the presence of reactive oxygen species or oxygen, sulfite can oxidize to sulfate, which is not detected with DTNB (Mader, 1958). Furthermore, in the presence of oxygen, $FMNH_2$ can oxidize to FMN while producing hydrogen peroxide (Massey, 1994). These together suggest that sulfite quantification is more accurate for shorter reactions where fewer $FMNH_2$ degradation products can oxidize sulfite. Sulfite produced by SsuD can be used to accurately assess substrate conversion and kinetic paraments if reaction times are limited to 2-3 minutes (Peng *et al*., 2019 and Zhan *et al*., 2008). In contrast, oxidizing all sulfite to sulfate and measuring sulfate concentrations is a more representative assessment of substrate turnover (as shown by Adak *et al*., 2016). Therefore, for the non-kinetic biochemical assessments of ISGA 1218, 1222 and SsuD, the quantified octanal or 6:2 FTSA depletion are more representative of the overall converted substrate.

### 4.3.2 Kinetic assessment

On the other hand, sulfite quantification during the kinetic assessment of SsuD was likely accurate due to the short reaction times and killing of the reaction just prior to analysis preventing further formation of $FMNH_2$. Carpenter *et al.* (2011) have reported $K_m$ values of octane sulfonate with respects to SsuD using a similar sulfite quantification method used in this study with SsuE and replicated the $K_m$ values described by Eichhorn *et al.* (1999), who followed NADH oxidation. The $K_m$ value calculated for octane sulfonate in this study were 1.45 times greater than those obtained

with SsuE by Eichhorn *et al.* (1999) and Carpenter *et al.* (2011). This likely suggests that $FMNH_2$ is not provided in excess in this study. Furthermore, PFOS had an apparent $K_m$ increasing effect on octane sulfonate and a contradictory $K_m$ reducing effect on 6:2 FTSA for SsuD. Furthering this, PFOS had an apparent uncompetitive inhibition on 6:2 FTSA while acting like a competitive inhibitor to octane sulfonate. These are contradictory since uncompetitive inhibition indicates PFOS binds the enzyme-substrate complex while competitive inhibition indicates PFOS can bind SsuD without substrate. The inconsistent effect of PFOS on SsuD kinetics, overlapping standard deviations of data points and higher than expected $K_m$ values is indicative that the methods used in the study require further refining. Dayal *et al.* (2015) found that $FMNH_2$ transfer between SsuE and SsuD does not require protein-protein interaction and therefore, $FMNH_2$ likely diffuses from the active site of SsuE to SsuD. Considering this, if $FMNH_2$ is in high enough concentration, the $K_m$ of octane sulfonate for SsuD should not change if provided by SsuE or another source.

### 4.3.3 High throughput *Escherichia coli* growth assay

In this study, a high throughput assessment of *E. coli* BL21(DE3) growth was desired in order to partially replicate the experiments of Eichhorn *et al.* (2000). In order to do this, a 96-well plate reader was set to measure $OD_{621}$ every hour for 48 hours placed in a 37°C incubator. Monitoring *E. coli* growth every hour for 48 hours using an automated system would have provided an accurate growth profile on different sulfur sources while avoiding awkward timing and time-consuming experiments. However, two issues arose during this attempted experiment. The first, wells bordering the edge of the 96-well plate having disproportionally larger $OD_{621}$ readings than the central wells and octane sulfonate containing wells measuring no growth. Growth was expected on octane sulfonate as it has been demonstrated several times before (Eichhorn *et al.* 1999 and 2000). The second issue was the attempt to produce a more oxygenated environment for *E. coli* BL21(DE3) growth. To achieve this same, the experiment was carried out without a lid to the 96-well plate. After 9 hours of incubation, the broth in the 96-well plate completely

evaporated leaving a short window where growth can be assessed. Attempts to produce a high throughput growth curve were unsuccessful and the qualitative scaled-up growth assay is more representative of *E. coli* growth when limited to different sulfur compounds.

## 4.4 Further studies

Moving forward, elucidating the roles of ISGA 1218 and 1222 in relation to 6:2 FTSA is critical in understanding why these enzymes are upregulated. Van Hamme *et al.* (2013) described the genes encoding ISGA 1218 and 1222 as nitrilotriacetate monooxygenases based on sequence similarity with other known monooxygenases. This, however, is unlikely considering Knobel *et al.* (1996) reported that structurally similar compounds such as citrate and EDTA were not transformed by NtaA. The only known Class C monooxygenase archetype which shared substrate archetype is EmoA which was capable of degrading nitrilotriacetate and EDTA (Bohuslavek *et al*. 2001). Foremost, identifying the native substrate of ISGA 1218 and 1222 is required. Given the results of the phylogenetic placement of ISGA 1218, it would be prudent to attempt to characterize ISGA 1218 in the presence of dibenzothiophene sulfone. Oshohiro *et al* (1999) and Adak *et al*. (2016) described HPLC parameters to detect 2-(2-hydroxy-phenyl)-benzenesulfinic acid, the degradation product of dibenzothiophene sulfone by DszA. ISGA 1222 on the other hand could be characterized in a similar manner as NtaA given EmoA shares substrate archetype with NtaA (Jun *et al*. 2016). *Aminobacter aminovorans* ATCC 29600, the renamed *Chelatobacter hentzii*, is available for purchase from the American Type Culture Collection and harbors the characterized gene encoding NtaA (Uetz *et al.*, 1992 and Kampfer *et al.*, 2002). If the NtaA characterized by Uetz *et al.* (1992) were cloned in a similar manner as SsuD, a positive control for nitrilotriacetate degradation could be developed. Uetz *et al.* (1992) described several analytical methods to confirm the activity of NtaA and could be adapted; most methods involve some use of HPLC, however, some GC methods have been used to detect esterified nitrilotriacetate or iminodiacetate (Parks *et al.*, 1981 and Chau and Fox, 1971). Once assigned to a monooxygenase archetype, the upregulation of ISGA 1218 and 1222 under sulfur limiting conditions in *Gordonia* NB4-1Y may be hypothesized.

Of the two *Gordonia* NB4-1Y monooxygenases cloned and purified in this study, ISGA 1218 remains an ideal candidate for further identification. This being because of its close alignment with DszA whereas all others aligned with non carbon-sulfur bond breaking enzymes. ISGA 08960 presented here is an ideal candidate for 6:2 FTSA degradation as it is the most closely related enzyme, by protein sequence, to the SsuD of *E. coli*, is near a transporter system but not a reductase. The characterized *ssu* operons in Gram-positive bacteria typically do not include a reductase (van der Ploeg *et al*. 1998, Kahnert *et al*. 2000 and Koch *et al*. 2005). Transcriptomic data (unpublished results, Van Hamme personal communication, 2020), however, indicate that the gene encoding ISGA 08960 is not upregulated in the presence of octane sulfonate or 6:2 FTSA as a sole sulfur source over $MgSO_4$ levels. Alternatively, a taurine dioxygenase may be responsible for aliphatic sulfonate and 6:2 FTSA degradation. Eichhorn *et al*. (1997) found that TauD and SsuD did share substrate, however, some aliphatic sulfonate had a significantly lower affinity for TauD than SsuD. *Gordonia* NB4-1Y does harbor ISGA 768 (Van Hamme *et al*. 2013), a putative TauD and could be the subject of further investigation, however, an analysis of the transcriptomics data is required moving forward.

Production of ISGA 1218 and 1222 was a problem with the vectors in this study. Of the protein production vectors used, pMAL1218 and pMAL1222 produced the lowest per gram of cell weight yields. Furthermore, when expressed with an MBP tag, several degradation products arose and ISGA 1218 and 1222 were insoluble with a histidine tag. Jun *et al*. (2016) found success expressing EmoA along with pGro7, a *groELS* chaperone protein which promotes protein folding. Expression of ISGA 1218 and 1222 could be attempted with an N-terminal histidine or T7 tag with the pET28b or pET23d vectors or, alternatively, an N-terminal glutathione S-transferase (GST) tag could be encoded with the pET41a vector. Furthermore, the above stated methods could be used to produce and purify ISGA 08960 and any future candidate monooxygenase responsible aliphatic sulfonate degradation in *Gordonia* NB4-1Y and consequently, likely 6:2 FTSA as well.

In light of the issues faced while attempting to produce an accurate *E. coli* growth curve when sulfur was limited to 6:2 FTSA, there is room for method refinement. A balance between oxygen permeability and retaining media volume within the wells must be reached. Instead of attempting to produce growth curves with a solid 96-well plate lid, oxygen permeable adhesive seals could be used instead. Zimmermann *et al*. (2003) reviewed several commercially available adhesive seals that were oxygen permeable whilst maintaining media volume within the wells of a 96-well plate. Similar films could be used in an attempt to produce accurate growth curves. Once accurate growth curves can be produced, understanding the fate of 6:2 FTSA in *E. coli* BL21(DE3) should be assessed. First, confirming the absence of contaminating sulfur sources in 6:2 FTSA stock is required. This can be achieved by analyzing stock aqueous 6:2 FTSA by LC-MS and searching for potential contaminants. Alternatively, if the contaminating sulfur source is hypothesized to be an aliphatic sulfonate, SsuD can be challenged with 6:2 FTSA and reactions can be extracted with ethyl acetate and analyzed by GC-MS. Any contaminating aliphatic sulfonate could be identified by the presence of its corresponding aldehyde. Following these experiments, 6:2 FTSA metabolites could be searched for following the methods by Van Hamme *et al*. (2013) and Shaw *et al*. (2019). Ascribing the initial degradation of 6:2 FTSA to the *ssu* operon would be completed by assessing the ability of *E. coli* mutants in *ssuA*, *B*, or *C* to grow on 6:2 FTSA. These genes have been implicated in the import of aliphatic sulfonate (Eichhorn *et al*. 1999) and mutagenesis of *E. coli* has been described by Eichhorn *et al.* (1999).

Development of an effective mutagenesis tool in *Gordonia* NB4-1Y is critical for understanding the roles of ISGA 1218 and 1222. Several Gram- positive shuttle vectors exist for gene deletion or manipulation as well as *E. coli/Gordonia* shuttle vectors. For example, the pT181 vector has been used for Gram-positive mutagenesis (Charpentier *et al.*, 2004) and the pNC9503 vector described by Arensk̈otter *et al.* (2003) has been used for mutagenesis in the *Gordonia* genus. During the course of this study, the pK18mobsacB was used and has been successfully used to

delete genes from Gram-positive and -negative bacterial genomes (Chan *et al.*, 2015 and Wang *et al.* 2015) and in the *Gordonia* sister genus, *Rhodococcus* (Otani *et al.* 2014). In this study, a small subset of conjugation and transformation methods were attempted. Veiga-Crespo *et al.* (2006) reported effective electroporation of *Gordonia* species by growing cells in penicillin G and isoniazid before being ultrasonicated; use of a cell wall inhibitor may decrease the overall integrity of the *Gordonia* NB4-1Y cell wall and promote electroporation. In addition, conjugation with *E. coli* S17.1 can be achieved with certain success in broths ranging from low to high salt concentrations and varying temperatures (Wang *et al.*, 2015); in this study, a consistent salt concentration and temperature were used. Once a mutagenesis method is developed, the most important genes to disrupt will be the genes encoding ISGA 1218, 1222. Disrupting the genes encoding ISGA 1218 and 1222 would shed light on their role in 6:2 FTSA metabolism.

Finally, the benefits of further exploring the substrate range and kinetics of SsuD would be two-fold: the substrate range of SsuD will support the hypothesis on its catalytic mechanism (Armacost *et al.,* 2015) and the kinetics of SsuD, with SsuE, would inform on the ability of *E. coli* to metabolize FTSA. Given Armacost *et al.* (2015) hypothesize the hydrogens bound to the alpha carbon of aliphatic sulfonates are crucial to regenerating of FMN during catalysis, challenging SsuD with 6:1 FTSA may shed light on the substrate limit of SsuD. In order to fully understand why SsuD can accept 6:2 FTSA as a substrate, substrate bound crystal structures must be produced. The catalytic mechanism of SsuD is thought to be sequential where $FMNH_2$ must first bind the active site before a substrate can bind and no amino acids are thought to directly mediate the nucleophilic attack of the c4a-peroxyflavin intermediate on the sulfur center (Armacost *et al.*, 2015 and Zhan *et al.,* 2008). This would necessitate the soaking of SsuD crystals with high enough concentrations of $FMNH_2$ and the addition of a non-substrate analogue. Possible analogues could be 1:7 or 2:6 FTSA where the backbone is hydrogenated but the alpha and beta carbons are saturated with carbon-fluorine bonds. On the other hand, the kinetics of SsuD have been well

established with its native reductase SsuE (Eichhorn *et al.*, 1999 and Carpenter *et al.,* 2011). With that said, the conditions used to evaluate the kinetic parameters of SsuD with SsuE could be applied to 6:2 FTSA and other AFFF breakdown products such as 4:2 and 8:2 FTSA (Harding-Marjanovic *et al.,* 2015). If correlated with appropriate transcriptomics data, this information could be used to estimate the ability of a biological system to metabolize AFFF or FTSA.

# 5.0 Conclusion

In summary, the alkanesulfonate monooxygenase (SsuD) from *Escherichia coli* BL21(DE3) seems to transform 6:2 fluorotelomer sulfonate (FTSA) to sulfite and an unidentified fluoroalkyl substance and, when 6:2 FTSA is prepared in water; *E. coli* BL21(DE3) may use it as a sulfur source, but the data available is preliminary at this time. Furthermore, genetic context comparison and phylogenetic placement of ISGA 1218, 1222, 205, 1666 and 1835 were unable to provide clear evidence to propose any monooxygenases as an alkanesulfonate monooxygenase. ISGA 08960, on the other hand, is strongly suggested to be an alkanesulfonate monooxygenase. This study represents the first biochemical evidence of fluorinated surfactant degradation by bacterial enzyme by identification of sulfite as a metabolite of 6:2 FTSA degradation. This study further posits the aliphatic sulfonate degradation pathway as the pathway which initially degrades 6:2 FTSA. The alkanesulfonate monooxygenase in *Gordonia* NB4-1Y produced when sulfur is limited to 6:2 FTSA is likely the enzyme responsible for the initial degradation of 6:2 FTSA, however, a taurine dioxygenase may also fill this role. To date, there remains a gap in the exact mechanism of biochemical transformation of fluorinated pollutants due to the lack of candidate genes from pure cultures. Although the previously hypothesized ISGA 1218 and 1222 did not demonstrate activity, this may be a comment on the instability of enzymes from the *Gordonia* genus rather than their lack of activity. Further investigation into the genes responsible for aliphatic sulfonate degradation in *Gordonia* NB4-1Y and development of effective mutagenesis tools for *Gordonia* NB4-1Y will shed light on the biochemical mechanisms used by *Gordonia* NB4-1Y to degrade 6:2 FTSA.

# 6.0 References

3M Company & Auer, C. Phase-out plan for POSF-based products. United States Environmental Protection Agency. (2000).

3M Company, Moore, J., Rodricks, J., Turnbull, D. & Warren-Hicks, W. Environmental and health assessment of perfluorooctane sulfonic acid and its salts. (2003).

Adak, S. & Begley, T. P. Dibenzothiophene catabolism proceeds via a flavin-N5-oxide intermediate. *Journal of the American Chemical Society* **138**, 6424–6426 (2016).

Alsmeyer, Y. W., Childs, W. V., Flynn, R. M., Moore, G. G. I. & Smeltzer, J. C. Organofluorine chemistry and its applications. in *Topics in Applied Chemistry* 121–143 (Springer, Boston, MA, 1994).

Apelberg, B. J. *et al.* Cord serum concentrations of perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA) in relation to weight and size at birth. *Environmental Health Perspectives* **115**, 1670–1676 (2007).

Arensk̈otter, M., Baumeister, D., Kalscheur, R. & Steinbuchel, A. Identification and application of plasmids suitable for transfer of foreign DNA to members of the genus *Gordonia*. *Applied and Environmental Microbiology* **69**, 4971–4974 (2003).

Arensk̈otter, M., Broker, D. & Steinbuchel, A. Biology of the metabolically divers genus *Gordonia*. *Applied and Environmental Microbiology* **70**, 3195–3204 (2004).

Arrieta-Cortes R., Farias P., Hoyo-Vadillo C., Kleiche-Dray M. Carcinogenic risk of emerging persistent organic pollutants perfluorooctane sulfonate (PFOS): A proposal of classification. *Regulatory Toxicology and Pharmacology*. **83**, 66-80 (2017).

Armacost, K., Musila, J., Gathiaka, S., Ellis, H. R., Acevedo O. Exploring the catalytic mechanism of alkanesulfonate monooxygenase using molecular dynamics. *Biochemistry* **53**, 3308–3317 (2014).

Arrieta-Cortes, R., Farias, P., Hoyo-Vadillo, C. & Kleiche-Dray, M. Carcinogenic risk of

emerging persistent organic pollutant perfluorooctane sulfonate (PFOS): A proposal of classification. *Regulatory Toxicology and Pharmacology* **83**, 66–80 (2017).

Arvaniti, O. S. & Stasinakis, A. S. Review on the occurrence, fate and removal of perfluorinated compounds during wastewater treatment. *Science of the Total Environment* 81–92 (2015).

Avendaño, S. M., Zhong, G. & Liu, J. Comment on "Biodegradation of perfluorooctanesulfonate (PFOS) as an emerging contaminant.". *Chemosphere* **138**, 1037–1038 (2015).

Basu, T., & Poddar, R. K. Effect of ethanol on *Escherichia coli* cells. Enhancement of DNA synthesis due to ethanol treatment. *Folia Microbiologica*, 39(1), 3–6 (1994).

Bedouelle, H. & Duplay, P. Production in *Escherichia coli* and one-step purification of bifunctional hybrid proteins which bind maltose. *European Journal of Biochemistry* **171**, 541–549 (1988).

Betts, K. S. Not Immune to PFOS Effects? *Environmental Health Perspectives* A290 (2008).

Boden R., Borodina E., Wood A. P., Kelly D. P., Murrell J. C., Schafer H. Purification and characterization of dimethylsulfide monooxygenase from *Hyphomicrobium sulfonivorans*. *Journal of Bacteriology*. **193**(5), 1250-1258 (2011).

Bohuslavek J., Payne J. W., Liu Y., Bolton H., Xun L. Cloning, sequencing and characterization of a gene cluster involved in EDTA degradation from the bacterium BNC1. *Applied and Environmental Microbiology*. **67**(2), 688-695 (2001).

Bykowski T., van der Ploeg J. R., Iwanicka-Nowicka R., Hryniewicz M. M. The switch from inorganic to organic sulfur assimilation in *Eschericha coli*: adenosine 5'-phosphsulfate (APS) as a signaling molecule for sulfate excess. *Molecular Microbiology* **43**(5), 1347-1358 (2002).

Buck, R. C. *et al.* Perfluoroalkyl and polyfluoroalkyl substances in the environment: terminology, classification, and origins. *Integrated Environmental Assessment and Management* **7**, 513–541 (2011).

Carpenter, R. A., Xiong, J., Robbins, J. M. & Ellis, H. R. Functional role of a conserved arginine residue located on a mobile loop of alkanesulfonate monooxygenase. *Biochemistry* **50**, (2011).

Chan, K. K. J. & O'Hagan, D. The rare fluorinated natural products and biotechnological prospects for fluorine enzymology. in *Natural Product Biosynthesis by Microorganisms and Plants* 219–235 (2012).

Chan, Y. C. Trifluoromethylation of carbonyl compounds with sodium trifluoroacetate. *Journal of Fluorine Chemistry* **126**, 937–940 (2005).

Chan, Y. C., Levar, C. E., Zacharoff, L., Badalamenti, J. P. & Bond, D. R. Scarless genome editing and stable inducible expression vectors for *Geobacter sulfurreducens*. *Applied Environmental Microbiology* **81**, 7178–7186 (2015).

Charpentier, E. *et al.* Novel cassette-based shuttle vector system for Gram-positive bacteria. *Applied and Environmental Microbiology* **70**, 6076–6085 (2004).

Chau, Y. K. & Fox, M. E. A GC method for the determination of nitrilotriacetic acid in lake water. *Journal of Chromatographic Science* **9**, 271–275 (1971).

City of Hamilton, Canada, Everson, N. & Sabo, R. Hamilton international airport - perfluorooctane sulphonate acid update (PED11223) (City Wide). (2011).

Collier, H. B. A note on the molar absorptivity of reduced Ellman's reagent, 3-carboxylato-4-nitrothiophenolate. *Analytical Biochemistry*, 56(1), 310–311 (1973).

Colosi, L. M., Pinto, R. A., Haung, Q. & Weber, W. J. Peroxide-mediated degradation of perfluorooctanoic acid. *Environmental Toxicology and Chemistry* **28**, 264–271 (2009).

D'Agostino, L. A. & Mabury, S. A. Aerobic biodegradation of 2 fluorotelomer sulfonamide-based aqueous film-forming foam components produces perfluoroalkyl carboxylates. *Environmental Toxicology and Chemistry* **36**, 2012–2021 (2017).

de solla, S. R., de Silva, A. O. & Letcher, R. J. Highly elevated levels of perfluorooctane sulfonate and other perfluorinated acids found in biota and surface water downstream of an

international airport, Hamilton, Ontario, Canada. *Environment International* **39**, 19–26 (2012).

Dayal, P. V., Singh H., Busenleher, L. S. & Ellis, H. R. Exposing the alkanesulfonate monooxygenase protein-protein interaction sites. *Biochemistry*. **54**(51), 7531-7538 (2015).

Denome S. A., Oldfield C., Nash L. J., Young K. D. Characterization of the desulfurization genes from *Rhodococcus* sp. Strain IGTS8. *American Society of Microbiology*. **176**(21), 6707-67126 (1994).

United States of America - Department of the Army. Limiting the use of aqueous film forming foam. (2016).

Eichhorn, E., Davey, C. A., Sargent, D. F., Leisinger, T. & Richmond, T. J. Crystal structure of *Escherichia coli* alkanesulfonate monooxygenase SsuD. *Journal of Molecular Biology* **324**, 457–468 (2002).

Eichhorn, E., van der Ploeg, J. R., Kertesz, M. & Leisinger, T. Characterization of alpha-ketoglutarate-dependent taurine dioxygenase from *Escherichia coli*. *Journal of Biological Chemistry* **272**, 23031–32036 (1997).

Eichhorn, E., van der Ploeg, J. R. & Leisinger, T. Characterization of a two-component alkanesulfonate monooxygenase from *Escherichia coli*. *Journal of Biological Chemistry.* **274**, 26639–26646 (1999).

Eichhorn, E., van der Ploeg, J. R. & Leisinger, T. Deletion Analysis of the *Escherichia coli* taurine and alkanesulfonate transport systems. *American Society for Microbiology* **182**, 2787–2795 (2000).

Eli Lilly Australia. PROZAC, Fluoxetine hydrochloride. (2016).

Ellis, H. R. The FMN-dependent two-component monooxygenase systems. *Archives of Biochemistry and Biophysics* **497**, 1–12 (2010).

Ellis, H. R. Mechanism for sulfur acquisition by the alkanesulfonate monooxygenase system. *Bioorganic Chemistry* **39**, 178–184 (2011).

Eschauzier, C., Beerendonk, E., Scholte-Veenendaal, P. & De Voogt, P. Impact of treatment

processes on the removal of perfluoroalkyl acids from the drinking water production chain. *Environmental Science & Technology* **46,** 1708-1715 (2011).

European Parliament. Amending for the 30th time Council Directive 76/769/EEC on the approximation of the laws, regulations and administrative provisions of the Member States relating to restrictions on the marketing and use of certain dangerous substances and preparations (perfluorooctane sulfonates). (2006).

Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783-791 (1985).

Feng L., Wang W., Cheng J., Ren Y., Zhao G., Gao C., Tang Y., Liu X., Han W., Peng X., Liu R., Wang L. Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proceedings of the National Academy of Science*. **104**(13), 5602-5607 (2006).

Ferrario, V. *et al.* Elucidating the structural and conformational factors responsible for the activity and substrate specificity of alkanesulfonate monooxygenase. *Journal of Biomolecular Structure and Dynamics* **30**, 74–88 (2012).

Fisher A. J., Thomspon T. B., Thoden J. B., Baldwin T. O., Rayment I. The 1.5-Å resolution crystal structure of bacterial luciferase in low salt conditions. *Journal of Biological Chemistry*. **271** (36), 21956-21968 (1996).

Gao, Y. *et al.* Perfluorooctanesulfonate (PFOS)-induced Sertoli cell injury through a disruption of F-actin and microtubule organization is mediated by Akt1/2. *Nature Scientific Reports* **24**, 1–14 (2017).

Girbble, G. W. Naturally occurring organofluorines. in *Handbook of Environmental Chemistry* 121–136 (2002).

Government of Canada. Virtual Elimination List - Canadian environmental protection act. (2009).

Hamilton, J. T. G., Muhammad, R. A., Harper, D. B. & O'Hagan, D. Biosynthesis of

fluoroacetate and 4-fluorothreonine by *Streptomyces cattleya*. Glycine and pyruvate as precursors. *Chemical Communications* 797–798 (1997).

Hanson, S. R., Best, M. D. & Wong, C. H. Sulfatases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angewandte Chemie International* **43**, 5736–5763 (2004).

Harding-Marjanovic, K. C. *et al.* Aerobic biotransformation of thioether amido sulfonate (Lodyne) in AFFF-amended microcosms. *Environmental Science & Technology* **49**, 7666–7674 (2015).

Henikoff S. and Henikoff J. G. Position-based sequence weights. *Journal of Molecular Biology*. **243**(4), 574-578 (1994).

Hill, P. J., Taylor, M., Goswami, P. & Blackburn, R. S. Substitution of PFAS chemistry in outdoor apparel and the impact on repellency performance. *Chemosphere* **181**, 500–507 (2017).

Hindson, J. V. Serine acetyltransferase of *Escherichia coli*: substrate specificity and feedback control by cysteine. *Biochemical Journal* **375**, 745–752 (2003).

Hogan, D. A., Smith, S. R., Saari, E. A., McCracken, J. & Hausinger, R. P. Site-directed mutagenesis of 2,4-dichlorophenoxyacetic acid/alpha-ketoglutarate dioxygenase. *Journal of Biological Chemistry* **275**, 12400–12409 (2000).

Houde, M. *et al.* Polyfluoroalkyl compounds in free-ranging bottlenose dolphins (*Tursiops truncatus*) from the Gulf of Mexico and the Atlantic Ocean. *Environmental Science & Technology* **39**, 6591–6598 (2005).

Huang, S. & Jaffe, P. R. Isolation and characterization of an ammonium-oxidizing iron reducer: *Acidimicrobiaceae* sp. A6. *PLoS One* **13**, 1–12 (2018).

Huang, S. & Jaffe, P. R. Defluorination of perfluorooctanoic acid (PFOA) and perfluorooctane sulfonate (PFOS) by *Acidimicrobium* sp. strain A6. *Environmental Science & Technology* **53**, 11410–11419 (2019).

Huijbers, M. M. E., Montersino, S., Westphal, A., Tischler, D. & van Berkel, W. J. H. Flavin dependent monooxygenases. *Archives of Biochemistry and Biophysics* **544**, 2–17 (2014).

Isanbor, C. & O'Hagan, D. Fluorine in medicinal chemistry: A review of anti-cancer agents. *Journal of Fluorine Chemistry* **127**, 303–319 (2006).

Iwaki H., Grosse S., Bergeron H., Leisch H., Morley K., Hasegawa Y., Lau P. C. K. Camphor pathway redux: function recombination expression of 2,5- and 3,6- diketochamphane monooxygenase of *Pseudomonas putida* ATCC 17543 with their cognate flavin reductase catalyzing Baeyer-Villiger reactions. *Applied and Environmental Microbiology*. **79**(10), 3282-3293 (2013).

Iwanicka-Nowicka R., Hryniewicz M. M. A new gene, *cbl*, encoding a member of the LysR family of transcriptional regulators belongs to *Escherichia coli cys* regulon. *Gene*. **166**(1), 11-17 (1995).

Iwanicka-Nowicka R., Zielak A., Cook A. M., Thomas M. S., Hryniewicz M. M. Regulation of sulfur assimilation pathways in *Burkholderia cenocepacia*: Identification of transcription factors CysB and SsuR and their role in control of target genes. **189**(5), 1675-1688 (2007).

Jun S. Y., Lewis K. M., Youn B., Xun L., Kang C. Structural and biochemical characterization of EDTA monooxygenase and its physical interaction with a partner flavin reductase. *Molecular Microbiology*. **100**(6), 989-1003 (2016).

Kahnert A., Bermeij P., Wietek C., James P., Leisinger T., Kertesz M. A. The *ssu* locus plays a key role in organosulfur metabolism in *Pseudomonas putida* S-313. *Journal of Bacteriology*. **182**(10), 2869-2878 (2000).

Kampfer, P., Neef, A., Salkinoja-Salonen, M. S. & Busse, H. J. *Chelotbacter heintzii* (Auling *et al.* 1993) is a later subjective synonym of *Aminobacter aminovorans* (Urakami *et al*. 1992). *International Journal of Systematic and Evolutionary Microbiology* **52**, 835–839 (2002).

Kertesz, M. A. Riding the sulfur cycle - metabolism of sulfonates and sulfate esters in Gram-negative bacteria. *FEMS Microbiology Reviews* **24**, 135–175 (1999).

Key B. D., Howell R. D., Criddle C. S., Defluorination of organofluorine sulfur compounds by *Pseudomonas* sp. Strain D2. *Environmental Science and Technology*. **32**(15) 2283-2287 (1998).

Kim, M. H., Wang, N., McDonald, T. & Chu, K. H. Biodefluorination and biotransformation of fluorotelomer alcohols by two alkane-degrading *Pseudomonas* strains. *Biotechnology and Bioengineering.* (2012).

Kim, S. B., Brown, R., Gilbert, S. C., Iliarionov, S. & Goodfellow, M. *Gordonia amicalis* sp. nov., a novel dibenzothiophene desulphurizing actinomycete. *International Journal of Systematic and Evolutionary Microbiology* **50**, 2031–2036 (2000).

Kim, S. B., Brown, R., Oldfield, C. & Gilbert, S. C. *Gordonia desulfuricans* sp. nov., a benzothiophene desulphurizing actinomycete. *International Journal of Systematic and Evolutionary Microbiology* **49**, 1845–1851 (1999).

Kim, S. K. & Kannan, K. Perfluorinated acids in air, rain, snow, surface runoff, and lakes: relative importance of pathways to contamination of urban lakes. *Environmental Science & Technology* **41**, 8328–8334 (2007).

Kissa, E. Fluorinated surfactants: Synthesis, properties and applications. (Marcel Dekker: New York., 1994).

Kissa, E. Fluorinated Surfactants and Repellents: Second Edition, Revised and Expanded Surfactants Science Series. vol. 97 (Marcel Dekker: New York., 2001).

Koch D. J., Ruckert C., Albersmeier A., Huser A. T., Tauch A., Puhler A., Kalinowski J. Transcriptional regulator SsuR activates expression of the *Corynebacterium glutamicum* sulphonate utilization genes in the absence of sulphate. *Molecular Microbiology*. **58**, 480-494 (2005).

Koch *et al.* Role of the *ssu* and *seu* genes of *Corynebacterium glutamicum* ATCC 13032 in utilization of sulfonates and sulfonate esters as sulfur sources. *Applied and Environmental Microbiology* **71**, 6104–6114 (2005).

Könnecker, G., Regelmann, J., Belanger, S., Gamon, K. & Sedlak, R. Environmental properties and aquatic assessment of anionic surfactants: physico-chemical, environmental fate and ecotoxicity properties. **74**, 1445–1460 (2011).

Kredich, N. M. Biosynthesis of Cysteine. *EcoSal Plus* **3**, (2008).

Kumar S., Stecher G., Li M., Knyaz C., and Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* **35**, 1547-1549 (2018).

Kwon, B. G. *et al.* Biodegradation of perfluorooctanesulfonate (PFOS) as an emerging contaminant. *Chemosphere.* (2013).

Lauble, H., Kennedy, M. C., Emptage, M. H., Beinert, H. & Stout, C. D. The reaction of fluorocitrate with aconitase and the crystal structure of the enzyme-inhibitor complex. *Proceedings of the National Academy of Science* **93**, 13699–13703 (1996).

Lehmler, H. J. Synthesis of environmentally relevant fluorinated surfactants - a review. *Chemosphere* **58**, 1471–1496 (2005).

Li L., Liu X., Yang W., Xu F., Wang W., Lu F., Bartla, M., Wang L., Rao Z. Crystal structure of long-chain alkane monooxygenase (LadA) in complex with coenzyme FMN: unveiling the long-chain alkane hydroxylase. *Journal of Molecular Biology*. **376**(2), 453-465 (2007).

Li, Y. *et al.* Half-lives of PFOS, PFHxS and PFOA after end of exposure to contaminated drinking water. *Occupational and Environmental Medicine* **75**, 46–51 (2018).

Liew, Z., Goudarzi, H. & Oulhote, Y. Developmental exposures to perfluoroalkyl substances (PFAS): An update of associate health outcomes. *Current Environmental Health* **5**, 1–19 (2018).

Liu, J. & Avendaño, S. M. Microbial degradation of polyfluoroalkyl chemicals in the environment: a review. *Environment International* **61**, 98–114 (2013).

Liu, J., Lee, L. S., Nies, L. F., Nakatsu, C. H. & Turco, R. F. Biotransformation of 8:2 fluorotelomer alcohol in soil and by soil bacteria isolates. *Environmental Science & Technology* **41**, 8024–8030 (2007).

Liu, J. *et al.* 6-2 Fluorotelomer alcohol aerobic biodegradation in soil and mixed bacterial culture. *Chemosphere* **78**, 437–444 (2010).

Loganathan, B. G., Sajwan, S. K., Sinclair, E., Kumar, K. S. & Kannan, K. Perfluoroalkyl

sulfonates and perfluorocarboxylates in two wastewater treatment facilities in Kentucky and Georgia. *Water Research* 4611–4620 (2007).

Mader, P. M. Kinetics of the hydrogen peroxide-sulfite reaction in alkaline solution. *Journal of the American Chemical Society.* **80**, 2634-2639 (1958).

Marais, J. C. S. The isolation of the toxic principle “K cymonate” from “Gifblaar”, *Dichapetalum cymosum*. *Onderstepoort Journal of Veterinary Science and Animal Industry* **18**, (1943).

Massey, V. Activation of molecular oxygen by flavins and flavoproteins. *Journal of Biological Chemistry.* **269**, 22459-22462, (1994).

Matsubara, T., Ohshiro, T., Nishina, Y. & Izumi, Y. Purification, characterization, and over expression of flavin reductase involved in dibenzothiophene desulfurization by *Rhodococcus erythropolis* D-1. *Applied and Environmental Microbiology* **67**, 1179–1184 (2001).

McAmmond, B. Subcloning of ISGA 205, 1666 and 1835 *ssuD* genes from *Gordonia* sp. NB4-1Y. Thompson Rivers University, (2017).

Milton-Wood, C. Subcloning of ISGA 1218 & 1222 NtaA genes from *Gordonia* sp. NB4-1Y for large scale protein production. Thompson Rivers University, (2016).

Mitchel, A., Spencer, M. & Edmiston, C. Jr. Role of healthcare apparel and other healthcare textiles in the transmission of pathogens: a review of the literature. *Journal of Hospital Infections* **90**, 285–292 (2015).

Moody, C. A. & Field, J. A. Determination of perfluorocarboxylates in groundwater impacted by fire-fighting activity. *Environmental Science & Technology* **33**, (1999).

Moody, C. A. & Field, J. A. Perfluorinated surfactants and the environmental implications of their use in fire-fighting foams. *Environmental Science & Technology* **34**, 3864-3870 (2000).

Moody, C. A., Herbery, G. N., Strauss, S. H. & Field, J. A. Occurrence and persistence of perfluorooctanesulfonate and other perfluorinated surfactants in groundwater at a fire-training area at Wurtsmith Airforce Base, Michigan, USA. *Journal of Environmental Monitoring* **5**, 341–

345 (2003).

Morgenthaler, M. *et al.* Predicting and tuning physicochemical properties in lead optimization: Amine Basicities. *ChemMedChem* **2**, 1100–1115 (2007).

Morton, G. O., Lancaster, J. E., Van Lear, G. E., Fulmor, W. & Meyer, W. E. Structure of nucleocidin. III. Revised structure. *Journal of the American Chemical Society.* **91**, 1535–1537 (1969).

Mueller J. W., Shafqat N. Adenosine-5′-phosphosulfate – a multigaceted modulator of bifunctional 3′-phospho-adenosine-5′-phosphosulfate synthases and related enzymes. **280**(13), 3050-3057 (2013).

Mukherjee T., Zhang Y., Abdelwahed S., Ealick S. E., Begley T. P. Catalysis of a flavoenzyme-mediated amide hydrolysis. *Journal of the American Chemical Society*. **132**(16), 5550-5551 (2010).

Muir, D. *et al.* Levels and trends of poly- and perfluoroalkyl substances in the arctic environment - An update. *Emerging Contaminants* **5**, 240–271 (2019).

Muller R. and Yingling V. Naming conventions and physical and chemical properties of per- and polyfluoroalkyl substances (PFAS). *Interstate Technology Regulation Council.* 1-15 (2017).

Murphy, C. D., Clark, B. R. & Amadio, J. Metabolism of fluoroorganic compounds in microorganisms: impacts for the environment and the production of fine chemicals. *Applied Microbial Biotechnology.* **84**, 617–629 (2009).

Novagen. pET System Manual. (2003).

O'Hagan, D. Understanding organofluorine chemistry. An introduction to the C-F bond. *Chemical Society Reviews* **37**, 308–319 (2008).

O'Hagan, D. & Harper, D. B. Fluorine-containing natural products. *Journal of Fluorine Chemistry* 127–133 (1999).

O'Loughlin, E. J., Traina, S. J. & Sims, G. K. Effects of sorption on the biodegradation of 2-methylpyridine in aqueous suspensions of reference clay minerals. *Environmental Toxicology and*

*Chemistry* **19**, 2168–2174 (2009).

Ochoa-Herrera, V., Field, J. A., Luna-Velasco, A. & Sierra-Alvarez, R. Microbial toxicity and biodegradability of perfluorooctane sulfonate (PFOS) and shorter chain perfluoroalkyl and polyfluoroalkyle substances (PFAS). *Environmental Science Processes & Impacts* **18**, 1236–1246 (2016).

Ochoa-Herrera, V. *et al.* Reductive defluorination of perfluorooctane sulfonate. *Environmental Science & Technology* **42**, 3260–3264 (2008).

Ohshiro, T., Kojima, T., Torii, T., Kawasoe, H. & Izumi, Y. Purification and characterization of dibenzothiophene (DBT) sulfone monooxygenase, an enzyme involved in DBT desulfurization, from *Rhodococcus erythropolis* D-1. *Journal of Bioscience and Bioengineering* **88**, 610–616 (1999).

Olsen, G. W. *et al.* Perfluorooctanesulfonate and other fluorochemicals in the serum of American Red Cross adult blood donors. *Environmental Health Perspectives* **111**, (2003).

Olsen, G. W. *et al.* Temporal trends of perfluoroalkyl concentrations in American Red Cross adult blood donors, 2000-2010. *Environmental Science & Technology* **46**, 6330–6338 (2012).

Otani H., Lee Y. E., Casabon I., Eltis L. D. Characterization of *p*-hydroxycinnamate catabolism in a soil actinobacterium. *Journal of Bacteriology*. **196**, 4293-4303 (2014).

Parkes, D. G., Caruso, M. G. & Spardling, J. E. Determination of nitrilotriacetic acid in ethylenediaminetetraacetic acid disodium salt by reversed-phase ion pair liquid chromatography. *Analytical Chemistry* **53**, 2154–2156 (1981).

Paul, A. G., Jones, K. C. & Sweetman, A. J. A first global production, emission and environmental inventory for perfluorooctane sulfonate. *Environmental Science & Technology* **43**, 386–392 (2009).

Payne, J. W., Bolton, H., Campbell, J. A. & Xun, L. Purification and Characterization of EDTA monooxygenase from the EDTA-degrading bacterium BNC1. *Journal of Bacteriology* **180**, 3823–3827 (1998).

Penden-Adams, M. M. *et al.* Developmental toxicity in white leghorn chickens following *in ovo* exposure to perfluorooctane sulfonate (PFOS). *Reproductive Toxicology* **27**, 307–318 (2009).

Peri K. G., Goldie H. Waygood E. B. Cloning and characterization of the N-acetylglucosamine operon of *Escherichia coli*. *Biochemistry and Cell Biology*. **68**(1), 123-127 (1990).

Peters, R. A., Hall, R. J., Ward, P. F. V. & Sheppard, N. The chemical nature of the toxic compounds containing fluorine in the seeds of *Dichapetalum toxicarium*. *Biochemical Jorunal* **77**, 17–22 (1960).

Peng C., Huang D., Shi Y., Zhang B., Sun L., Li M., Deng X., Wang W. Comparative transcriptomics analysis revealed the key pathways responsible for organic sulfur removal by thermophilic bacterium *Geobacillus thermoglucosidasius* W-2. *Science of Total Environment*. **676**, 639-650 (2019).

Place, B. J. & Field, J. A. Identification of novel fluorochemicals in aqueous film forming foams (AFFF) used by the US military. *Environmental Science & Technology* **46**, 7120–7127 (2012).

Poulson, P. B. *et al.* Substitution of PFOS for use in non-decorative hard chrome plating. *Danish environmental protection agency*, (2011).

Purser, S., Moore, P. R., Swallow, S. & Gouverneur, V. Fluorine in medicinal chemistry. *Chemical Society Reviews* **37**, 320–330 (2007).

Raran-Kurussi, S. & Waugh, D. S. The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperon-mediated. *PLoS One* **7**, (2012).

Rey D. A., Nentwich S. S., Koch D. J., Ruckert C., Puhler A., Tauch A., Kalinowski J. The McbR repressor modulated by the effector substance *S*-adenosylhomocystein controls directly the transcription of a regulon involved in sulphur metabolism of *Corynebacterium glutamcum* ATCC 13032. *Molecular Microbiology*. **56**(4), 871-887 (2005).

Robbins, M. R. & Ellis, H. R. Identification of critical steps governing the two-component alkanesulfonate monooxygenase catalytic mechanism. *Biochemistry* **51**, 6378–6387 (2012).

Robbins, M. R. & Ellis, H. R. Steady-state kinetic isotope effects support a complex role of Arg226 in the proposed desulfonation mechanism of alkanesulfonate monooxygenase. *Biochemistry* **53**, 161–168 (2014).

Romero, E., Castellanos, J. R. G., Gadda, G., Fraaije, M. W. & Mattevi, A. Same substrate, many reactions: oxygen activation in flavoenzymes. *Chemical Reviews* **118**, 1742–1769 (2018).

Ruckert C., Milse J., Albersmeier A., Kock D. J., Puhler A., Kalinowski J. The dual transcriptional regulator CysR in *Corynebacteriuma glutamicum* ATCC 13032 controls a subset of genes of the McbR regulon in response to the availability of sulphide acceptor molecules. *BMC Genomiocs*. **9**(1) (2008).

Saitou N. and Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425 (1987).

Sawayama, S. Possibility of anoxic ferric ammonium oxidation. *Journal of Bioscience and Bioengineering* **101**, 70–72 (2006).

Schafer, A. *et al.* Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* **145**, 69–73 (1994).

Schellenberger, S. *et al.* Highly fluorinated chemicals in functional textiles can be replaced by re-evaluating liquid repellency and end-user requirements. *Journal of Cleaner Production* **217**, 134–143 (2019).

Schmitz, R. P. H. *et al.* Evidence for a radical mechanism of the dechlorination of chlorinated propenes mediated by the tetrachloroethene reductive dehalogenase of *Sulfurospirillum multivorans*. *Environmental Science & Technology* **41**, (2007).

Schultz M., M. M., Barofsky, D. F. & Field, J. A. Quantitative determination of fluorotelomer sulfonates in groundwater by LC MS/MS. *Environmental Science & Technology* 1828–1835

(2004).

Scott, B. F. *et al.* Analysis for perfluorocarboxylic acids/anions in surface waters and precipitation using GC–MS and analysis of PFOA from large-volume samples. *Environmental Science & Technology* **40**, 6405–6410 (2006).

Scott, C. *et al.* A global response to sulfur starvation in *Pseudomonas putida* and its relationship to the expression of low-sulfur-content proteins. *Federation of European Microbiological Societies, Microbiology Letters* **267**, 184–193 (2006).

Shankar, A., Xiao, J. & Ducatman, A. Perfluoroalkyl chemicals and chronic kidney disease in US adults. *American Journal of Epidemiology* **174**, 893–900 (2011).

Shaw, D. M. J. *et al.* Degradation and defluorination of 6:2 fluorotelomer sulfonamidoalkyl betaine and 6:2 fluorotelomer sulfonate by *Gordonia* sp. strain NB4-1Y under sulfur-limiting conditions. *Science of the Total Environment* **647**, 690–698 (2019).

Sinclair, E. & Kannan, K. Mass loading and fate of perfluoroalkyl surfactants in wastewater treatment plants. *Environmental Science & Technology* 1408–1414 (2006).

Smart, B. E. Fluorine substituent effects (on bioactivity). *Journal of Fluorine Chemistry* **109**, 3–11 (2001).

Smithwick, M. *et al.* Temporal trends of perfluoroalkyl contaminants in polar bears (*Ursus maritimus*) from two locations in the North American Arctic, 1972-2002. *Environmental Science & Technology* **40**, 1139–1143 (2006).

Stec E., Witkowska-Zimny M., Hryniewicz M. M., Neumann P., Wilkinson A. J., Brzozowski A. M., Verma C. S., Zaim J., Wyzocki S., Burjacz G. D. *E. coli*: Cyrstal structure and mutational analysis of the cofactor-binding domain of the cbl transcriptional regulator. *Journal of Molecular Biology*. **364**(3), 309-322 (2006).

Sullivan, M. Addressing perfluorooctane Sulfonate (PFOS) and perfluorooctanoic acid (PFOA). (2018).

Taves, D. R. Evidence that there are two forms of fluoride in human serum. *Nature* **217**,

1050–1051 (1968).

Tchobanoglous, G. & Burton, F. L. Wastewater characteristics. in *Wastwater Engineering - Treatment, Disposal and Reuse* 47–119 (McGraw-Hill, 1991).

Thibaut, D. *et al.* Purification of the two-enzyme system catalyzing the oxidation of D-proline residue of pristinamycin IIb during the last step of pristinamycin IIa biosynthesis. *American Society for Microbiology* **177**, 5199–5205 (1995).

Thomas D., Surdin-Kerjan Y. Metabolism of sulfur amino acids in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology reviews*. **61**(4), 503-532 (1997).

Titgemeyer F., Reizer J., Reizer A., Saier M. H. Evolutionary relationship between sugar kinases and transcriptional repressors in bacteria. *Microbiology*. **140**(9), 2349-2354 (1994).

Toesch, M., Schober, M. & Faber, K. Microbial alkyl- and aryl-sulfatases: mechanism, occurrence, screening and stereoselectivities. *Applied Microbial Biotechnology.* **98**, 1485–1496 (2014).

Tralau *et al.* Transcriptomic analysis of the sulfate starvation response of *Pseudomonas aeruginosa*. *American Society for Microbiology* **189**, (2007).

Uetz, T., Schneider, R., Snozzi, M. & Egli, T. Purification and characterization of a two-component monooxygenase that hydroxylates nitrilotriacetate from “Chelatorbacter” strain ATCC 29600. *Journal of Bacteriology* **174**, 1179–1188 (1992).

Unite States environmental protection agency. Long-chain perfluorinated chemicals (PFCs) action plan. (2009).

van Berkel, W. J. H., Kamerbeek, N. M. & Fraaije, M. W. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *Journal of Biotechnology* **124**, 670–689 (2006).

van de Vijver, K. I. *et al.* Tissue distribution of perfluorinated chemicals in harbor seals (*Phoca vitulina*) from the Dutch Wadden sea. *Environmental Science & Technology* **39**, 6978–6984 (2005).

van der Ploeg J. R., Cummings N. J., Leisinger T., Connerton I. F. *Bacillus subtilis* genes for

the utilization of sulfur from aliphatic sulfonates. *Microbiology*. **144**, 2555-2561 (1998).

van der Ploeg, J. R., Eichhorn, E. & Leisinger, T. Sulfonate-sulfur metabolism and its regulation in *Escherichia coli*. *Archives of Microbiology* **176**, 1–8 (2001).

van der Ploeg J. R, Iwanicka-Nowicka R., Bykowski T., Hryniewicz M. M., Leisinger T. The *Escherichia coli ssuEADCB* gene cluster is required for the utilization of sulfur from aliphatic sulfoantes and is regulated by the transcriptional activator Cbl. *Journal of Biological Chemistry*. **274**(43), 29358-29365 (1999).

van der Ploeg J. R., Iwanicka-Nowicka R., Kertesz M. A., Leisinger T., Hryniewicz M. M. Involvement of CysB and Cbl regulatory proteins in expression of the *tauABCD* operon and other sulfate starvation-inducible genes in *Escherichia coli*. *Journal of Bacteriology*. **179**(24), 7671-7678 (1997).

Van Hamme, J. D., Bottos, E. M., Bilbey, N. J. & Brewer, S. E. Genomic and proteomic characterization of *Gordonia* sp. NB4-1Y in relation to 6:2 fluorotelomer sulfonate biodegradation. *Microbiology* **159**, 1618–1628 (2013).

Veiga-Crespo, P., Feijoo-Siota, L., de Miguel, T. & Villa, T. G. Proposal of a method for the genetic transformation of *Gordonia jacobaea*. *Journal of Applied Microbiology* **100**, 608–614 (2006).

Verreault, J. *et al.* Perfluorinated alkyl substances in plasma, liver, brain and eggs of glaucous gulls from the Norwegian Arctic. *Environmental Science & Technology* **39**, 7439–7445 (2005).

Virk, R. K., Ramaswamy, G. N., Bourham, M. & Bures, B. L. Plasmas and antimicrobial treatment of nonwoven fabrics for surgical gowns. *Textile Research Journal* **74**, 1073–1079 (2004).

Wang, J. *et al.* Fluorine in pharmaceutical industry: Fluorine-containing drugs introduced to the market in the last decade (2001-2011). *Chemical Reviews* **114**, 2432–2506 (2014).

Wang, N. *et al.* 6:2 Fluorotelomer sulfonate aerobic biotransformation in activated sludge

wastewater treatment plants. *Chemosphere* 853–858 (2010).

Wang, P. *et al.* Development of an efficient conjugation-based genetic manipulation system for *Pseudoaltermonas*. *Microbial Cell Factories* **14**, 1–11 (2015).

Wang, Z., Cousins, I. T., Scheringer, M., Buck, R. C. & Hungerbuhler, K. Global emission inventories for C4-C14 perfluoroalkyl carboxylic acid (PFCA) homologues from 1951 to 2030, Part I: production and emissions from quantifiable sources. *Environment International* **70**, 62–75 (2014).

Washino, N. *et al.* Correlations between prenatal exposure to perfluorinated chemicals and reduced fetal growth. *Environmental Health Perspectives* **117**, 660–667 (2009).

Westgate, P. J. & Park, C. Evaluation of proteins and organic nitrogen in wastewater treatment effluents. *Environmental Science & Technology* **44**, 5352–5357 (2010).

Wong, D. T., Bymaster, F. P. & Engleman, E. A. Prozac (Fluoxetine, Lilly 110140), the first selective serotonin uptake inhibitor and an antidepressant drug: twenty years since its first publication. *Life Sciences* **57**, 411–441 (1995).

Wong, F. *et al.* Assessing temporal trends and source regions of per- and polyfluoroalkyl substances (PFAS) in air under the Arctic monitoring and assessment program (AMAP). *Atmospheric Environment* **172**, 65–73 (2018).

Xiao, F. Emerging poly- and perfluoroalkyl substances in the aquatic environment: A review of current literature. *Water Research* **124**, 482–495 (2017).

Xu, Y. *et al.* Cloning, sequencing, and analysis of a gene cluster from *Chelatobacter heintzii* ATCC 2960 encoding nitrilotriacetate monooxygenase and NADH:flavin mononucleotide oxidoreductase. *Journal of Bacteriology* 1112–1116 (1997).

Yamashita, N. *et al.* A global survey of perfluorinated acids in oceans. *Marine Pollution Bulletin* **51**, 658–668 (2005).

Yeung, L. W. Y. *et al.* Perfluorinated compounds and total and extractable organic fluorine in human blood samples from China. *Environmental Science & Technology* **42**, 8140–8145

(2008).

Yoo, H. *et al.* Perfluoroalkyl acids in the egg yolk of birds from Lake Shihwa, Korea. *Environmental Science & Technology* **42**, 5821–5827 (2008).

Zhan, X., Carpenter, R. A. & Ellis, H. R. Catalytic importance of the substrate binding order for the $FMNH_2$-dependent alkanesulfonate monooxygenase enzyme. *Biochemistry* **47**, 2221–2230 (2008).

Zhang, P. *et al.* Extracellular protein analysis of activated sludge and their functions in wastewater treatment plant by shotgun proteomics. *Nature Scientific Reports* **5**, (2015).

Zhang, S., Lu, X., Wang, N. & Buck, R. C. Biotransformation potential of 6:2 fluorotelomer sulfonate (6:2 FTSA) in aerobic and anaerobic sediment. *Chemosphere* **154**, 224–230 (2016).

Zimmermann H. F., John G. T., Trauthwein H., Dingerdissen U., Huthmacher K. Rapid evaluation of oxygen and water permeation through microplate sealing tapes. *Biotechnology Progress* **19,** 1061-1063 (2003).

Zuckerkland E. and Pauling L. Evolutionary divergence and convergence in proteins. Edited in Evolving Genes and Proteins by V. Bryson and H.J. Vogel, pp. 97-166. *Academic Press, New York* (1965).

# 7.0 Appendix

## 7.1 Calibration curves

### 7.1.1 Octanal calibration curve

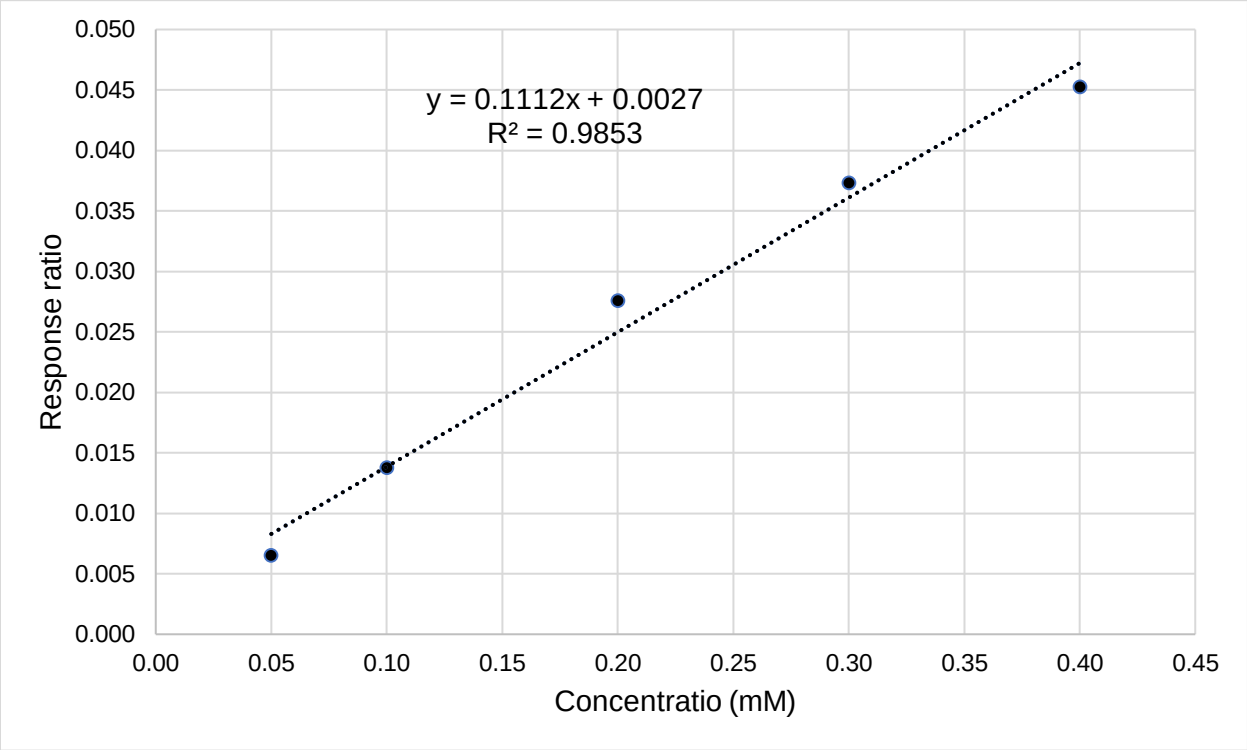


Figure S1. Octanal standard curve of 50 to 400 μM calculated from the response ratio of octanal to decanal. Standards were prepared by diluting octanal working stocks to concentrations of 50, 100, 200, 300 and 400 μM in ethyl acetate. Error bars represent standard deviation (n=3).

### 7.1.2 Octanol calibration curve

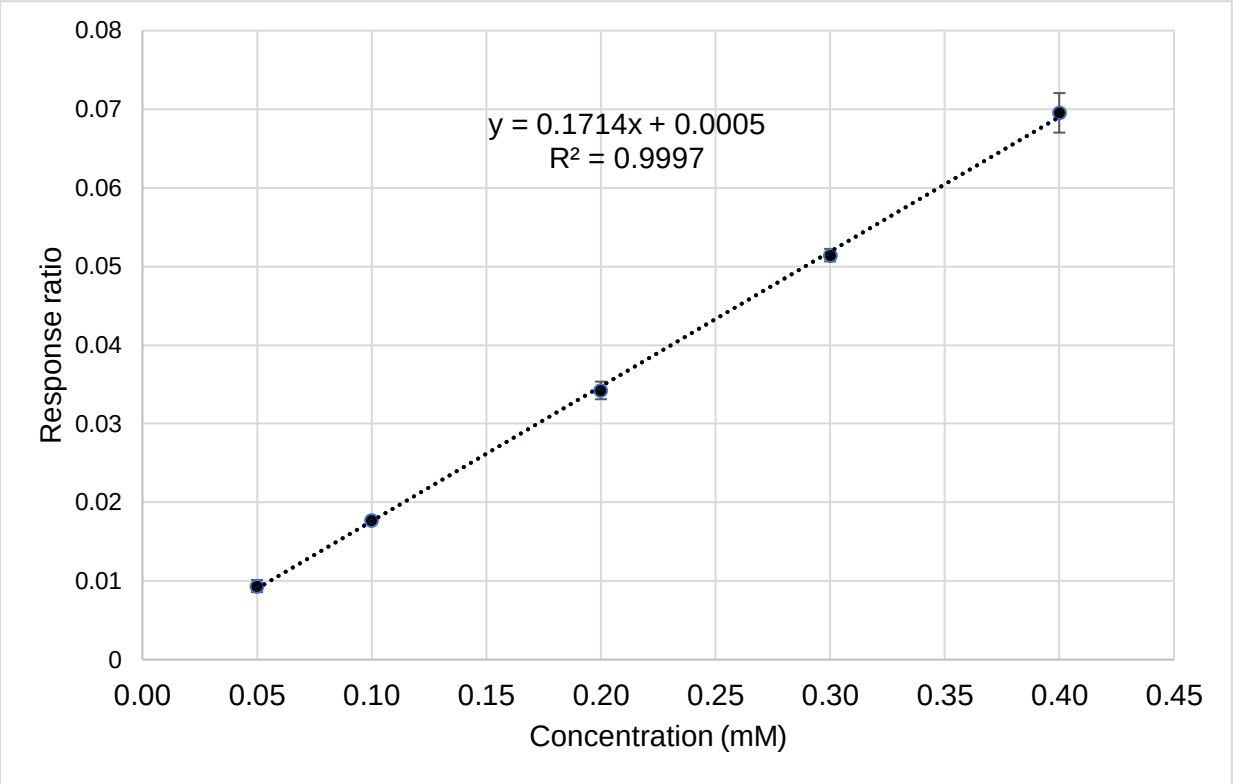


Figure S2. Octanol standard curve of 50 to 400 µM calculated from the response ratio of octanol to decanal. Standards were prepared by diluting octanol working stocks to concentrations of 50, 100, 200, 300 and 400 µM in ethyl acetate. Error bars represent standard deviation (n=3).

### 7.1.3 Sulfite calibration curve

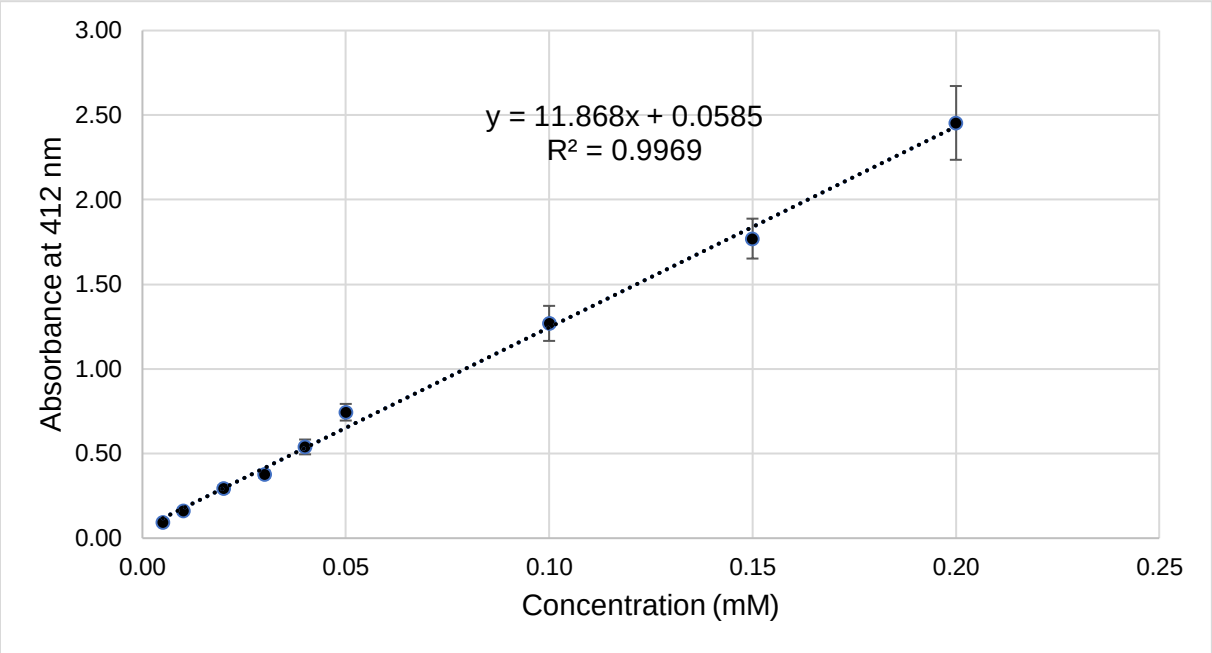


Figure S3. Sulfite standard curve of 5 to 200 µM. Standards were prepared by reacting $N_2$-purged sodium sulfite at concentrations of 5, 10, 20, 30, 40, 50, 100, 150 and 200 µM with excess DTNB and measured at 412 nm. Error bars represent standard deviation (n=3)

When predicted with the calibration curve, MBPSsuD and SsuDH produced 31.3 +/- 1.54 and 51.62 +/- 8.93 µM of sulfite when challenged with octane sulfonate and 25.27 +/- 5.73 and 33.35 +/- 3.11 µM when challenged with 6:2 FTSA. When predicted with a molar extinction coefficient of 14.15 $mM^{-1}cm^{-1}$, MBPSsuD and SsuDH generated 34.52 +/- 1.28 and 51.56 +/- 7.49 µM of sulfite when challenged with octane sulfonate and 29.47 +/- 4.79 and 34.52 +/- 2.60 µM when challenged with 6:2 FTSA.

## 7.2 Gas chromatography – mass spectrometry chromatograms

Analytical standards and reactions extracts were separated on a DB-5 column in line with an Agilent Technologies (Santa Clara, CA) 5977A mass selective detector and mass to charge (m/z) ratios were monitored. Peak m/z profiles were searched against the National Institute of Standards and Technology (USA) database and peak compound identity reported in percent probability.

7.2.1 Analytical standards and reaction extracts sample chromatograms

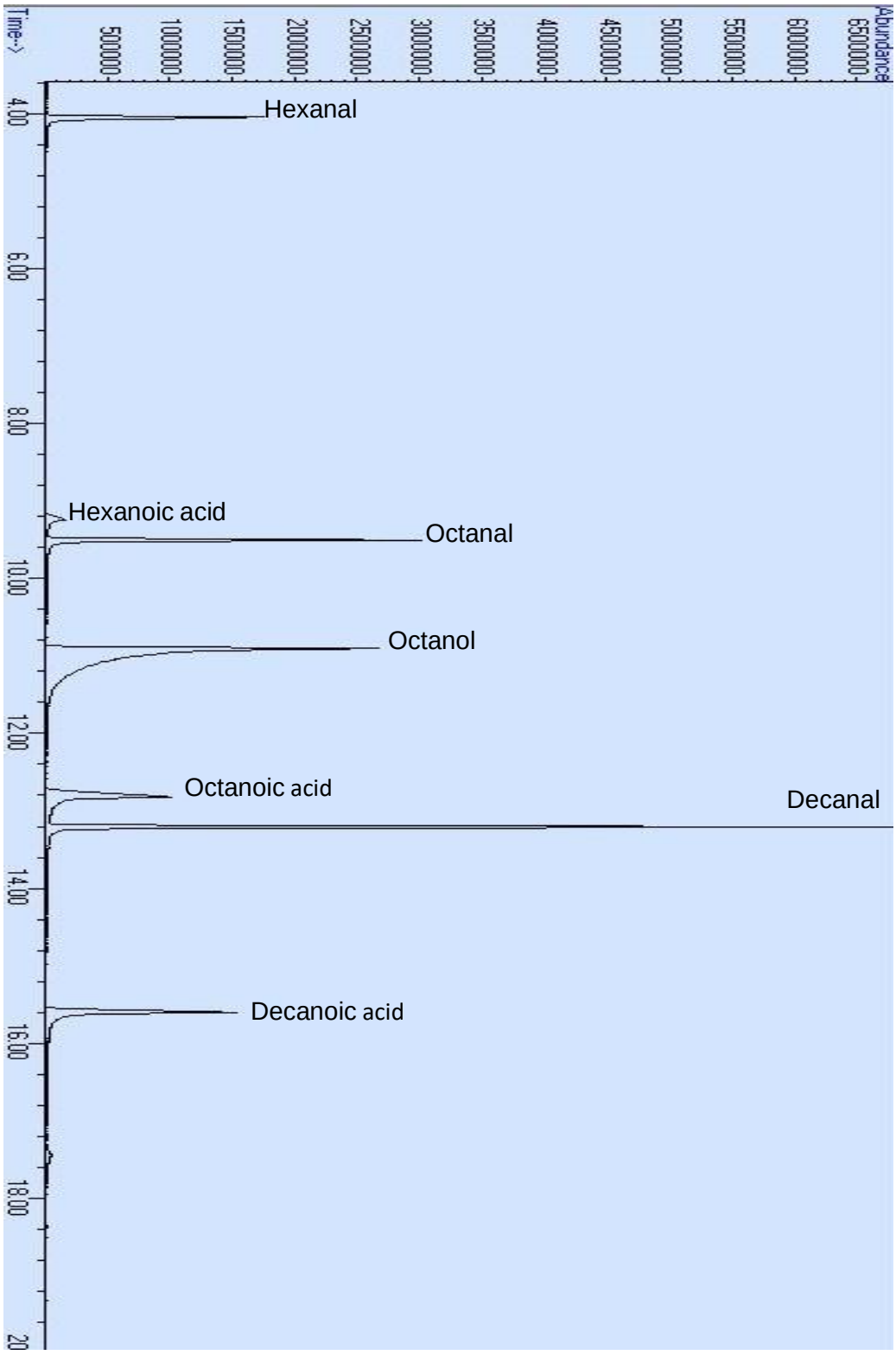


Figure S4. GC-MS chromatogram of all organic analytical standards used in this study. Percent probability of each peak when search against the National Institute of Standards and Technology (NIST) mass spectra library are as follows: hexanal – 89.1%, hexanoic acid – 85.8%, octanal – 88.5%, octanoic acid – 95.1%, octanol – 49.5%, decanal – 83.4% and decanoic acid – 92.2%.

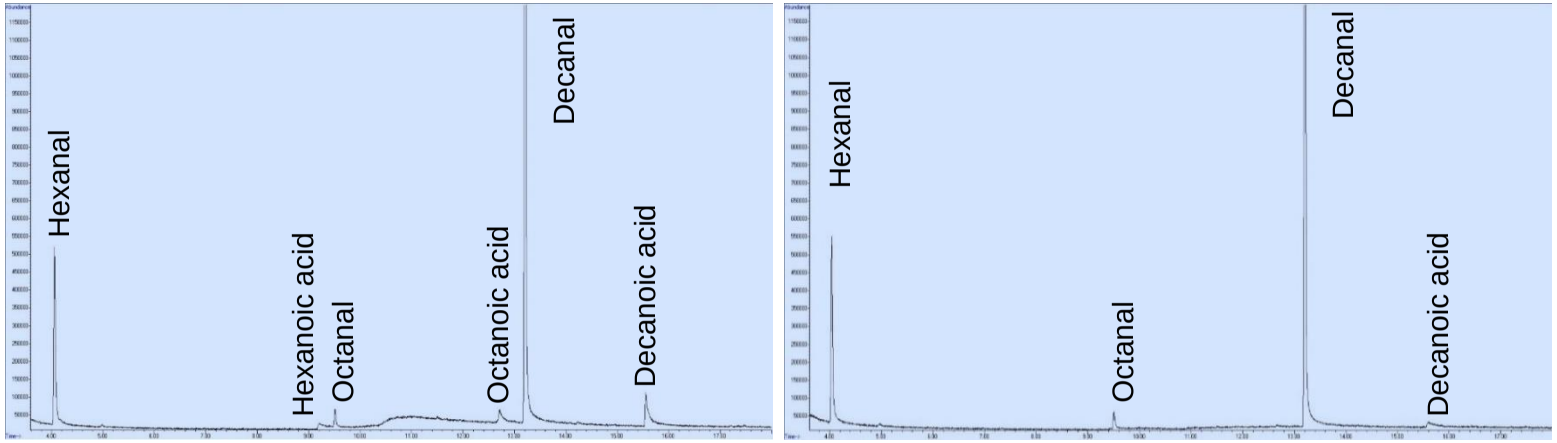


Figure S6. GC-MS chromatogram of full MBPSsuD (Left) and SsuDH (Right) reactions challenged with octane sulfonate. Percent likelihood of identified peaks are as follows: Hexanal, 87-89.3%; decanal, 77.8-82.6%; octanal 65.1-77.7%; decanoic acid, 15-70.1%; hexanoic acid, 51.1%; octanoic acid, 33.3%.

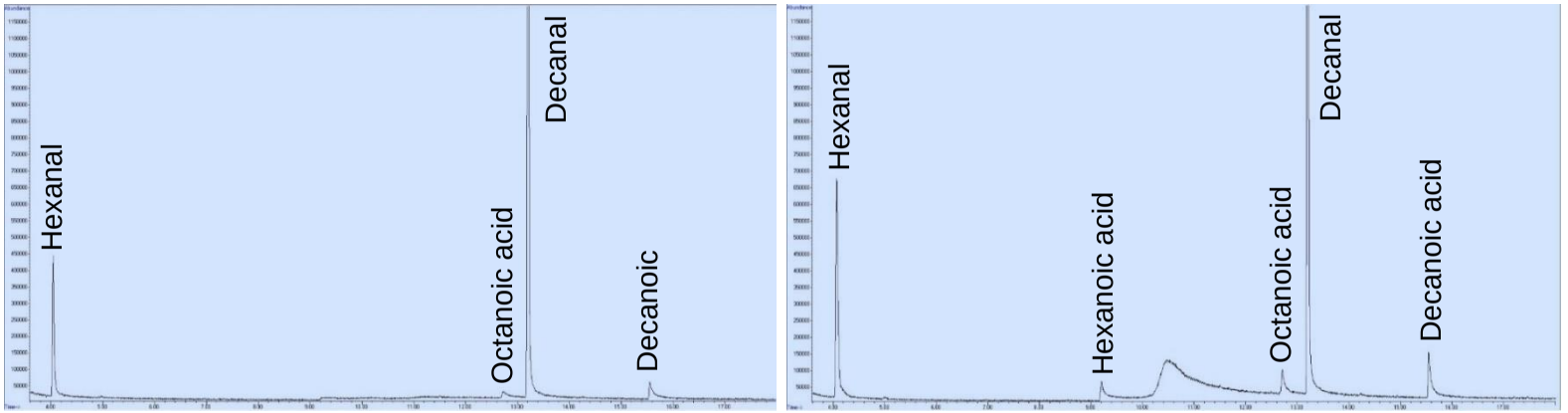


Figure S5. GC-MS chromatograms of full MBP1218 (Left) and MBP1222 (Right) reactions challenged with octane sulfonate. Percent likelihood of identified peaks are as follows: Hexanal, 87.6-88.6%**;** decanal, 78.4-79.8%**;** octanoic acid, 32.6-79.4%**;** decanoic acid, 67.1-81.9%**;** hexanoic acid 57.9%.

### 7.2.2 Retention times of analytical standards

Table S1. Retention times of analytical standards and identified peaks by GC-MS.

| Analytical standard | Retention time |
|---|---|
| Hexanal | 4 minutes |
| Hexanoic acid | 9.25 minutes |
| Octanal | 9.5 minutes |
| Octanoic acid | 12.8 minutes |
| Octanol | 10.9 minutes |
| Decanal | 13.2 minutes |
| Decanoic acid | 15.6 minutes |

## 7.3 Gas chromatography – flame ionization detection chromatograms

### 7.3.1 Analytical standards and reaction extract sample chromatograms

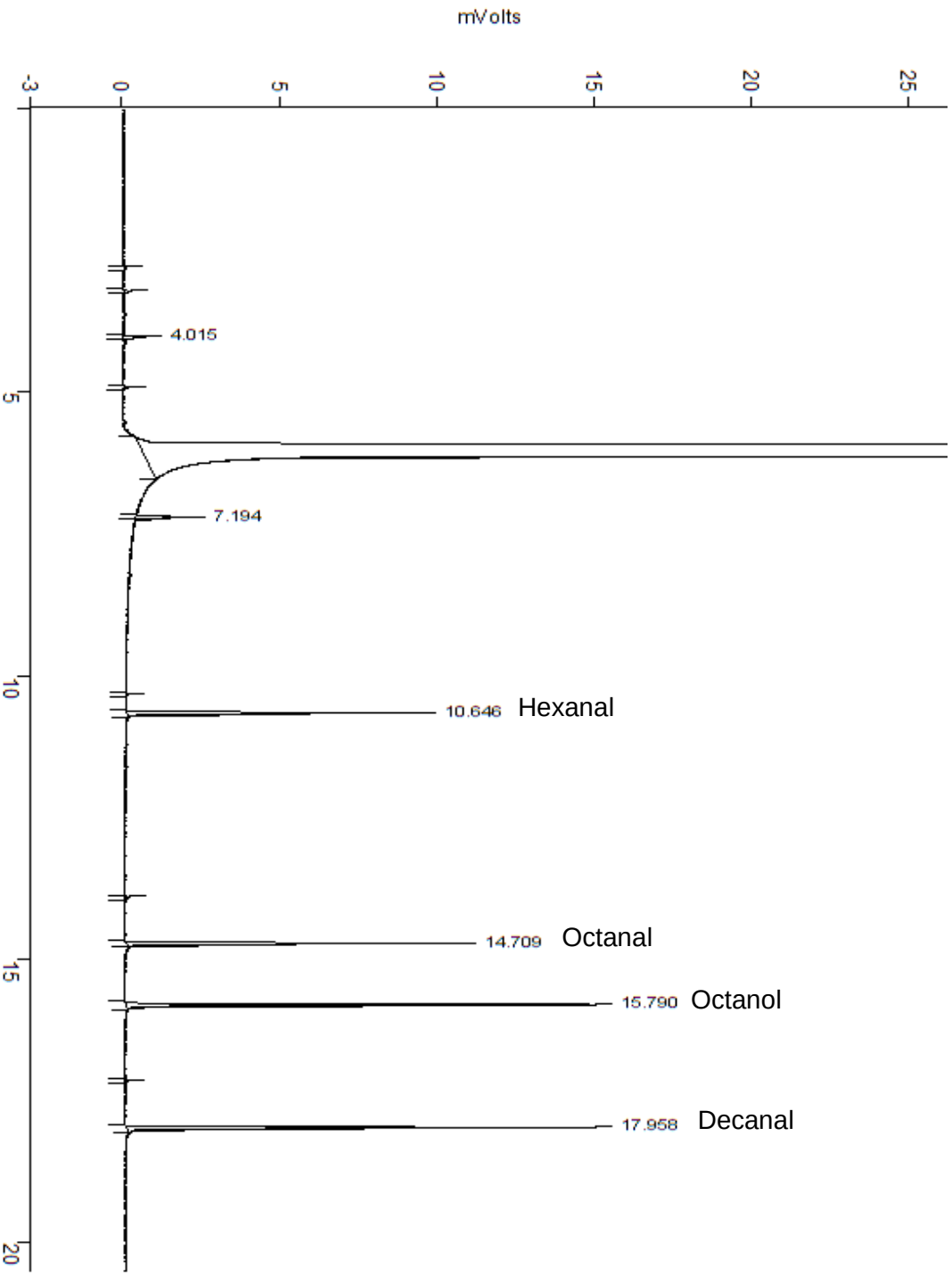


Figure S7. GC-FID chromatogram of all organic analytical standards used in this study. Analytical standards were prepared in a single vial separated on a single run. Retention times are as follows: Hexanal – 10.62, Octanal – 14.7, Octanol 15.9 and Decanal – 17.9 minutes.

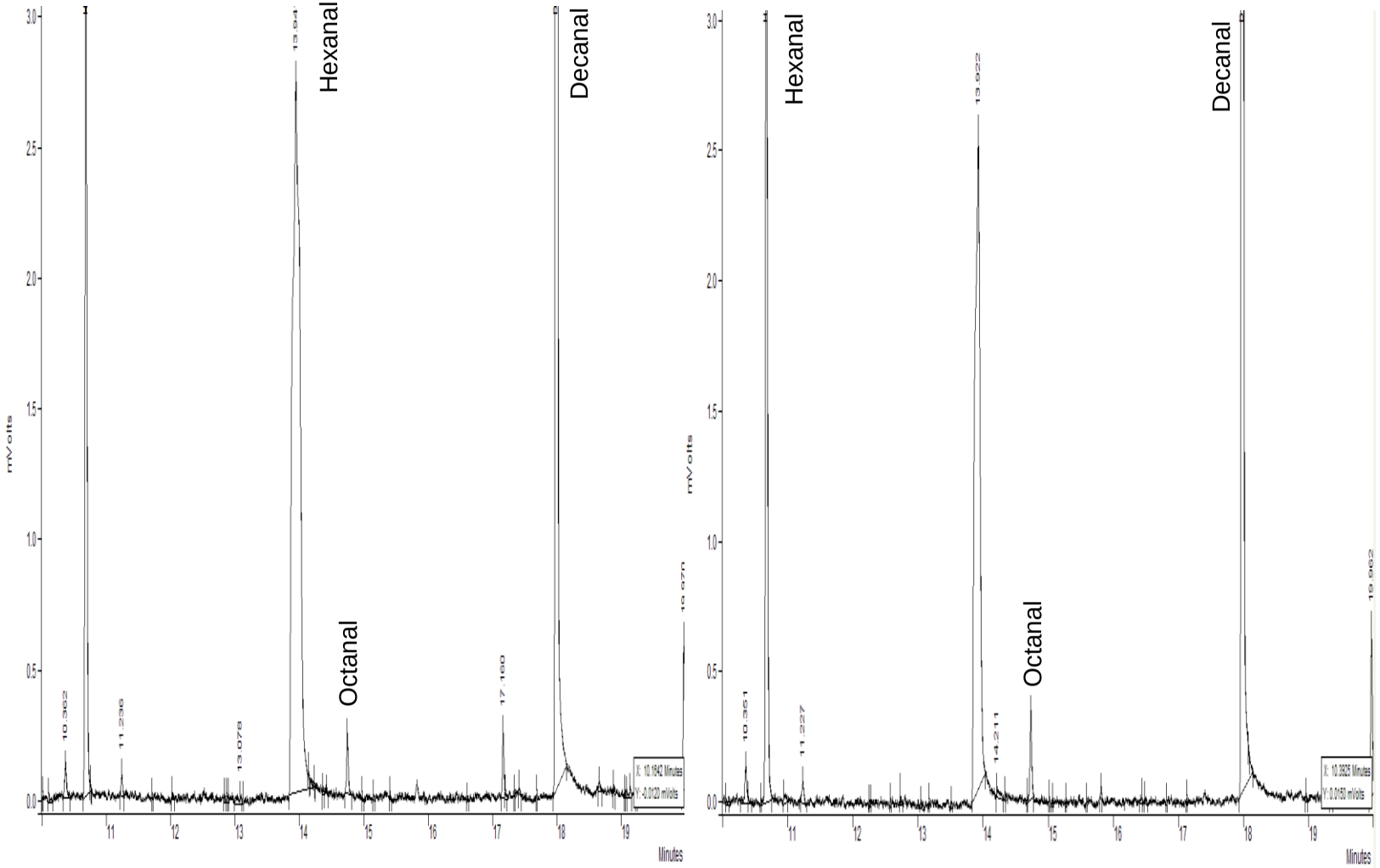


Figure S8. GC-FID chromatograms of full MBPSsuD (left) or SsuDH (right) reactions challenged with octane sulfonate. Octanal was detected at 14.75 minutes.

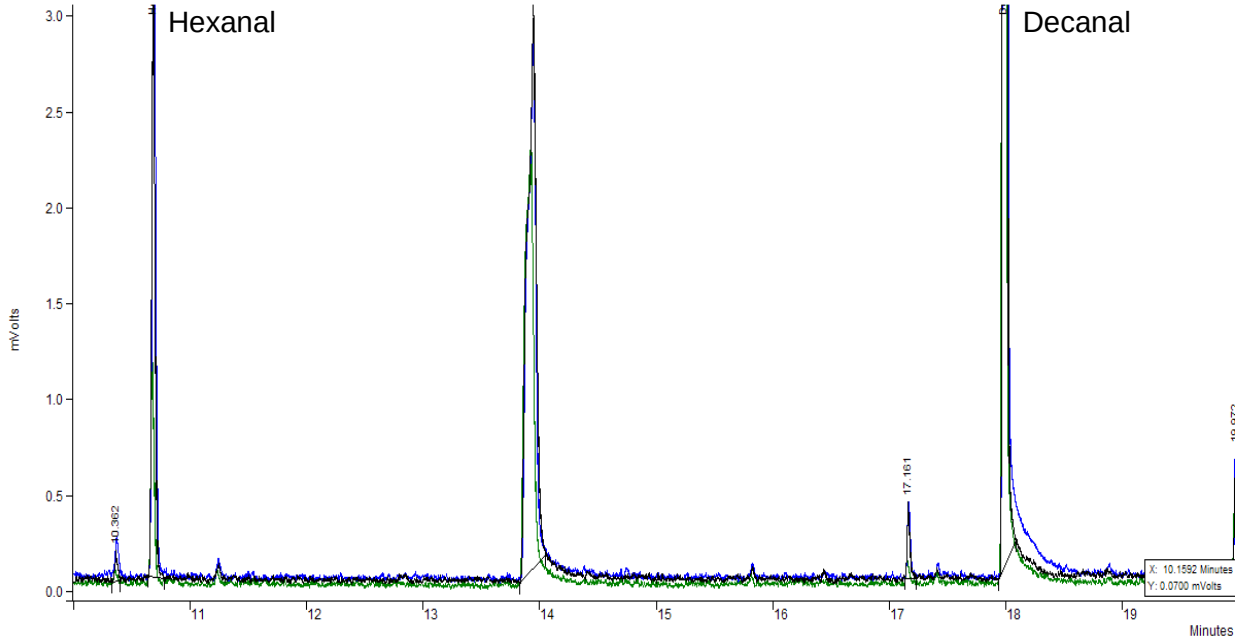


Figure S9. Overlaid GC-FID chromatograms of full MBP205 (green), MBP1666 (black) or MBP1835 (blue) reactions challenged with octane sulfonate. No octanal was detected.

### 7.3.2 Retention times of analytical standards

Table S2. Retention times of analytical standards by GC-FID.

| Analytical standard | Retention time |
|---|---|
| Hexanal | 10.6 minutes |
| Octanal | 14.7 minutes |
| Octanol | 15.8 minutes |
| Decanal | 17.9 minutes |

## 7.4 Sample SDS-PAGE

### 7.4.1 Time course protein production assay for MBP1218

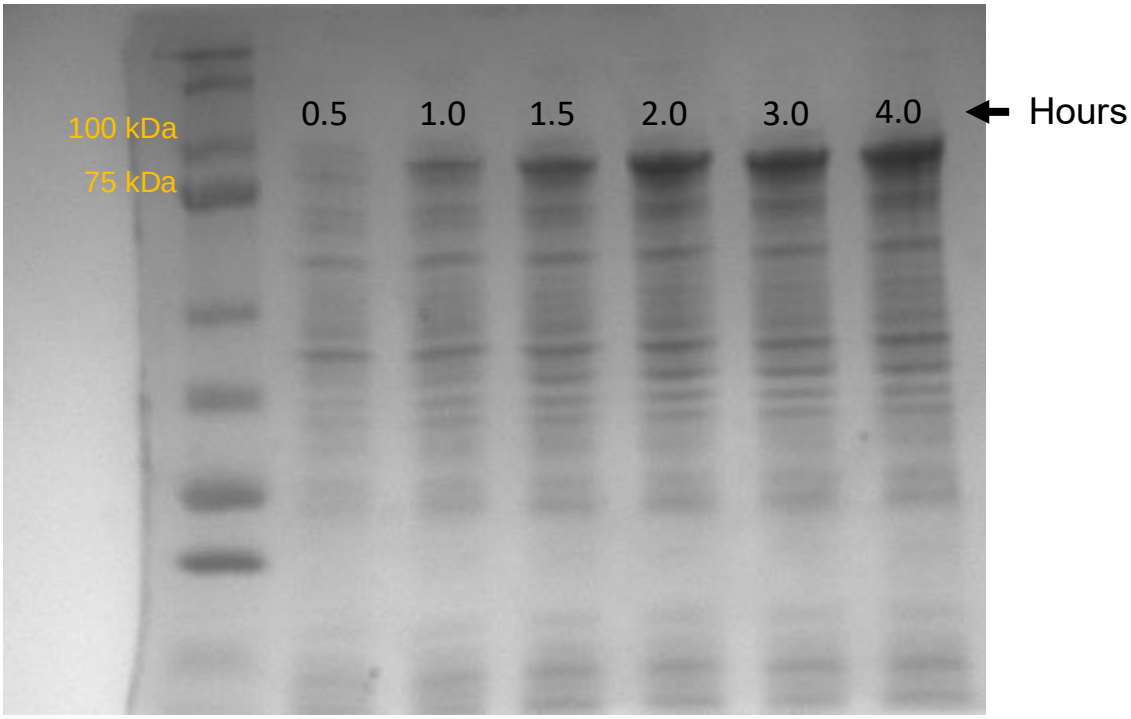


Figure S10. Time course protein production assay of pMAL1218 producing MBP1218 for up to 4 hours. *E. coli* BL21(DE3) was induced at $OD_{660}$ 0.5 with 0.3 mM of IPTG and protein production was carried out at 37°C.

7.4.1 Small scale protein production assay of MBP1218, MBP1222 and SsuD

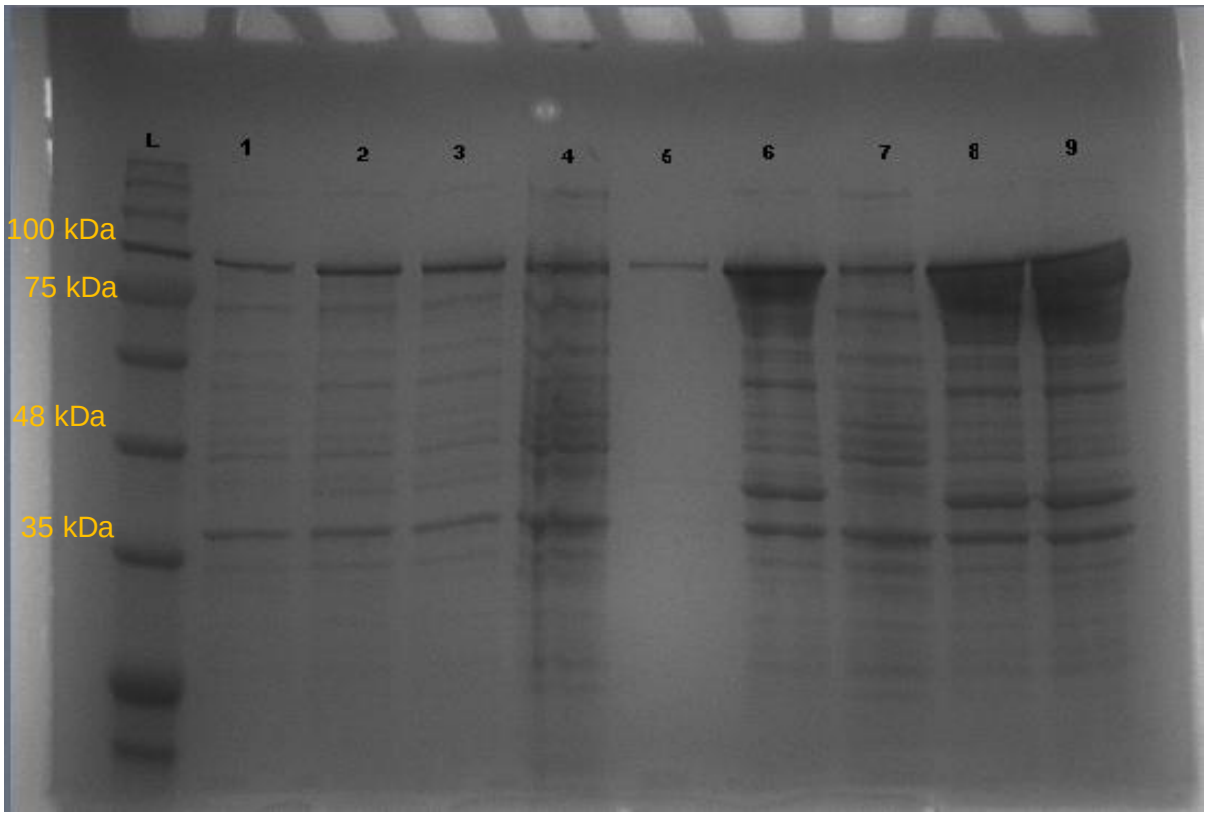


Figure S11. Small scale protein production assay of MBP1218. Labels are as follows: **L** – BLUelf prestained protein ladder (FroggaBio, Toronto, ON), **1**-**3** – pMAL1218 induced with 0, 0.3 and 0.6 mM respectively IPTG and incubated at 18°C, **4-6 –** pMAL1218 induced with 0, 0.3 and 0.6 mM respectively IPTG and incubated at 30°C and **7-9 –** pMAL1218 induced with 0, 0.3 and 0.6 mM respectively IPTG and incubated at 37°C

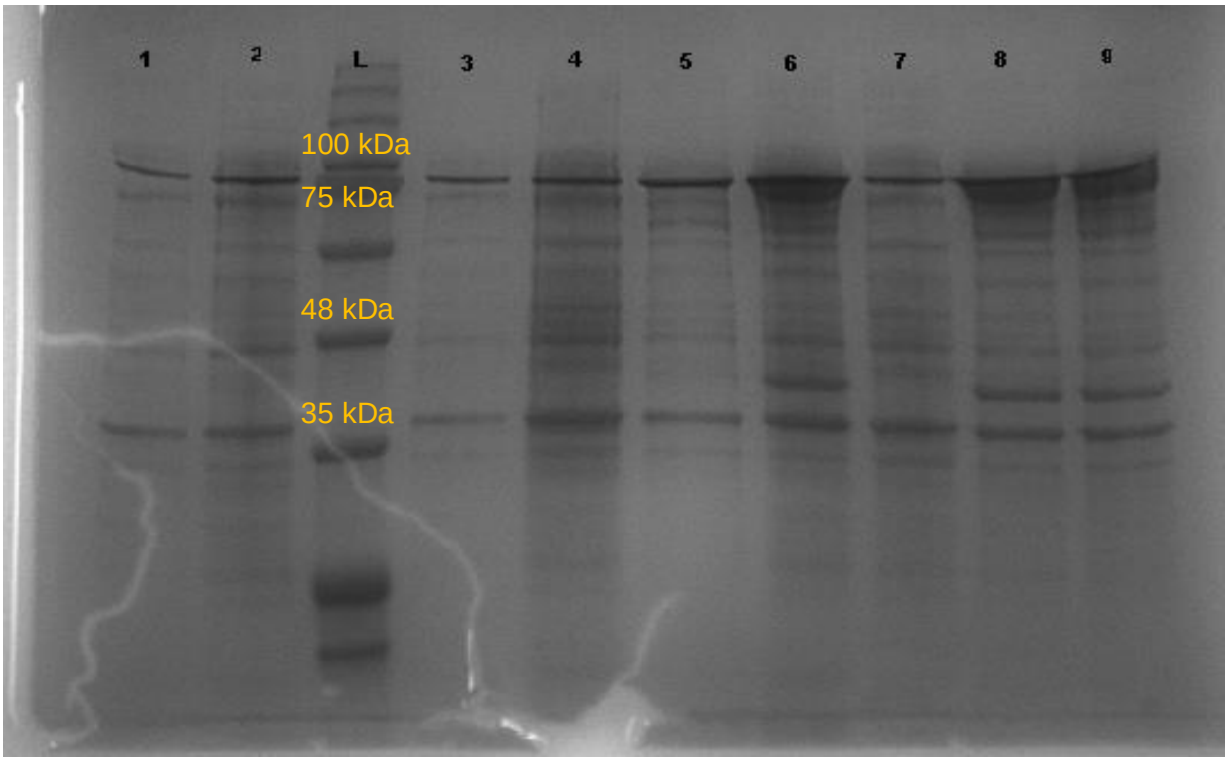


Figure S12. Small scale protein production assay of MBP1222. Labels are as follows: **L** – BLUelf prestained protein ladder (FroggaBio, Toronto, ON), **1**-**3** – pMAL1222 induced with 0, 0.3 and 0.6 mM respectively IPTG and incubated at 18°C, **4-6 –** pMAL1222 induced with 0, 0.3 and 0.6 mM respectively IPTG and incubated at 30°C and **7-9 –** pMAL1222 induced with 0, 0.3 and 0.6 mM respectively IPTG and incubated at 37°C.

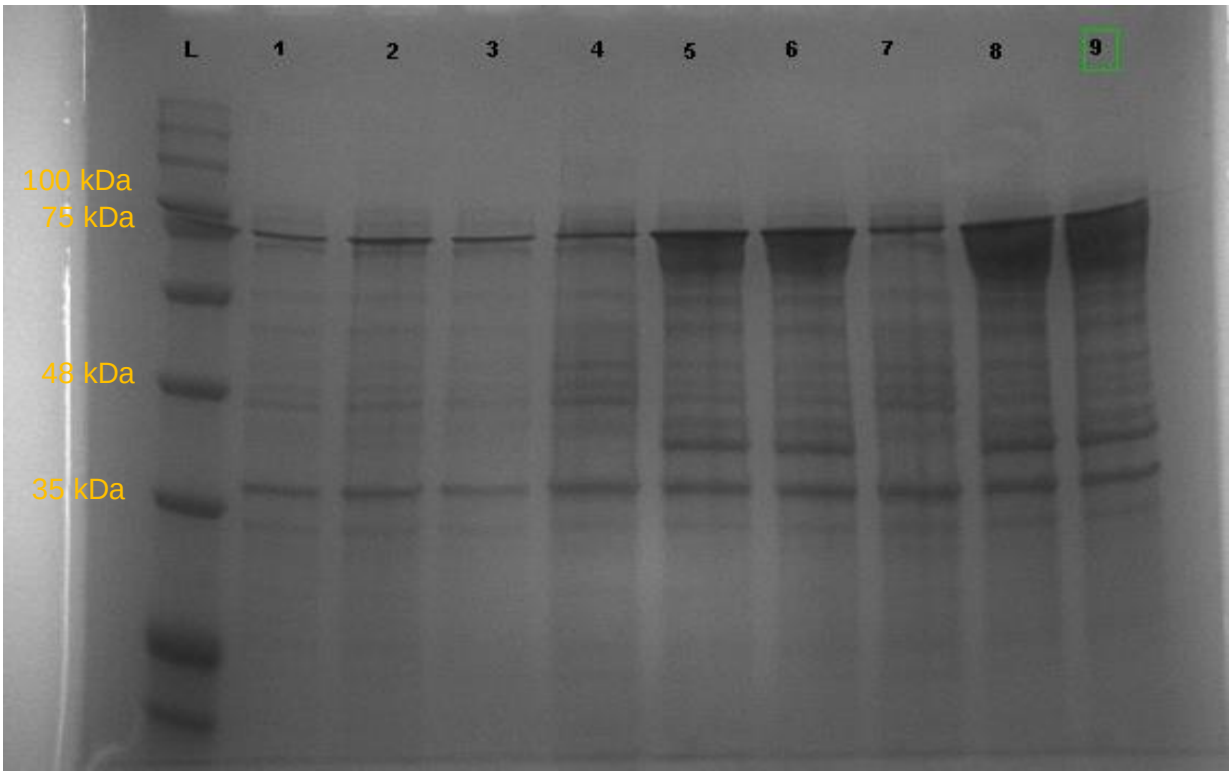


Figure S13. Small scale protein production assay of MBPSsuD. Labels are as follows: **L** – BLUelf prestained protein ladder (FroggaBio, Toronto, ON), **1-3** – pMALSsuD induced with 0, 0.3 and 0.6 mM respectively IPTG and incubated at 18°C, **4-6 –** pMALSsuD induced with 0, 0.3 and 0.6 mM respectively IPTG and incubated at 30°C and **7-9 –** pMALSsuD induced with 0, 0.3 and 0.6 mM respectively IPTG and incubated at 37°C.

7.4.2 Size exclusion chromatography peak identity

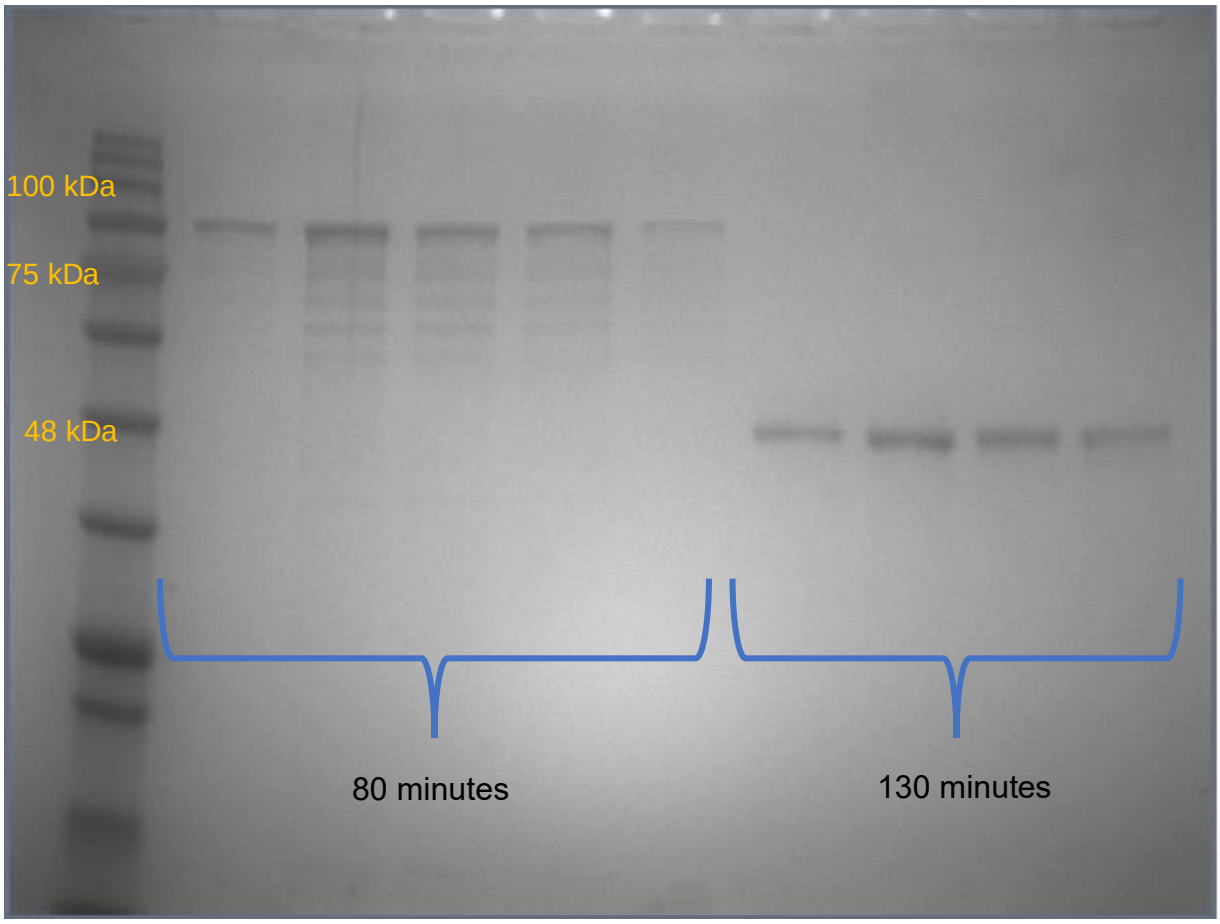


Figure S14. Size exclusion chromatography of MBP1218. The left most bracketed fractions appeared as the first peak around 80 minutes and the right most bracketed fractions appeared as the second peak near 130 minutes post column application on the UV chromatogram.

7.4.3 Washed versus unwashed nickel resin elution profile

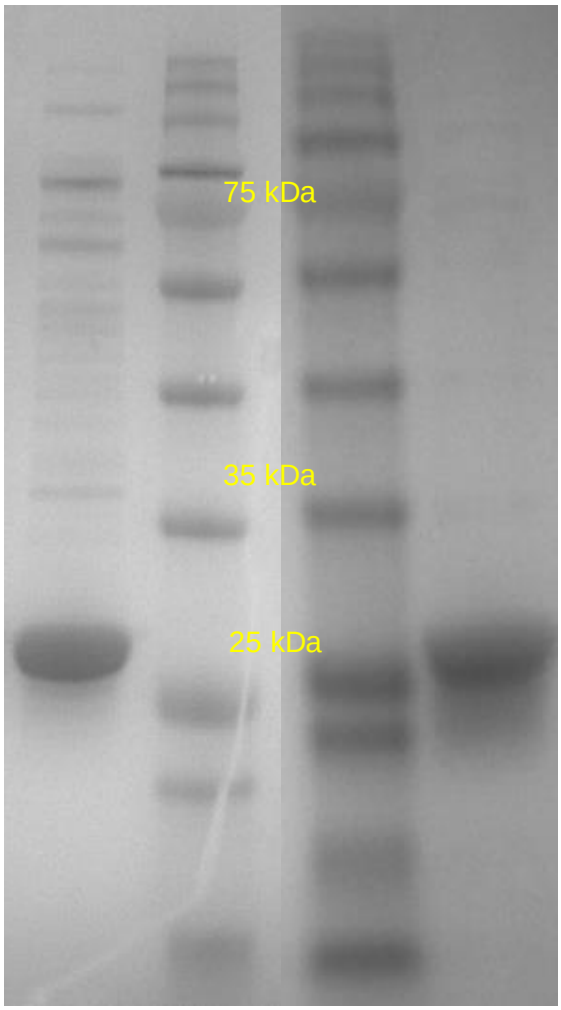


Figure S15. Comparison of the of imidazole washes on FreH purity. Left, FreH containing lysate was applied with binding buffer containing 10 mM imidazole and was washed with 3-5 column volumes of 20 followed by 40 mM of imidazole prior to elution with 500 mM imidazole. Right, Fre containing lysate was applied with binding buffer containing 20 mM imidazole and was washed with 3 column volumes of 40 followed by 80 mM of imidazole prior to elution with 500 mM imidazole.

## 7.5 Sample UV chromatograms

### 7.5.1 Sample UV chromatogram of MBP tagged protein application and elution

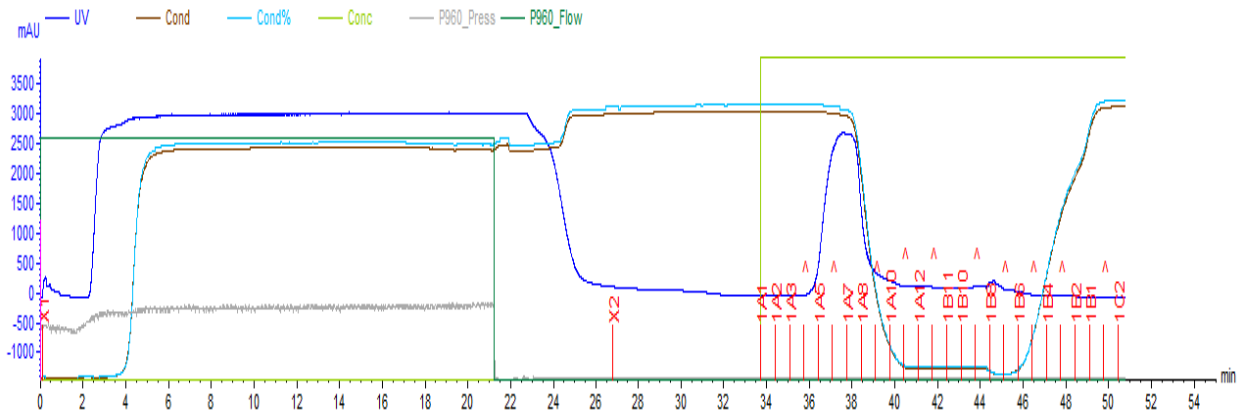


Figure S16. Purification by amylose resin of MBP1218. Lysate containing MBP1218 was applied with a sample P960 sample pump (green trace) and the column washed to baseline. UV signal (blue trace) was monitored following a 10 mM isocratic gradient of binding buffer with maltose (yellow trace) and fractions containing protein were collected (red trace A5-A10).

### 7.5.2 Sample UV chromatogram of size exclusion chromatography

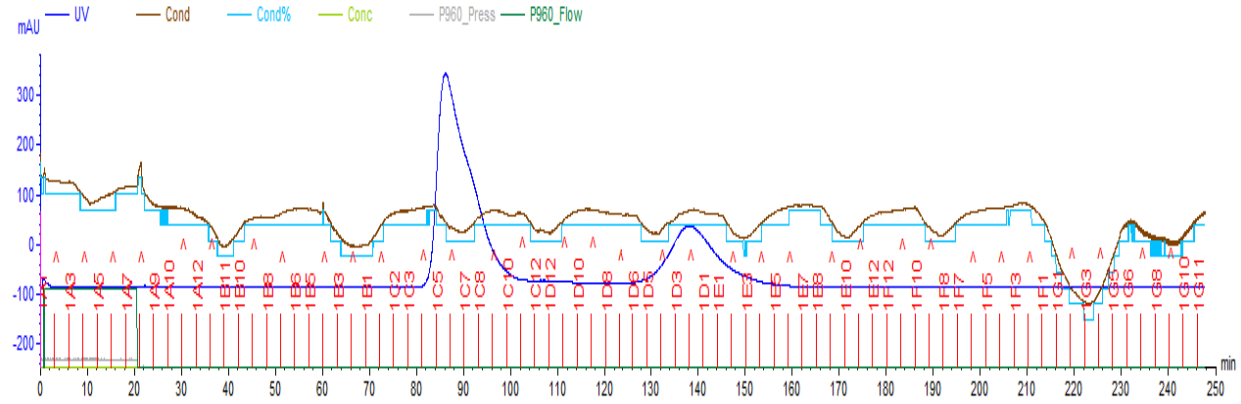


Figure S17. Size exclusion chromatography UV chromatogram of MBP1218. Protein to be separated was added at time 0 minutes with a P960 sample pump and UV signal (blue trace) was monitored every fraction (red bars) for four hours.

7.5.3 Sample UV chromatogram of His tagged protein application and elution

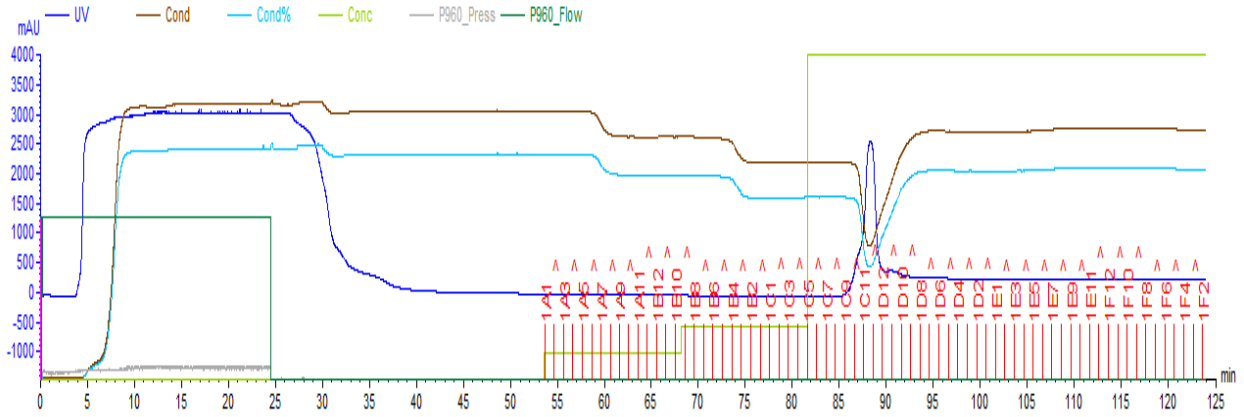


Figure S18. Purification by nickel resin of FreH. Lysate containing FreH was applied with a sample P960 sample pump (green trace) and the column washed to baseline. UV signal (blue trace) was monitored following a 40 mM (first step – yellow trace), 80 mM (second step – yellow trace) and 500 mM (third step – yellow trace) isocratic gradient of imidazole and fractions containing protein were collected (red trace C7-D10).

## 7.6 Statistical analysis of raw data and means

### 7.6.1 Statistical analysis of sulfite quantification from octane sulfonate challenged reactions

Table S3. Ryan-Joiner normality test of sulfite quantified from octane sulfonate challenged reactions

| Reaction | p-value | Ryan-Joiner value |
|---|---|---|
| SsuDH | >0.1 | 0.998 |
| SsuDH NC | N/D | N/D |
| MBPSsuD | >0.1 | 0.996 |
| MBPSsuD NC | N/D | N/D |
| MBP1218 | >0.1 | 0.894 |
| MBP1218 NC | N/D | N/D |
| MBP1222 | >0.1 | 0.995 |
| MBP1222 NC | N/D | N/D |

Table S4. Levene's two sample variance, two sample T-test and Welch T-test of sulfite quantified from octane sulfonate challenged reactions

| Reaction pair | Levene's p-value | T-test p-value | Welch T-test p-value |
|---|---|---|---|
| SsuDH/SsuDH NC | N/D | N/D | N/D |
| MBPSsuD/MBPSsuD NC | N/D | N/D | N/D |
| MBP1218/MBP1218 NC | N/D | N/D | N/D |
| MBP1222/MBP1222 NC | N/D | N/D | N/D |

### 7.6.2 Statistical analysis of 6:2 FTSA quantified from 6:2 FTSA challenged reactions

Table S5. Ryan-Joiner normality test of sulfite quantified 6:2 FTSA challenged reactions

| Reaction | p-value | Ryan-Joiner value |
|---|---|---|
| SsuDH | 0.038[1] | 0.875 |
| SsuDH NC | >0.1 | 0.931 |
| MBPSsuD | >0.1 | 0.990 |
| MBPSsuD NC | >0.1 | 0.997 |
| MBP1218 | >0.1 | 0.949 |
| MBP1218 NC | >0.1 | 0.935 |
| MBP1222 | >0.1 | 0.906 |
| MBP1222 NC | >0.1 | 1.000 |

[1]The p-value for SsuDH was found to be marginally close to the 0.05 cutoff and used in the subsequent Levene's test and T-test. T-tests performed on non-normally distributed data tend to increase type-1 error (false positive) and for this population specifically, the data was assumed to be normally distributed.

Table S6. Levene's two sample variance, two sample T-test and Welch T-test of 6:2 FTSA quantified from 6:2 FTSA challenged reactions

| Reaction pair | Levene's p-value | T-test p-value | Welch T-test p-value |
|---|---|---|---|
| SsuDH/SsuDH NC | 0.337 | 0.012 | N/D |
| MBPSsuD/MBPSsuD NC | 0.251 | 0.073 | N/D |
| MBP1218/MBP1218 NC | 0.418 | 0.883 | N/D |
| MBP1222/MBP1222 NC | 0.151 | 0.850 | N/D |

### 7.6.3 Statistical analysis of sulfite quantification from 6:2 FTSA challenged reactions

Table S7. Ryan-Joiner normality test of sulfite quantified from 6:2 FTSA-based reactions

| Reaction | p-value | Ryan-Joiner value |
|---|---|---|
| SsuDH | >0.1 | 0.997 |
| SsuDH NC | N/D | N/D |
| MBPSsuD | >0.1 | 0.996 |
| MBPSsuD NC | N/D | N/D |
| MBP1218 | >0.1 | 0.894 |
| MBP1218 NC | N/D | N/D |
| MBP1222 | >0.1 | 0.979 |
| MBP1222 NC | >0.1 | 0.994 |

Table S8. Levene's two sample variance, two sample T-test and Welch T-test of sulfite quantified from 6:2 FTSA-based reactions

| Reaction pair | Levene's p-value | T-test p-value | Welch T-test p-value |
|---|---|---|---|
| SsuDH/SsuDH NC | N/D | N/D | N/D |
| MBPSsuD/MBPSsuD NC | N/D | N/D | N/D |
| MBP1218/MBP1218 NC | N/D | N/D | N/D |
| MBP1222/MBP1222 NC | 0.917 | 0.325 | N/D |

### 7.6.4 Statistical analysis of $OD_{660}$ readings

Table S9. Ryan-Joiner normality test of $OD_{660}$ under different growth conditions in an oxygen permissive environment.

| Growth condition | p-value | Ryan-Joiner value |
|---|---|---|
| No Sulfur | >0.1 | 0.962 |
| $MgSO_4$ | >0.1 | 0.936 |
| Octane sulfonate | >0.1 | 0.940 |
| 6:2 FTSA | >0.1 | 0.932 |

Table S10. Levene's two sample variance, two sample T-test and Welch T-test of $OD_{660}$ under different growth conditions combination in an oxygen permissive environment

| Growth condition pair | Levene's p-value | T-test p-value | Welch T-test p-value |
|---|---|---|---|
| No sulfur/6:2 FTSA | 0.356 | 0.001 | N/D |
| No sulfur/Octane sulfonate | 0.022 | N/D | 0.01 |
| Octane sulfonate/6:2 FTSA | 0.041 | N/D | 0.061 |

Table S11. Ryan-Joiner normality test of $OD_{660}$ under different growth conditions in an oxygen restrictive environment.

| Growth conditions | p-value | Ryan-Joiner value |
|---|---|---|
| No Sulfur | >0.1 | 0.941 |
| $MgSO_4$ | >0.1 | 0.923 |
| Octane sulfonate | >0.1 | 0.957 |
| 6:2 FTSA | 0.095 | 0.909 |

Table S12. Levene's two sample variance, two sample T-test and Welch T-test of $OD_{660}$ under different growth conditions combination in an oxygen restrictive environment

| Growth condition pair | Levene's p-value | T-test p-value | Welch T-test p-value |
|---|---|---|---|
| No sulfur/6:2 FTSA | 0.724 | <0.000 | N/D |
| No sulfur/Octane sulfonate | 0.685 | <0.000 | N/D |
| Octane sulfonate/6:2 FTSA | 0.880 | 0.576 | N/D |

## 7.7 Sanger sequencing results and *in silico* constructed protein production vectors

Below are the Sanger sequencing results of protein production and mutagenesis vectors amplified with gene amplification primers, sequencing primers or custom primers. Protein production and mutagenesis vectors assembled *in silico* were reported as 1200 base pairs up or downstream of the gene to be produced or AB fragment. Nucleotides reported as N are any nucleotide.

Sequence data is provided in FASTA format with the title following a 'fragment/gene'_'primer'_'direction' paradigm. For example, the sanger sequence data produced by the pET28b1218 vector and T7T primer would be reported as pET28b1218_T7T_reverse complement. Constructs produced *in silico* are denoted with the plasmid naming format. For example, pET28b1218 is the *in silico* construct of ISGA 1218 inserted into the pET28b vector.

Reverse complements were produced using the Reverse Complement calculator from Bioinformatics.org and reported for directionality congruence between *in silico* constructs and sanger sequencing data. Bolded characters represent enzyme amino acid start and stop sites and underlined characters represent vector protein production start and stop site.

>pMAL1218

GCACTTCACCAACAAGGACCATAGATTATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAA
GGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCC
GGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACG
ACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTG
TATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCG
CTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACT
GAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTG
ACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGC
GAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGC
AGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCA
GCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTG
AGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGA
TGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTG
GCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGAT
GTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCC
CTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGGATCGAGGG
AAGGATTTCAGAATTC**ATG**AACGTAAACGTTGTTGGCGGAATCTCTCAATGGGACACTGTTCAGGCCGATGCGT
GGTGGTCGGCTGCAGTCGGTCCAGCAACGAACGGAGAGATCTCGATGTCGGCACATGGGCGTCCAATTCACTT
GGGCGGGTTTTTGATTGCAGGGAATGTAACCCACAGTCATCCTTCGTGGCGTCATCCACGCAGTGATCCCGGGT
TTCTCACACCGGAGTACTACCAGCACCTCGGTAGGGTTTTCGAACGCGCGAAATTGGACTTCGTCTTCTTTGCC

GACAATTCTGCGACTCCTGCCAGCTACCGCAACGATATTCGTGACCCGCTCGCTCGCGGTACTCAGAGTGCAGC
CGGGTTGGATCCCCGCTTCGTCGTTCCTGTCGTCGCGGGTGTCACGCGCAACCTGGGGATCGTATCGACCACG
TCGGCGACGTTCTACTCGCCATACGACCTCGCCCGGAGCTTTGCCACTCTGGATCATCTGACCCACGGCCGCG
TCGGTTGGAATGTTGTGACTTCCAATACGACCGTCGAGGCGCAGAACTTCGGGCTTGCCCGACACCTCGACCAT
GACGTGCGATACGACCGTGCCGAGGAATTGCTTGAGGTCGCGTTCAGGCTGTGGGCCAGTTGGGACGATGGA
GCTCTGATCCAGGATAAAGAGGCGGGTGTCTTCGCCGACCCGGACCTGATTCACAGGCTCGATCATCACGGGG
AGAACTTCGATGTTCGGGGCCCGCTGTCGGTTCCCCGCTCACCGCAGGGACGGCCGGTCATCTTTCAAGCGGG
ATCATCCACCCGCGGTCGGGATTTTGCTGCGCGCTGGGCAGAAGCGATTTTCGAGATCGACCCGACGTCTGTG
GGGCGTAAGGCCTACTACGACGACATCAAGTCGCGAGCCTCCGACTTCGGTCGTGATCCCGACGGCGTCAAGA
TACTCCCGTCGTTCATTCCGTTTGTGGGTGAGACCGAGTCGATCGCACGGGAAAAGCAGGCGTTCCACAACGAA
CTGGCCGATCCGACCGATGGATTGATCACGCTGTCGGTGCACACCGACCATGATTTCTCCGGCTATGACCTCGA
CGCTGTGATCGCCGACATCGATGTTCCAGGGACGAAGGGGCTTTTCGAAGTCGCTCGGAGTCTGAGTGTGAAC
GAGAACCTGACGCTGCGCGATATCGGAAAGCTGTACGCCCAGGGCGTGTTATTGCCGCAGTTCGTGGGTACCG
CGGCTCAGGTGGCCGACCAGATCGAGGCTGCCGTCGACGGTGGAGAGGCTGATGGGTTCCTCTTTTCGGCCG
GGTATACGCCTGGCGGATTCGAGGAGTTCGCCGATCTCGTCATCCCGGAACTGCAGCGGCGAGGGCGGTTTCG
TACGGAGTACACGGGTTCGACGCTGCGTGAACATCTGGGTCTACCCGCTGATGCGAATCTTGTGCCCGTTCCGC
GCAAGGCAGTGGGGGCGGCG**TGA**AAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTG
GCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACC
GATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCAGCTTGGCTGTTTTGGCGGATGAGATAAGAT
TTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCG
CGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCC
CCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTT
ATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCA
ACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCT
GACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCAT
GAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCC
CTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTG
AAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGC
CCCGAAGAACGTTCTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCC
GGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAA
GCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCA
ACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACT
CGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGC

>pMAL1218_21M13_reverse complement

NNTTNNNNGANTCCNNGCCCAGCTACNGCAANGATNNTTCGNNGNCCCGCTNCNCTCNNCGGTANTCAGAGTG
CAGCNGGNNNNNCCCCGNTNGTCGTTCCTGTCGTCNCGGGTGTCACGCGCANNNGGGGATCGTATCGNCCAC
NTCGGCGACGTTCTACTCNCCATACGACCTCGCCCGGGAGCTTTGCCANTCTGGATCATCTGACCCACGGCCG
CGTCGGTTGGAATGTTGTGACTNCCAATACGACCGTCGAGGCGCAGAACTTCGGGCTTGCCCGACACCTCGAC
CATGACGTGCGATACGACCGTGCCGAGGAATTGCTTGAGGTCGCGTTCAGGCTGTGGGCCAGTTGGGACGATG
GAGCTCTGATCCAGGATAAAGAGGCGGGTGTCTTCGCCGACCCGGACNTGATTCACAGGCTCGATCATCACGG
GGAGAACTTCGATGTTCGGGGCCCGCTGTCGGTTCCCCGCTCACCGCAGGGACGGCCGGTCATCTTTCAAGCG
GGATCATCCACCCGCGGTCGGGATTTTGCTGCGCGCTGGGCAGAAGCGATTTTCGAGATCGACCCGACGTCTG
TGGGGCGTAAGGCCTACTACGACGACATCAAGTCGCGAGCCTCCGACTTCGGTCGTGATCCCGACGGCGTCAA
GATACTCCCGTCGTTCATTCCGTTTGTGGGTGAGACCGAGTCGATCGCACGGGAAAAGCAGGCGTTCCACAAC
GAACTGGCCGATCCGACCGATGGATTGATCACGCTGTCGGTGCACACCGACCATGATTTCTCCGGCTATGACCT
CGACGCTGTGATCGCCGACATCGATGTTCCAGGGACGAAGGGGCTTTTCGAAGTCGCTCGGAGTCTGAGTGTG
AACGAGAACCTGACGCTGCGCGATATCGGAAAGCTGTACGCCCAGGGCGTGTTATTGCCGCAGTTCGTGGGTA
CCGCGGCTCAGGTGGCCGACCAGATCGAGGCTGCCGTCGACGGTGGAGAGGCTGATGGGTTCCTCTTTTCGG
CCGGGTATACGCCTGGCGGATTCGAGGAGTTCGCCGATCTCGTCATCCCGGAACTGCAGCGGCGAGGGCGGT
TTCGTACGGAGTACACGGGTTCGACGCTGCGTGAACATCTGGGTCTACCCGCTGATGCGAATCTTGTGCCCGTT
CCGCGCAAGGCAGTGGGGCGGCNNANNANNNNNNNNN

>pMAL1218_MBP-F

NNNNNNNNTNNNNNNNCACATAACAATAACAACAACCTCGGGATCGAGGGAAGGATTTCAGAATTCATGAACGT
AAACGTTGTTGGCGGAATCTCTCAATGGGACACTGTTCAGGCCGATGCGTGGTGGTCGGCTGCAGTCGGTCCA
GCAACGAACGGAGAGATCTCGATGTCGGCACATGGGCGTCCAATTCACTTGGGCGGGTTTTTGATTGCAGGGA
ATGTAACCCACAGTCATCCTTCGTGGCGTCATCCACGCAGTGATCCCGGGTTTCTCACACCGGAGTACTACCAG

CACCTCGGTAGGGTTTTCGAACGCGCGAAATTGGACTTCGTCTTCTTTGCCGACAATTCTGCGACTCCTGCCAG
CTACCGCAACGATATTCGTGACCCGCTCGCTCGCGGTACTCAGAGTGCAGCCGGGTTGGATCCCCGCTTCGTC
GTTCCTGTCGTCGCGGGTGTCACGCGCAACCTGGGGATCGTATCGACCACGTCGGCGACGTTCTACTCGCCAT
ACGACCTCGCCCGGAGCTTTGCCACTCTGGATCATCTGACCCACGGCCGCGTCGGTTGGAATGTTGTGACTTCC
AATACGACCGTCGAGGCGCAGAACTTCGGGCTTGCCCGACACCTCGACCATGACGTGCGATACGACCGTGCCG
AGGAATTGCTTGAGGTCGCGTTCAGGCTGTGGGCCAGTTGGGACGATGGAGCTCTGATCCAGGATAAAGAGGC
GGGTGTCTTCGCCGACCCGGACCTGATTCACAGGCTCGATCATCACGGGGAGAACTTCGATGTTCGGGGCCCG
CTGTCGGTTCCCCGCTCACCGCANGGACGGCCGGTCATCTTTCAAGCGGGATCATCCACCCGCGGTCGGGATT
TTGCTGCGCGCTGGGCAGAAGCGATTTTCGAGATCGACCCGACGTCTGTGGGGCGTAAGNCCTACTACGACGA
CATCAAGTCGCGAGCCTCCGACTTCGGTCGTGATCCCGACGGCGTCAAGATACTCCCGTCGTTCATTCCGTTTG
TGGGTGAGACCGANTCGATCGCACGGGAAAAGCAGGNNTTCCACACGAANTGGNNGATCCGACCGATGGATTG
ATCACGCTGTCGNGCACNNGACNTGANTTCTNNNCTATGACNTCNACGCTNNGATCNCNANNTCNATGNTNCAG
GGNCNNNNGGNNTTTTCNNN

>pMAL1222

GCACTTCACCAACAAGGACCATAGATTATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAA
GGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCC
GGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACG
ACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTG
TATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCG
CTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACT
GAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTG
ACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGC
GAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGC
AGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCA
GCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTG
AGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGA
TGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTG
GCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGAT
GTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCC
CTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGGATCGAGGG
AAGGATTTCAGAATTC**ATG**GCTGATCGAGAGCTCCATCTGGGCGTCAATGTCCTCTCGGACGGTATGCACCCAG
CCGCGTGGCAGTATCCGAGTTCCGATCCGTCGTGGTTCACGGATCCGGCGTACTGGATTCGTGTTGCGCAGAT
CGCGGAGCGAGGAACCCTCGACGCGGTCTTCCTCGCCGACAGTCCGTCGTTGTTCCAGCCGCCCGACCAGCC
GCTGAGTGCGCCACCGTTGGCCCTGGACCCGATCGTGTTGTTGTCGACACTGGCATCGGTGACCACACACATC
GGACTCATCGGTACGGTGTCGACCTCGTTCGAGGAGCCGTACAACGTCGCCCGCCGATTCTCGACGCTGGACC
ACCTCAGCCGGGGTCGTGTGGCATGGAACGTCGTGACGAGTAGTGATCGGTATGCCTGGAACAATTTCGGTGG
TGGTGAACAACCCGACCGCGCTACTCGATACGAGCGGGCCGGCGAGTTCATCGAAGTCGTCCGGGCATTGTGG
GATTCGTGGGACGACGACGCAGTTGTCGCCGACAAGTCCACGGGTGCGTTCAGTAAGGTCGGTGCGATACGAC
CGATCCGGCATCGCGGTGGGCACTTCTCGGTGGACGGGCCGTTGACTCTACCCAGATCCCCACAGGGGCATCC
GGTGTTGTTTCAGGCAGGCGGTTCCACCGGCGGGTTGGATCTGGCGGCGAAGTACGCCGACGGGGTCTTTGC
GGCACAGGCCTCGCTCGAGGATGCGCTGTCCAACGCGCAGGAGCTGCGGAGTCGGTTGATCGCGCATGGCCG
TCCCGCCGAGGCGATCCGAATCATGCCTGGCTTGTCGTTCGTGCTCGGCAGTACGGAGGCAGAGGCCAGGTCG
CGAAACGACGAATTGAACGAGCTCGCCGGGGATCGACGCCTGGCACATCTGGCTGGTCAACTCAGCGTCGATG
TGGCGGAGCTGAAGTGGGACAAGCCGCTTCCCGGTTGGCTCCTCGAGGGCGCGGCGCCGATCAGCGGTTCCC
AGGGAGCTCGCGACATCGTCGTCAACATCGCTCGGCGGGAGAACCTGACCGTGCGTCAGCTGCTCGATCGGGT
GATCACGTGGCACCGCTTCGTGGTCGGATCGCCTGAACAGATCGCCGATGCCATCGAGGACTGGTTCGTTGCG
GGCGCTGTCGACGGCTTCAACCTGATGCCGGATGTCTTCCCGTCGGGTCTCGAGTTGTTCGTCGACCACGTCG
TACCGATCCTCCGGGACCGAGGGTTGTTCCGGCGGGAGTACACATCGACGACATTGCGTGGGCATCTGGGCCT
CGAGCGCACCCCAGACCGGCCGTCGTCGGGTTCGATCCGCCGGACCGGT**TAG**AAGCTTGGCACTGGCCGTCG
TTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCA
GCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCA
GCTTGGCTGTTTTGGCGGATGAGATAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGAT
AAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGC
CGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAG
GCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCG
CCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCC

AGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTTGTTTATTTTT
CTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAG
AGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACC
CAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTC
AACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTCTCCAATGATGAGCACTTTTAAAGTTCTGCTA
TGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGA
CTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGC
CATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTT
TTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGC

>pMAL1222_21M13_reverse complement

NNNNNNTTNNTNNTTNGCCGNNNANNCNTNGTNGTTCCNNCCGCCCGNCNAGCCGCTGANNNNNGCCNCCNNN
CCCTGGNCCCGATCGNGNTNNNTCGNCACNGGCATCGGTGNCCACACACATCGGANTCATCGGTACGGNGTCG
ACCTCGTTCGAGGAGCCGTACAACGTCGCCCGCCGATTCTCGACGCTGGACCANTCAGCCGGGGTCGTGTGGC
ATGGAACGTCGTGACGAGTAGTGATCGGTATGCCTGGAACAATTTCGGTGGTGGTGAACAACCCGACCGCGCT
ACTCGATACGAGCGGGCCGGCGAGTTCATCGAAGTCGTCCGGGCATTGTGGGATTCGTGGGACGACGACGCA
GTTGTCGCCGACAAGTCCACGGGTGCGTTCAGTAAGGTCGGTGCGATACGACCGATCCGGCATCGCGGTGGG
CACTTCTCGGTGGACGGGCCGTTGACTCTACCCAGATCCCCACAGGGGCATCCGGTGTTGTTTCAGGCAGGCG
GTTCCACCGGCGGGTTGGATCTGGCGGCGAAGTACGCCGACGGGGTCTTTGCGGCACAGGCCTCGCTCGAGG
ATGCGCTGTCCAACGCGCAGGAGCTGCGGAGTCGGTTGATCGCGCATGGCCGTCCCGCCGAGGCGATCCGAA
TCATGCCTGGCTTGTCGTTCGTGCTCGGCAGTACGGAGGCAGAGGCCAGGTCGCGAAACGACGAATTGAACGA
GCTCGCCGGGGATCGACGCCTGGCACATCTGGCTGGTCAACTCAGCGTCGATGTGGCGGAGCTGAAGTGGGA
CAAGCCGCTTCCCGGTTGGCTCCTCGAGGGCGCGGCGCCGATCAGCGGTTCCCAGGGAGCTCGCGACATCGT
CGTCAACATCGCTCGGCGGGAGAACCTGACCGTGCGTCAGCTGCTCGATCGGGTGATCACGTGGCACCGCTTC
GTGGTCGGATCGCCTGAACAGATCGCCGATGCCATCGAGGACTGGTTCGTTGCGGGCGCTGTCGACGGCTTCA
ACCTGATGCCGGATGTCTTCCCGTCGGGTCTCGAGTTGTTCGTCGACCACGTCGTACCGATCCTCCGGGACCG
AGGGTTGTTCCGGCGGGAGTACACATCGACGACATTGCGTGGGCATCTGGGCCTCGAGCGCACCCCAGACCG
GCCGTCGTCGGGTTCGATCCGCCGGACCGGTTAGTCNAGANNCNNNNNNNNNNN

>pMAL1222_MBP-F

NNNNNNNNNNNNNNANNCNCACATAACAATAACAACAACCTCGGGATCGAGGGAAGGATTTCAGAATTCATGGC
TGATCGAGAGCTCCATCTGGGCGTCAATGTCCTCTCGGACGGTATGCACCCAGCCGCGTGGCAGTATCCGAGT
TCCGATCCGTCGTGGTTCACGGATCCGGCGTACTGGATTCGTGTTGCGCAGATCGCGGAGCGAGGAACCCTCG
ACGCGGTCTTCCTCGCCGACAGTCCGTCGTTGTTCCAGCCGCCCGACCAGCCGCTGAGTGCGCCACCGTTGGC
CCTGGACCCGATCGTGTTGTTGTCGACACTGGCATCGGTGACCACACACATCGGACTCATCGGTACGGTGTCGA
CCTCGTTCGAGGAGCCGTACAACGTCGCCCGCCGATTCTCGACGCTGGACCACCTCAGCCGGGGTCGTGTGGC
ATGGAACGTCGTGACGAGTAGTGATCGGTATGCCTGGAACAATTTCGGTGGTGGTGAACAACCCGACCGCGCT
ACTCGATACGAGCGGGCCGGCGAGTTCATCGAAGTCGTCCGGGCATTGTGGGATTCGTGGGACGACGACGCA
GTTGTCGCCGACAAGTCCACGGGTGCGTTCAGTAAGGTCGGTGCGATACGACCGATCCGGCATCGCGGTGGG
CACTTCTCGGTGGACGGGCCGTTGACTCTACCCAGATCCCCACAGGGGCATCCGGTGTTGTTTCAGGCAGGCG
GTTCCACCGGCGGGTTGGATCTGGCGGCGAAGTACGCCGACGGGGTCTTTGCGGCACAGGCCTCGCTCGAGG
ATGCGCTGTCCAACGCGCAGGAGCTGCGGAGTCGGTTGATCGCGCATGGCCGTCCCGCCGAGGCGATCCGAA
TCATGCCTGGCTTGTCGTTCGTGCTCGGCAGTACGNAGGCAGAGGCCAGGTCGCGAAACGACGAATTGAACGA
GCTCGCCGGGGATCGACGCCTGGCACATCTGGCTGGTCAACTCAGCGTCGATGTGGCGGAGCTGAAGTGGGA
CAAGCCNCTTCCCGGTTGGCTCCTCNAGGGCGCNGCGCCGATCAGCGGTTCCCNGGNGCTCGCGACATCNTC
GTNACATCGNTCGNNGGANNNNGACGTGCGTCANCTGCTCGATCGGNNGATCNCNNNGCNNCNNNTTNNNGNN
NNNATCGCCNNAACNNATNNNCNATNCCATCNNNGNANTNNNN

>pMALSsuD

GCACTTCACCAACAAGGACCATAGATTATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAA
GGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCC
GGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACG
ACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTG
TATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCG
CTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACT
GAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTG

ACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGC
GAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGC
AGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCA
GCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTG
AGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGA
TGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTG
GCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGAT
GTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCC
CTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGGATCGAGGG
AAGGATTTCAGAATTC**ATG**AGTCTGAATATGTTCTGGTTTTTACCGACCCACGGTGACGGGCATTATCTGGGAAC
GGAAGAAGGTTCACGCCCGGTTGATCACGGTTATCTGCAACAAATTGCGCAAGCGGCGGATCGTCTTGGCTATA
CCGGTGTGCTAATTCCAACGGGGCGCTCCTGCGAAGATGCGTGGCTGGTTGCCGCATCGATGATCCCGGTGAC
GCAGCGGCTGAAGTTTCTTGTCGCCCTGCGTCCCAGCGTAACCTCACCTACCGTTGCCGCCCGCCAGGCCGCC
ACGCTTGACCGTCTCTCAAATGGACGTGCGTTGTTTAACCTGGTCACAGGCAGCGATCCACAAGAGCTGGCAGG
CGACGGAGTGTTCCTTGATCATAGCGAGCGCTACGAAGCCTCGGCGGAATTTACCCAGGTCTGGCGGCGTTTAT
TGCAGAGAGAAACCGTCGATTTCAACGGTAAACATATTCATGTGCGCGGAGCAAAACTGCTCTTCCCGGCGATT
CAACAGCCGTATCCGCCACTTTACTTTGGCGGATCGTCAGATGTCGCCCAGGAGCTGGCGGCAGAACAGGTTG
ATCTCTACCTCACCTGGGGCGAACCGCCGGAACTGGTTAAAGAGAAAATCGAACAAGTGCGGGCGAAAGCTGC
CGCGCATGGACGCAAAATTCGTTTCGGTATTCGTCTGCATGTGATTGTTCGTGAAACTAACGACGAAGCGTGGC
AGGCCGCCGAGCGGTTAATCTCGCATCTTGATGATGAAACTATCGCCAAAGCACAGGCCGCATTCGCCCGGAC
GGATTCCGTAGGGCAACAGCGAATGGCGGCGTTACATAACGGCAAGCGCGACAATCTGGAGATCAGCCCCAAT
TTATGGGCGGGCGTTGGCTTAGTGCGCGGCGGTGCCGGGACGGCGCTGGTGGGCGATGGTCCTACGGTCGCT
GCGCGAATCAACGAATATGCCGCGCTTGGCATCGACAGTTTTGTGCTTTCGGGCTATCCGCATCTGGAAGAAGC
GTATCGGGTTGGCGAGTTGCTGTTCCCGCTTCTGGATGTCGCCATCCCGGAAATTCCCCAGCCGCAGCCGCTG
AATCCGCAAGGCGAAGCGGTGGCGAATGATTTTATCCCCCGTAAAGTCGCGCAAAGC**TAA**AAGCTTGGCACTG
GCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCC
TTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGC
GAATGGCAGCTTGGCTGTTTTGGCGGATGAGATAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGC
GGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAG
TGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAA
AACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGA
CAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCAT
AAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTTG
TTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAA
AAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTT
GCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACT
GGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTCTCCAATGATGAGCACTTTTAAAGT
TCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTC
AGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCA
GTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTA
ACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAG

>pMALSsuD_EcoliSsuDF

GGGCATTATCTGGGAACGGAAGAAGGTTCACGCCCGGTTGATCACGGTTANNNNNNANAAATTGCGCAAGCGG
CGGATCGTCTTGGCTATACCGGTGTGCTAATTCCAACGGGGCGCTCCTGCGAAGATGCGTGGCTGGTTGCCGC
ATCGATGATCCCGGTGACGCAGCGGCTGAAGTTTCTTGTCGCCCTGCGTCCCAGCGTAACCTCACCTACCGTTG
CCGCCCGCCAGGCCGCCACGCTTGACCGTCTCTCAAATGGACGTGCGTTGTTTAACCTGGTCACAGGCAGCGA
TCCACAAGAGCTGGCAGGCGACGGAGTGTTCCTTGATCATAGCGAGCGCTACGAAGCCTCGGCGGAATTTACC
CAGGTCTGGCGGCGTTTATTGCANAGAGAAACCGTCGATTTCAACGGTAAACATATTCATGTGCGCGGAGCAAA
ACTGCTCTTCCCGGCGATTCAACAGCCGTATCCGCCACTTTACTTTGGCGGATCGTCAGATGTCGCCCAGGAGC
TGGCGGCAGAACAGGTTGATCTCTACCTCACCTGGGGCGAACCGCCGGAACTGGTTAAAGAGAAAATCGAACA
AGTGCGGGCGAAAGCTGCCGCGCATGGACGCAAAATTCGTTTCGGTATTCGTCTGCATGTGATTGTTCGTGAAA
CTAACGACGAAGCGTGGCAGGCCGCCGAGCGGTTAATCTCGCATCTTGATGATGAAACTATCGCCAAAGCACAG
GCCGCATTCGCCCGGACGGATTCCGTAGGGCAACAGCGAATGGCGGCGTTACATAACGGCAAGCGCGACAATC
TGGAGATCAGCCCCAATTTATGGGCGGGCGTTGGCTTAGTGCGCGGCGGTGCCGGGACGGCGCTGGTGGGCG
ATGGTCCTACGGTCGCTGCGCGAATCAACGAATATGCCGCGCTTGGNATCGACAGTTTTGTGCTTTCGGNTATC
CGCATCTGGNNAAGCGTATCGGGNNNGNGANTTGCTNTNCCCGCTTCTGGATGNCNCCNTCCCGGANNTCCCC

ANCCNCNNNNGCTGNANCNNCAGNNNAANCNNNGNCNNANGATTTATCCCCCGTAAGNCGCGCAAAGCTNNNN
NTTGNNNNTNNNNNCNNTTTANNANNNNNNGNCNGGNNNAANNNNNNGNNNNNNNNTTNNNNNNNNNNANNNN
NNNCNCNN

>pMALSsuD_EcoliSsuDR_reverse complement

NNNNNNTCNNTNNNNNNANNNNGCNNNNNNNNNNNNGNCNNNNNNNNNNNNNNNNNNNCNNCNTCGGNATNGNN
GNNNATTCAGNNNNNNNNNNTCNNANNTNTTNNGTTTNNNNNCCNNGNGANNGGNCNTNNTCTGNNNGAAGAA
GNTTNNNGCCCNNTNATCNCNNTNATNNGNNNAAANTGNGCNAGNNGNGGATCGTNTNGCTATNCCNNNNNGC
TAATTCCANCGGGNNNNTCCNGCGAAGATGCGNNCTGGTTGCCGCATCGATGATCCCGGTGACGCAGCGGCTG
AAGTTTCTTGTCNCNTGCGTCCCAGCGTAACCTCACCTACCGTTGCCGCCCGCCAGGCCGCCNCGCTTGACCG
TCTCTCAAATGGACGTGCGTTGTTTAACCTGGTCACAGGCAGCGATCCACAAGAGCTGGCAGGCGACGGAGTG
TTCCTTGATCATAGCGAGCGCTACGAAGCCTCGGCGGAATTTACCCAGGTCTGGCGGCGTTTATTGCAGAGAGA
AACCGTCGATTTCAACGGTAAACATATTCATGTGCGCGGAGCAAAACTGCTCTTCCCGGCGATTCAACANCCGTA
TCCGCCACTTTACTTTGGCGGATCGTCAGATGTCGCCCNGGAGCTGGCGGCAGAACAGGTTGATNTNTACCTCA
CCTGGGGCGAACCGCCGGAACTGGTTAAAGAGAAAATCGAACAAGTGCGGGCGAAAGCTGCCGCGCATGGAC
GCAAAATTCGTTTCGGTATTNGTNTGCATGTGATTGTTCGTGAAACTAACGACGAAGCGTGGCAGGCCGCCGAG
CGGTTAATCTCGCATNTTGATGATGAAACTATCGCCAAAGCNCAGGCCGCATTCGCCCGGACGGATTCCGTAGG
GCAACAGCGAANGGNGGCGTTACATAACGGCAAGCGCGACAATNTGGAGATCAGCCCCAATTTATGGGCGGGC
GTTGGCTTAGTGCGCGGCGGTGCCGGGACGGCGCTGGTGGGCGATGGTCCTACGGTCGCTGCGCGAATCAAC
GAATATGCCGCGCTTGGCATCGACAGTTTTGTGCTTTCGGGCTATCCGCATNTGGAAGAAGCGTATCGGGTTGG
CGAGTTGCTGTTCCCGCTTNNNGTTNTNNCCATCCCGGAAATTCCCCAGCCGCAGCCGCNGAATCCGCAAGGC
GAAGCGGNN

>pMAL205

GCACTTCACCAACAAGGACCATAGATTATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAA
GGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCC
GGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACG
ACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTG
TATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCG
CTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACT
GAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTG
ACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGC
GAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGC
AGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCA
GCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTG
AGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGA
TGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTG
GCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGAT
GTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCC
CTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGGATCGAGGG
AAGGATTTCAGAATTC**ATG**CCCACCACCCCGCATCGCCCACTGATCCTCAACGCGACCGACATGGCCACCGCCA
ACCACATCGCCTTCGGGCTCTGGCGTCTGGCCGACCCGGACAAGCCCGACTACACAACACTGCGGTTCTGGAC
CGATCTCGCGATCGAACTGGAACAGTCCGGATTCGACGCACTGTTCCTCACCGACGCGCTCGGGCAGCTCGAC
ACCTACACCGCCAGTGCCGACCCGGCGCTGCGCACCGCGACGCAGACACCCCTCGACGATCCGCTCCTCGCG
GTATCGGCGATGGCCGCGGTGACCGAACAGCTCGGCTTCGCGGTGACCGTCTCGGCCACCTACGAGCATCCCT
ACCTGCTCGCCCGCAAGTTCACCACCCTCGACCATCTGACCGACGGCCGCATCGGATGGAACATCGTGACCTC
GCAACTCGACAGCGCGGCAAGGAACCTCGGCCTGGAGCGGCAGATCCCCCACGACGAACGCTACGAACGTGC
CGAGGAGTTCCTCACGGTGGCGTACAAACTGTGGGAGGGATCGTGGGATGAGGGCGCAGTGCTCCGCGACCG
TGGGACCGGGGTCTACGCCGATCCCTCCAGGGTCCATGCGATCGGACACCACGGCCGCTACTTCTCGGTTCCC
GGCGCGGCCCTGAGCGAACCATCCCGACAACGCACCCCGGTTCTGTACCAGGCCGGAACATCACCCCGTGGA
AGTCTTTTCGCCGCTCGTCACGCCGAGATCGTCTTCGTCGCGGGCCACGAGCCCGACGTCCTGCGCCGAAACA
TCGACCGGATTCGTGTGCTGGCACGCGAGCAGGGTCGCGAACCCGACGACATCAAGTTCGTCGCGTCGGCGC
TGGTGATCACCGATGAGACGGATGCGGGCGCCGAAGCCAAACTACGCAGGTATCAGGACGCCTACAGCATCGA
GGGCGCCCTGACCCACTTCTCGGCCATCACCGGTATCGACTGGTCCGAGTACGACATCGACGCACCCCTCTCC
TACATCGAAACCGACAGCAACCGTTCGATTCTCGCCTCTCTCACCACCGACGCCCCACCCGGCAGTGTCTGGAC
TCTCCGCCGTCTGCTCGCACCCGCCCGCGGGGTCTCCTATGCCGATGCCGTCGTCGGGTCGGGGACGACGGT
GGCCGACCGGCTCGAGAAGCTCGCCGACGAGACCGGCGTCGACGGCTTCAACCTCTCTGCAGCCGTCGCGCA
CGAGAGCTACCGGGACATCGCCGATCACGTCATCCCGGTGCTCCGTGACCGCGGACGGATACGTCGGCCCGC
ATCGGCGACGGCGTCATTGCGCGAGAAACTGTTCGAGACGACAGAACCCCACATCGGTTCGCGTCATCCGGCC

TCCCGATACCGCAACGCGTTCACCGGGCTGCCGTCGGCGGCACCGCGCCCGGTCGCCGCATCC**TGA**AAGCTT
GGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCAC
ATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCT
GAATGGCGAATGGCAGCTTGGCTGTTTTGGCGGATGAGATAAGATTTTCAGCCTGATACAGATTAAATCAGAACG
CAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAAC
TCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCAT
CAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTG
AGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGC
CCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACT
CTTTTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAAT
ATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCC
TGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACA
TCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTCTCCAATGATGAGCACTT
TTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACAC
TATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAA
TTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAA
GGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATG
AAGC

>pMAL205_205F

NCGACCGNNATGGCNNCCGCCNACCACNTCGCCTTCGGGCTCTGGCGTCNGGCCNACCCGGNCAAGCCCGAC
TACACAACACTGNGGTTCTGGACCGATCTCNCGATCGAACTGGAACAGTCCGGATTCGACGCACTGTTCCTCAC
CGACGCGCTCGGGCAGCTCGACNCCTACNCCGCCNGTGCCGACCCGGCGCTGCGCACCGCGACGCANACNC
CCCTCGACGATCCGCTCCTCGNGGTATCGGCGATGGCCGCGGNGACCGAACAGCTCGGCTTCGCGGNGACCG
TCTCGGCCACCTACNAGCATCCCTACCTGCTCGCCCGCAAGTTCACCNCCCTCGACCATCTGACCGACGGCCG
CATCGGATGGAACATCGNGACCTCGCAACTCGACAGCGCGGCAAGGAACCTCGGCCTGGAGCGGCANATCCC
CCACNACNAACGCTACGAACGTGCCGAGGAGTTCCTCACGGNGGCGTACAAACTGNGGGAGGGATCGNGGGA
TGAGGGCGCANTGCTCCGCGACCGNGGGACCGGGGTCTACGCCNATCCCTCCAGGGTCCATGCGATCGGACA
CCACGGCCGCTACTTCTCGGTTCCCGGCGCGGCCCTGANCGAACCATCCCGACAACGCACCCCGGTTCTGTAC
CAGGCCGGAACATCACCCCGTGGAAGTCTTTTCGCCGCTCGTCACGCCNAGATCGTCTTCGTCGCGGGCCACN
AGCCCGACGTCCTGCGCCGAAACATCGACCGGATTCGTGTGCTGGCACGCGAGCAGGGTCGCGAACCCGACG
ACATCAAGTTCGTCGCGTCGGCGCTGGTGATCACCGATGAGACGGATGCGGGCGCCGAAGCCAAACTACGCAG
GTATCAGGACGCCTACAGCATCGAGGGCGCCCTGACCCACTTCTCGGCCATCACCGGTATCGACTGGTCCGAG
TACGACATCGACGCACCCCTNTNCTACNTCGAANCGANNGCNNNNGTTCGATTNTCGCNNNNNTNACCNNNACN
NCCCACCCGGCNNTGNCTGGACTNNTCNNCNTCNNNTNNNACCCNNNGNNGNNNNNCCTNNNNNATNCNNCNN
NGGNNGGGNNNANGNNNNNACNNNNNNNNNCNCNNNANNNNNNCGNNNNNGNTTNNNNNCNNNNGCNNNNN
NCNNNNNNNANNNNNNNNNN

>pMAL205_205R_reverse_complement

NNNNNANCNANCCNNNNNNNNNNNNNNNNNNNCNNNNNNNNNANNNCNNNANGNNNGNNCCTNNNGNNTTN
NNGATNNNNCGNNNNCCNNNANNTNNGNNTTNNNNGTGNCNNNNTCNNNNNCTACGAGCANCCCTNCNNNNNG
CCNGCAAGTTCACCNCCCTCGNCCNTCTGNNNNCNGNCGCATCGNATGNAACANNGTGNCCTNGCAANTNGAC
AGCGNNGCANGNAACNTNGGCCTGGAGCGGCAGANCCCCNCGACGAACGCTACGAACGTGCCGAGGAGTTCC
TCNCGGTGGCGTACAAACTGTGGGAGGGATCGTGGGATGAGGGCGCAGTGCTCCGCGACCGTGGGACCGGG
GTNTACGCCGATCCCTCCAGGGTCCATGCGATCGGACACCACGGCCGCTACTTCTCGGTTCCCGGCGCGGCCC
TGAGCGAACCATCCCGACAACGCACCCCGGTTCTGTACCAGGCCGGAACATCACCCCGTGGAAGTCTTTTCGC
CGCTCGTCACGCNGAGATCGTNTTNGTCGCGGGCCACGAGCCCGACGTCCTGCGCNGAAACATCGACCGGATT
CGTGNGCTGGCNCGCGAGCAGGGTCGCGAACCCGACGACATCAAGTTCGTCGCGTNGGCGCTGGTGATCACC
GATGAGACGGATGCGGGCGCCGAAGCCAAACTACGCAGGTATCAGGACGCCTACAGCATNGAGGGNGCCCTG
ACCCACTTNTNGGCCATCACCGGTATCGACTGGTCCGAGTACGACATNGACGCNCCCCTCTCNTACATNGAAAC
CNACAGCAACCGTTCGATTCTCGCCTNTCTCACCNCCGACGCCCCNCCCGGCAGTGTNTGGACTNTCCGCCGT
NTGCTCGCNCCCGCCCGNGGGGTNTCCTATGCCNANGCCGTCGTCGGGTCGGGGACGACGGNGGCCGACCG
GCTCGAGAAGCTNGCCGACGAGACCGGNGTNGACGGCTTCAACCTNTNTGCAGCCGTCGCGCNCGAGAGCTA
CCGGGACATCGCCGATCACGTCATCCCGGTGCTCCGTGACCGCGGACGGATACGTCGGCCCGCATCGGNGAC
GGNGTCATTGCGCGAGAAACTGTTCGAGACGACAGAACCCCACATCGGTTCGCGTCATCCGGCCTCCCGATAC
CGCAACGCNNCACCGGGCTNNN

>pMAL1666

GCACTTCACCAACAAGGACCATAGATTATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAA
GGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCC
GGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACG
ACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTG
TATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCG
CTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACT
GAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTG
ACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGC
GAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGC
AGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCA
GCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTG
AGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGA
TGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTG
GCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGAT
GTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCC
CTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGGATCGAGGG
AAGGATTTCAGAATTC**ATG**ACCGTGGCACGTCCTCGGATGATCTTCAACGCGTTCAACATGTTCACCGTCTCCCA
TCACGACCAGGGGATGTGGGCATGGCCGGGAAGCCGTCAGCGCGAATACAACTCGGTCGACTACTGGGTGGA
CGTCGCACGCCTGCTCGAGCGTGGCCATTTCGACACGCTGTTCTTCGCCGATGTGCTCGCCCCGTACGACACC
TTCGGTGACAGCAGCGAGCGGGCCATCTCGAGTGGAATGCAGTTCCCGGTCAACGATCCCGGCACCCTGATCC
CCGCCCTGGCCCACGCCACCGACGACCTGGGTTTCGTCCTGACCCAGAACATCCTGCAGGAACCGCCGTATGC
CTTCGCGCGCAAGATGTCGTCGCTCGATCACCTCACCCGCGGACGGATCGCCTGGAACATCGTCACCACGTTC
CTGCCCGGTGCGGGCCGCAACCTGGGGTTCGCCGGATTGCCCGACCACGCCGAACGGTATGCGCGCGCAGAC
GACTTCGTCGACGTCGTCTACAAGTTGTGGGAGGCGTCCTGGGAGGACGACGCCGTCATCGCCGACGCCGCG
ACCGGCCGATACAACGACCCGGCCAAGATCCATCGCATCGATCACACCGGGCCCTATTACGACGTCGTCGGCC
CGCATCTGTGTGAACCGTCGCCGCAGCGCACACCGTTTCTCGTGCAGGCCGGGGTGTCCGCGCGCGGACGTG
ACTTCGCCGGCCGCAACGCGGAGGCATTGTTCATCAATGCGTTGAGTCCGCAGGAGGCCGCACCCGTCGTCGC
CGACGTCCGGGCCGCCGCGGCACGTCACGGGCGTGATCCGGCGAGCGTGGTGCTGTTCGGCATCCTCGGTTT
CGTCGTCGGCAGCACGGAGGCCGAGGCCAAGCGATTGCAGGAGGAGATCACCGACTTCCAGAGCATCGACGC
CCACCTCGCCAAGCAGAGCGTCTTCCTCGGATACGACTTCGGCCAACTCGACCCGAACGAGCCGATCGGTGAG
ATCGCCAAACGGCCCGAAGGAAAGGAAGGCGTCGTCGGGCAACTCATCGCGATGTCGCCGAACGATCGCTTCA
CCATCGGTGAGCTGGTCCGCTGGTACGGCAACCTGCGGGTGGTGGGCACCCCGGAACAGATCGCCGACCACA
TCGAGGCCTGGCAGGATGCCGGCGTCGGTGGGATGAACGTCCAGTACGTGGTGTCGCCGGGCACCTTCGAGG
ACTTCGTCGACCATGTCGCCCCCGAACTCGAACGCCGCGGGATCATGCAGGATCGCTACCGGCCGGGAACGTT
GCGGGAAAAGATCTTTCCCGGCAACGGGCCGTACCTCCCCGAAGCGCATCCGGCGCGTGGACACCGTCGGGC
GGCATTCGGTGTGACCGTC**TGA**AAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGC
GTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGA
TCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCAGCTTGGCTGTTTTGGCGGATGAGATAAGATTT
TCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCG
GTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCC
ATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTAT
CTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAAC
GGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGA
CGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGA
GACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCT
TATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAA
GATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCC
CGAAGAACGTTCTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGG
GCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGC
ATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACT
TACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGC
CTTGATCGTTGGGAACCGGAGCTGAATGAAGC

>pMAL1666_1666F

GCGTTNANATGTTCNCCGTCTCCCATCACGACCAGGGGATGNGGGCATGGCCGGGNAGCCGTCAGNGNGAATA
CAACTCGGTCGACTACNGGGNGGACGTCNCNCNCCTGCTCGAGCGTGGCCNTTTCGACNCNCTGTTCTTCNCC
NATGTGCTCNCCCCGTACNACNCCTTCGGNGACAGCAGCGAGNGGGCCATCTCGNGNGGAATGCANTTCCCGG
TCAACGATCCCGGCNCCCTGATCCCCNCCCTGGCCCACNCCNCCGACNACCTGGGTTTCGTCCTGACCCANAA
CATCCTGCAGGAACCGCCGTATGCCTTCNCGCGCAANATGTCNTCNCTCGATCNCCTCNCCCGCGGACGGATC
NCCTGGAACATCNTCNCCNCNTTCCTGCCCGGNGNGGGCCGCAACCTGGGGTTCNCCGGATTGCCCGACCNC
NCCNAACGGTATGCGCGCGCANACNACTTCGTCNACGTCGTCTACAAGTTGNGGGAGGNGTCCTGGGAGGACN

ACNCCGTCATCNCCNACNCCNCGACCGGCCNATACAACGACCCGGCCAANATCCATCGCATCGATCACACCGG
GCCCTATTACNACGTCNTCGGCCCGCATNTGTGTGAACCGTCNCCGCANCGCACACCGTTTCTCGTGCAGGCC
GGGGTGTCCGCGCGCGGACGTGACTTCNCCGGCCGCAACGCGGAGGCATTGTTCATCAATGCGTTGAGTCCG
CAGGAGGCCGCACCCGTCGTCNCCGACGTCCGGGCCGCCGCGGCACGTCACGGGCGTGATCCGGCGAGCGT
GGTGCTGTTCGGCATCCTCGGTTTCGTCGTCGGCAGCACGGAGGCCGAGGCCAAGCGATTGCAGGAGGAGAT
CACCNACTTCCNNAGCATCGACGCCCACCTCGCCAAGCANAGCGTCTTCCTCGGATACGACTTCGGCCACTCG
ACCCGAACGAGNCGATCGGTGNNNTNNCNNACGGCCNNAGGAAANGNAGGNGTCGTCGGGCAACTCATCNNG
ATGTCNCNAANGATCGCTTTCNCNTCGGNGNNCTNGNCNGCTNNANGNAACNNNNNGGGTNGGNGGNNNCCN
NAANNNANNNNNAANNNNTCGNNNNNTGNNGNTNCNGNNNGGTGGNNNAACNNCNNNNANNNNNTNNCNNNN
NCCTNNNNNNTNNNNNATGTN

>pMAL1666_1666R_reverse complement

NGNNTTNGNNNNNNNNNNNNCNGNNTGTNNNCGNCCNNNNGNNNCNNNGNGNNNGNNNNGANNGGNNNNNN
NNANTGGAANNNNGNNCCGGTCACNANCCNNNNNCCNGNTNCCCGCCNNGCCCANNCNNCGACGNCNNNNTT
TCGTCCTGACCCAGNNCATNNNGCAGNANCCGCCNTNNGCCNTTNGCGCGCAANATGNNGTCGCTCGATCNCC
TCACCCGCGGACGGATCGCCTGGAACATCGTCNCACGTTCCTGCCCGGTGCGGGCCGCAACCTGGGGTTCGC
CGGATTGCCCGACCACGCCGAACGGTATGCGCGCGCAGACGACTTCGTNGACGTCGTCTACAAGTTGTGGGAG
GCGTCCTGGGAGGACGACGCCGTCATCGCNGACGCCGCGACCGGCCGATACAACGACCCGGCCAAGATCCAT
CGCATCGATCACACCGGGCCCTATTACGACGTCGTCGGCCCGCATCTGTGTGAACCGTCGCCGCAGCGCACAC
CGTTTCTCGTGCAGGCCGGGGTGTCCGCGCGCGGACGTGACTTCGCCGGCCGCAACGCGGAGGCATTGTTCA
TCAATGCGTTGAGTCCGCAGGAGGCCGCACCCGTCGTCGCCGACGTCCGGGCCGCCGCGGCACGTCACGGGC
GTGATCCGGCGAGCGTGGTGCTGTTCGGCATCCTCGGTTTNGTNGTCGGCAGCACGGAGGCCGAGGCCAAGC
GATTGCAGGAGGAGATCACCGACTTCCAGAGCATCGACGCCCACCTCGCCAAGCAGAGCGTCTTCCTNGGATA
CGACTTCGGCCAACTCGACCCGAACGAGCCGATCGGTGAGATCGCCAAACGGCCCGAAGGAAAGGAAGGNGT
CGTCGGGCAACTCATCGCGATGTCGCCGAACGATCGCTTCACCATCGGTGAGCTGGTCCGCTGGTACGGCAAC
CTGCGGGTGGTGGGCACCCCGGAACAGATCGCCGACCACATCGAGGCCTGGCAGGATGCCGGNGTCGGTGG
GATGAACGTCCAGTACGTGGNGTCGCCGGGCACCTTCGAGGACTTCGTCGACCATGTCGCCCCCGAANTNNAA
CGCCGCGGGATCATGCAGGATCGCTACCGGCCGGGAACGTNNNNGAAAAGATCTTTCCCGGCAACGGGCCGT
ACCNCCCCGAAGCGCNNCCGGCGNN

>pMAL1835

CAAGGACCATAGATTATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGT
CTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGA
AGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTG
GCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACC
TGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAAC
AAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAG
GTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTAT
GCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTC
TGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCT
TTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAAT
TATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTAT
TAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGG
AAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCA
CGTATTGCCGCCACCATGGAAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTG
GTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCG
CAGACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGGATCGAGGGAAGGATTTCAGAA
TTCGGATCCTCTAGA**ATG**TCCCGACCGCCCATCTACAACGGGTTCCTGCACCTGACGCCCAACCATCACAGCCA
CGGCTTCTGGCGTACCCCGGAGGGCGCGGTCCAATACGGGTACAGCAAACTCGATCCCTACGTCGATGTCGTT
CAGACACTCGAACGAGGATTGTTCGATACGCTCTTCATCGCCGACGTCGTCGGGGTCTACGACCTGGATTTCGG
TGACGGCACCACCACCATTCGCGCAGGATCACAGTTTCCCGAACCAGATCCCGTCACCATTGTCTCCGCGCTCG
GACACGCAACCGAGCATCTCGGAATCGCGGTGACCAGCAATATCATTCAAAGTCACCCGTTCACTTTTGCGCGT
CAACTCTCATCGTTGGATCACTTCACTGACGGGCGCGTCGCCTGGAACATCGTCACCTCCTACCTGTCCAACGG
GTTCCGCAACTACGGCTACGACTCGATCGTCGGACACGACGAGCGATACGCGTGGGCGCAGGAGTACGCAGA
CGTGACTTACAAACTCTGGGAACACTCTTGGCAGGACGGTGCAGTGATCCACGATCCGGCCACCAACCGGTTCT
TCGACCCCGACAAGATCCGCACAATCGATCACGTCGGGCCGCGCTACCAGGTCCAGGGACCCCACATCGTCGA
ACCTTCACCGCAGCGCACACCGGTTCTGTTCCAAGCCGGAAACTCCTCCGCAGGACGCGAATTCGCGGTCAAC
AACGCCGAGGTAACCTTCCTTCCCTCGCAGACCCCCGCAACCGCACGCGAAGATATCGCTGTACTCGACGCCC

TGGCTCGCGAGAAGGGCCGCAATCCGGCTTCACTCAAGAAGATCGTCACCTTGTCCACGGTGATCGGATCCAC
CGAGGAAGAAGCGAAGCGTAAGCAACAGTATTTCCGAGACAACATCGATTTCGAAGCGTTGCAGGCATTCTGGA
GTGGTGGCAGCGGCGTCGACCTGACGTCGGTCGACCCGGAGACGCCGCTTGCCGAGCTCGCTCAGAGGGCG
CAACTCGGCGACCACGTCCGGTCCATCTTCCGAGCAGCCGCGCAATCGCAGGACGAACCAGAGTCAGTCTCGT
GGCGTGACTATCTACTCGCGCAGGGACTCCTGCCCGGTCGATTTGCCGGTACACCGGAACAGATCGCCGACCA
CGTAGCCGAATGGGTCGAATCCGGCGTCGACGGTTTCAACGTTGTACCCATCACCACGCTCGGTTGGTGGGAC
GAGTGGGTCGACCACGTTGTACCTGTGCTGCAGGACCGCGGATTAGCACAGCGTGAGTACCATCAGGGAACCC
TGCGCAACAAGCTGTTTCGCTCGGGAGATGCGCTCGACCCGACGCACCGCGGACGGCAGATCCGACTGGCTG
ATGTCATCGGGGCCAAGGGA**TGA**AAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTG
GCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACC
GATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCAGCTTGGCTGTTTTGGCGGATGAGATAAGAT
TTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCG
CGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCC
CCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTT
ATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCA
ACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCT
GACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCAT
GAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCC
CTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTG
AAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGC
CCCGAAGAACGTTCTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCC
GGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAA
GCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCA
ACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACT
CGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGC

>pMAL1835_1835F

NNNNNGCCNNACCNTCACAGCCNCGGNTTCNGGNGTACCCCGGAGGGNGNGGTCCAATANGGGTACNGCAAA
CTCGATCCCTACGTCNATGTCGTTCANACNCTCNAACGAGGATTGTTCGATACNCTCTTCATCGCCNACGTCGTC
GGGGTNTACNACCTGGATTTCGGNGACGGCNCCNCCNCCNTTCNCGCAGGATCNCAGTTTCCCNAACCANATC
CCGTCNCCNTTGTCTCCGCGCTCGGACACNCAACCGAGCATCTCGGAATCGNGGNGACCAGCAATATCATTCAA
AGTCNCCCGTTCANTTTTGCGCGTCAACTCTCATCGTTGGATCACTTCACTGACGGGNGCGTCNCCNGGAACAT
CGNCNCCTCCTACCTGTCCAACGGGTTCCGCAACTACGGNTACNACTCGATCGTCGGANACNACNAGCGATAC
NCGNGGGCNCAGGAGTACNCANACGTGACTTACAAACTCNGGGAACNCTCTNGGCNGGACGGNGCAGNGATC
CACGATCCGGCCACCAACCGGTTCTTCNACCCCNACAANATCCGCACAATCNATCACGTCGGGCCGCGNTACC
AGGTCCAGGGACCCCACATCGTCNAACCTTCACCGCAGCGCACACCGGTTCTGTTCCAAGCCGGAAACTCCTC
CGCAGGACGCGAATTCGCGGTCAACAACGCCNAGGTAACCTTCCTTCCCTCGCANACCCCCGCAACCGCACGC
GAAGATATCGCTGTACTCGACGCCCTGGCTCGCGAGAAGGGCCGCAATCCGGCTTCACTCAANAAGATCGTCA
CCTTGTCCACGGNGATCGGATCCACCGAGGAANAAGCGAAGCGTAAGCAACAGTATTTCCNAGACAACATCGAT
TTCGAAGCGTTGCAGGCATTCTGGAGTGGNGGCAGCGGCGTCGACCTGACGTCGGNCGANNCGGAGANNNNN
NTNGNNNAGCTCGCTCNNAGGGNGNANTCGGCGACNACGTNCGGNCNNNTTNCGANNNGCNNNNNNTCGNNN
NCNNANCANANTCANNNTCGGNNNNNNGACTNTNNNCTCNNNNNGGNNTNNNNNNNCGATNNNNACANNNNNN
NTNNCGACNNNNANNCNANGGNNCGAANNNGNNCNANNNNNANNTNNNNNNNNACNNNNNNNNNNNNNNNNN
NNNNNCTNNNNTN

>pMAL1835_1835R_reverse complement

NNTNTNNNANNNNTTTNNNNCNNNNNNTNNNNNNNNNNTNNNTTCGGNNNNNNCNCNNNNNTTNNNNNGANNN
NNNNNNACNAGANNCNNNNCATNNTCNNNNNNNNNNNNNNNGCATTNNNNNNNNNNNNNNNNGCAANNTCATTCA
AAGTNNNCGTTCACTTTTGNNNNNCAACTNTNANNGTTGGANNACTTCACTNNCGGGCGCGNNNCCNGNANNNN
CGTCANNNCCTNCNTGNNNANCGNNNTCNNNAACTACGGCTACGACTCGATNGTNGGACACGACNANCGATAC
GCGTGGGCGCAGGAGTACGCAGACGTGACTTACAAACTCTGGGANCACTCTTGGCAGGACGGNGCAGTGATCC
ACGATCCGGCCNCCAACCGGNTNTTNGACCCCGACAAGATCCGCACAATNGATCACGTCGGGCCGCGCTACCA
GGTCCAGGGACCCCACATNGTNGAACCTTCACCGCAGCGCACACCGGTTNTGTTNCAANCCGGAAACTCCTCC
GCAGGACGCGAATTNGCGGTCAACAACGCCGAGGTAACCTTCCTTCCCTNGCANACCCCCGCAACCNCNCGNG
AANANATNGCTGTACTNGACGCCCTGGCNCNCGANAANGGCCNCAATCCGGCTTCNCTCAANAANATNGTCNC
CTTNTCCNCGGNGATNGGATCCNCCGAGGAAGAANNGAAGNGTAANCAACAGTATTTNCGAGACAACATNGATT
TNGAANCGTTGCAGGCATTNTGGAGNGGNGGCAGNGGNGTNGACCNGACGTCGGTNGACCCGGAGACNCCNC
TTGCCGAGCTCGCTCAGAGGGNGCAACTNGGNGNCCNCGTCCGGTCCATNTTNCGAGCANCCGNGCAATNNCA

GGACGAACCAGAGTCAGTNTCGNGGCGNGACTATNTACTCGCGCAGGGACTCCNNCCCGGTCGATTTGCCGGT
ACNCCGGAACANATCNCCGACCNCGTANCCGAANGGGTNGAATCCGGNGTNGACGGNTTCAACGTTGTACCCN
NCNCCNCNCTCGGTTGGNGGGACGAGNGGGTNGACCNCGTTGTACCTGTGCTGCAGGNCCNCGGATTAGCNC
AGCGTGAGTACCATCAGGGAACCCTGNGCAACAANCTGTTTCNCTCGGGAGANGCGCTCGACCCGACNCNCCG
CGNACGGCAGNNC

>pET28bSsuD

CGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTT
CCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAG
ACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTC
GCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGA
ACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGAC
GCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCA
CGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACT
GGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCA
GCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAA
ACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAAT
TGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGAC
GCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGC
AAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCC
GAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCA
GCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCG
CGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTT
TAAGAAGGAGATATACCATGGCC**ATG**AGTCTGAATATGTTCTGGTTTTTACCGACCCACGGTGACGGGCATTATC
TGGGAACGGAAGAAGGTTCACGCCCGGTTGATCACGGTTATCTGCAACAAATTGCGCAAGCGGCGGATCGTCTT
GGCTATACCGGTGTGCTAATTCCAACGGGGCGCTCCTGCGAAGATGCGTGGCTGGTTGCCGCATCGATGATCC
CGGTGACGCAGCGGCTGAAGTTTCTTGTCGCCCTGCGTCCCAGCGTAACCTCACCTACCGTTGCCGCCCGCCA
GGCCGCCACGCTTGACCGTCTCTCAAATGGACGTGCGTTGTTTAACCTGGTCACAGGCAGCGATCCACAAGAG
CTGGCAGGCGACGGAGTGTTCCTTGATCATAGCGAGCGCTACGAAGCCTCGGCGGAATTTACCCAGGTCTGGC
GGCGTTTATTGCAGAGAGAAACCGTCGATTTCAACGGTAAACATATTCATGTGCGCGGAGCAAAACTGCTCTTCC
CGGCGATTCAACAGCCGTATCCGCCACTTTACTTTGGCGGATCGTCAGATGTCGCCCAGGAGCTGGCGGCAGA
ACAGGTTGATCTCTACCTCACCTGGGGCGAACCGCCGGAACTGGTTAAAGAGAAAATCGAACAAGTGCGGGCG
AAAGCTGCCGCGCATGGACGCAAAATTCGTTTCGGTATTCGTCTGCATGTGATTGTTCGTGAAACTAACGACGAA
GCGTGGCAGGCCGCCGAGCGGTTAATCTCGCATCTTGATGATGAAACTATCGCCAAAGCACAGGCCGCATTCG
CCCGGACGGATTCCGTAGGGCAACAGCGAATGGCGGCGTTACATAACGGCAAGCGCGACAATCTGGAGATCAG
CCCCAATTTATGGGCGGGCGTTGGCTTAGTGCGCGGCGGTGCCGGGACGGCGCTGGTGGGCGATGGTCCTAC
GGTCGCTGCGCGAATCAACGAATATGCCGCGCTTGGCATCGACAGTTTTGTGCTTTCGGGCTATCCGCATCTGG
AAGAAGCGTATCGGGTTGGCGAGTTGCTGTTCCCGCTTCTGGATGTCGCCATCCCGGAAATTCCCCAGCCGCA
GCCGCTGAATCCGCAAGGCGAAGCGGTGGCGAATGATTTTATCCCCCGTAAAGTCGCGCAAAGCAAGCTTGCG
GCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGG
CTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTG
AAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGG
TGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTT
CTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTT
ACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTT
TTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTAT
CTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAA
AATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAAC
CCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCA
TCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGA
GAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATC
AATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCC
GGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAA
TCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAA
GGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGA
ATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGA

>pET28bSsuD_T7

GGGGGAACATTCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGCCATGAGTCTGAATATGT
TCTGGTTTTTACCGACCCACGGTGACGGGCATTATCTGGGAACGGAAGAAGGTTCACGCCCGGTTGATCACGGT
TATCTGCAACAAATTGCGCAAGCGGCGGATCGTCTTGGCTATACCGGTGTGCTAATTCCAACGGGGCGCTCCTG
CGAAGATGCGTGGCTGGTTGCCGCATCGATGATCCCGGTGACGCAGCGGCTGAAGTTTCTTGTCGCCCTGCGT
CCCAGCGTAACCTCACCTACCGTTGCCGCCCGCCAGGCCGCCACGCTTGACCGTCTCTCAAATGGACGTGCGT
TGTTTAACCTGGTCACAGGCAGCGATCCACAAGAGCTGGCAGGCGACGGAGTGTTCCTTGATCATAGCGAGCG
CTACGAAGCCTCGGCGGAATTTACCCAGGTCTGGCGGCGTTTATTGCAGAGAGAAACCGTCGATTTCAACGGTA
AACATATTCATGTGCGCGGAGCAAAACTGCTCTTCCCGGCGATTCAACAGCCGTATCCGCCACTTTACTTTGGCG
GATCGTCAGATGTCGCCCAGGAGCTGGCGGCAGAACAGGTTGATCTCTACCTCACCTGGGGCGAACCGCCGGA
ACTGGTTAAAGAGAAAATCGAACAAGTGCGGGCGAAAGCTGCCGCGCATGGACGCAAAATTCGTTTCGGTATTC
GTCTGCATGTGATTGTTCGTGAAACTAACGACGAAGCGTGGCAGGCCGCCGAGCGGTTAATCTCGCATCTTGAT
GATGAAACTATCGCCAAAGCACAGGCCGCATTCGCCCGGACGGATTCCGTAGGGCAACAGCGAATGGCGGCGT
TACATAACGGCAAGCGCGACAATCTGGAGATCAGCCCCAATTTATGGGCGGGCGTTGGCTTAGTGCGCGGCGG
TGCCGGGACGGCGCTGGTGGGCGATGGTCCTACGGTCGCTGCGCGATCAACGATTATGCCGCCGCTTGGCAT
CGACAGTTTTGTGCTTTCGGGCTATCCGCATCTGGGAGAGCGTTATCGGGTTGGCGAGTTGCTGTTCCCGCTTC
TGATGTCCGCCATCCCGGAAATTCCCCAGCCGCCAGACGGCTGAATCGCAAGGCTAGCGTGGCATGATTATCC
CCGTAGTCGGCAGCAAGCTTGCGCCGACTCGGATCCCCACCATCCACATGAATCCGGGTGCTACAAGTCGTAG
GGACTGGATTGGCTGCTGCAACCGTGAACATAACTAGCTACTTGCGTCTACCGCTGAGCTGCTAGGGGACTATC
CGGTAT

>pET28bSsuD_T7T_reverse complement

GCGAACGACTGGCCGTAATCGCAGTTCCGTTGAGTCGAGTCCGATCCGAATATGCATCATACGCATGGAGCGAT
ACCATCCCTTCGAAAATATTTGTTAACTAGAGGAGAATACCATGCATAGAGTCGATTAGTTTCGTTTACGACCCAC
GTGACGGCATATTCGGACGAAGAAGTTCGCCCGGTGATCACGTATTTGCACAATGGCCAGCGGCGGATCGTCT
GGCTATACGGTGGTGCTAATCACGGGGCGCTCCTGCGAAGATGCGTGCTGGTTGCGCATCGATGATCCCGGTG
ACGCAGCGGCTGAAGTTTCTTGTCGCCTGCGTCCCAGCGTAACCTCACCTACCGTTGCCGCCGCCAGGCCGCC
ACGCTTGACCGTCTCTCAAATGGACGTGCGTTGTTTAACCTGGTCACAGGCAGCGATCCACAAGAGCTGGCAGG
CGACGGAGTGTTCCTTGATCATAGCGAGCGCTACGAAGCCTCGGCGGAATTTACCCAGGTCTGGCGGCGTTTAT
TGCAGAGAGAAACCGTCGATTTCAACGGTAAACATATTCATGTGCGCGGAGCAAAACTGCTCTTCCCGGCGATT
CAACAGCCGTATCCGCCACTTTACTTTGGCGGATCGTCAGATGTCGCCCAGGAGCTGGCGGCAGAACAGGTTG
ATCTCTACCTCACCTGGGGCGAACCGCCGGAACTGGTTAAAGAGAAAATCGAACAAGTGCGGGCGAAAGCTGC
CGCGCATGGACGCAAAATTCGTTTCGGTATTCGTCTGCATGTGATTGTTCGTGAAACTAACGACGAAGCGTGGC
AGGCCGCCGAGCGGTTAATCTCGCATCTTGATGATGAAACTATCGCCAAAGCACAGGCCGCATTCGCCCGGAC
GGATTCCGTAGGGCAACAGCGAATGGCGGCGTTACATAACGGCAAGCGCGACAATCTGGAGATCAGCCCCAAT
TTATGGGCGGGCGTTGGCTTAGTGCGCGGCGGTGCCGGGACGGCGCTGGTGGGCGATGGTCCTACGGTCGCT
GCGCGAATCAACGAATATGCCGCGCTTGGCATCGACAGTTTTGTGCTTTCGGGCTATCCGCATCTGGAAGAAGC
GTATCGGGTTGGCGAGTTGCTGTTCCCGCTTCTGGATGTCGCCATCCCGGAAATTCCCCAGCCGCAGCCGCTG
AATCCGCAAGGCGAAGCGGTGGCGAATGATTTTATCCCCCGTAAAGTCGCGCAAAGCAAGCTTGCGGCCGCAC
TCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGAAGCGAGTC

>pET28bFre

CAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGT
CGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGC
GCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGC
CCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAAC
GCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCC
ACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACA
CCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGC
CAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATG
TAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCAC
GCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCA
CCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGG
ATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCG
CCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATAC
CCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATA

GGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTC
GATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTG
TTTAACTTTAAGAAGGAGATATACCATGGCC**ATG**ACAACCTTAAGCTGTAAAGTGACCTCGGTAGAAGCTATCAC
GGATACCGTATACCGTGTCCGAATCGTACCAGACGCGGCCTTTTCTTTTCGTGCTGGTCAGTATTTGATGGTAGT
GATGGATGAGCGTGACAAACGTCCGTTCTCAATGGCTTCGACGCCAGATGAAAAAGGGTTCATCGAGCTGCATA
TTGGCGCTTCTGAAATCAACCTTTACGCGAAAGCAGTCATGGACCGCATTCTTAAAGATCATCAAATCGTGGTCG
ATATTCCGCACGGTGAGGCATGGCTGCGCGATGATGAAGAACGTCCGATGATTTTGATTGCGGGTGGCACCGG
GTTCTCTTATGCGCGCTCGATTTTGCTGACGGCGCTGGCGCGTAACCCAAACCGTGATATCACCATTTACTGGG
GCGGGCGTGAAGAGCAGCATCTGTATGATCTCTGCGAGCTTGAGGCGCTTTCGCTTAAGCATCCTGGTCTGCAA
GTGGTGCCGGTGGTTGAACAACCGGAAGCGGGCTGGCGTGGGCGTACTGGCACTGTGTTAACGGCGGTATTG
CAGGATCATGGTACGCTAGCAGAGCATGATATCTATATTGCCGGACGTTTTGAGATGGCGAAAATTGCCCGTGA
CCTGTTTTGCAGTGAGCGTAATGCGCGGGAAGATCGCCTGTTTGGCGATGCGTTTGCATTTATCAAGCTTGCGG
CCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGC
TGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGA
AAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGT
GGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTC
TCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTA
CGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTT
TCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTAT
CTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAA
AATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAAC
CCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCA
TCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGA
GAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATC
AATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCC
GGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAA
TCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAA
GGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCT

>pET28bFre_T7

GGGGTACGTTACCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGCCATGACAACCTTAAGCTG
TAAAGTGACCTCGGTAGAAGCTATCACGGATACCGTATACCGTGTCCGAATCGTACCAGACGCGGCCTTTTCTTT
TCGTGCTGGTCAGTATTTGATGGTAGTGATGGATGAGCGTGACAAACGTCCGTTCTCAATGGCTTCGACGCCAG
ATGAAAAAGGGTTCATCGAGCTGCATATTGGCGCTTCTGAAATCAACCTTTACGCGAAAGCAGTCATGGACCGC
ATTCTTAAAGATCATCAAATCGTGGTCGATATTCCGCACGGTGAGGCATGGCTGCGCGATGATGAAGAACGTCC
GATGATTTTGATTGCGGGTGGCACCGGGTTCTCTTATGCGCGCTCGATTTTGCTGACGGCGCTGGCGCGTAACC
CAAACCGTGATATCACCATTTACTGGGGCGGGCGTGAAGAGCAGCATCTGTATGATCTCTGCGAGCTTGAGGCG
CTTTCGCTTAAGCATCCTGGTCTGCAAGTGGTGCCGGTGGTTGAACAACCGGAAGCGGGCTGGCGTGGGCGTA
CTGGCACTGTGTTAACGGCGGTATTGCAGGATCATGGTACGCTAGCAGAGCATGATATCTATATTGCCGGACGT
TTTGAGATGGCGAAAATTGCCCGTGACCTGTTTTGCAGTGAGCGTAATGCGCGGGAAGATCGCCTGTTTGGCGA
TGCGTTTGCATTTATCAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACA
AAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAA
CGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGG
CGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGC
TCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGGCTCC
CTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAACTTGATTAGGGTGATGGTTCACGTAGTGGG
CCATCGCCCTGATAGACGATTTCGCCCTTGACGTTGGAGTCCACGTTCTTTATAGTGACTCTGTCCAAACTGGAC
AACCTCACCCTATTCTTCGTCTATTCTTGATTATAGGATTTGGCGATTCGCTATGCTAAAATGACTGAATACAAATT
TAACGGATTACGAATTACGCTTACATTAGTGCCTTGCCGGAATGGCGCGACCCCGATTGCTTACTTC

>pET28bFre_T7T_reverse complement

CGAGAGTTGCACGCAATTCGCACGATGTGCGCAAGTGTTGCATGCGGTTGAAGTAATCAGTCGCATGCCGTCCA
CTTTCTGCGTTTGCAGAACGTGGCTGCTGTCACCACGCGGATCGTTCTGATAAGAGAACACCGCATCTCTGCGG
ACATCGTATTAACGTACTGGTTTCACATTCACACTTGAATTGACTTCTCTTCCGGGCGCTATCATGCATACGCGAA
AGGTTTTGCGCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCTGCATTAGGAAGCAGCCCA
GTAGTAGGTTGAGGCCGTTGAGCACGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCC
CCCGGCCACGGGGCCTGCCACCATACCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCT
TCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATG

CGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGAT
AACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGCCATGACAACCTTAAGCTGT
AAAGTGACCTCGGTAGAAGCTATCACGGATACCGTATACCGTGTCCGAATCGTACCAGACGCGGCCTTTTCTTTT
CGTGCTGGTCAGTATTTGATGGTAGTGATGGATGAGCGTGACAAACGTCCGTTCTCAATGGCTTCGACGCCAGA
TGAAAAAGGGTTCATCGAGCTGCATATTGGCGCTTCTGAAATCAACCTTTACGCGAAAGCAGTCATGGACCGCAT
TCTTAAAGATCATCAAATCGTGGTCGATATTCCGCACGGTGAGGCATGGCTGCGCGATGATGAAGAACGTCCGA
TGATTTTGATTGCGGGTGGCACCGGGTTCTCTTATGCGCGCTCGATTTTGCTGACGGCGCTGGCGCGTAACCCA
AACCGTGATATCACCATTTACTGGGGCGGGCGTGAAGAGCAGCATCTGTATGATCTCTGCGAGCTTGAGGCGCT
TTCGCTTAAGCATCCTGGTCTGCAAGTGGTGCCGGTGGTTGAACAACCGGAAGCGGGCTGGCGTGGGCGTACT
GGCACTGTGTTAACGGCGGTATTGCAGGATCATGGTACGCTAGCAGAGCATGATATCTATATTGCCGGACGTTTT
GAGATGGCGAAAATTGCCCGTGACCTGTTTTGCAGTGAGCGTAATGCGCGGGAAGATCGCCTGTTTGGCGATG
CGTTTGCATTTATCAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAA
GCCCGAAAGAAGCGAGTT

>pET28b1218

CAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGT
CGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGC
GCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGC
CCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAAC
GCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCC
ACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACA
CCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGC
CAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATG
TAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCAC
GCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCA
CCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGG
ATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCG
CCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATAC
CCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATA
GGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTC
GATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTG
TTTAACTTTAAGAAGGAGATATACCATGGTC**ATG**AACGTAAACGTTGTTGGCGGAATCTCTCAATGGGACACTGT
TCAGGCCGATGCGTGGTGGTCGGCTGCAGTCGGTCCAGCAACGAACGGAGAGATCTCGATGTCGGCACATGG
GCGTCCAATTCACTTGGGCGGGTTTTTGATTGCAGGGAATGTAACCCACAGTCATCCTTCGTGGCGTCATCCAC
GCAGTGATCCCGGGTTTCTCACACCGGAGTACTACCAGCACCTCGGTAGGGTTTTCGAACGCGCGAAATTGGAC
TTCGTCTTCTTTGCCGACAATTCTGCGACTCCTGCCAGCTACCGCAACGATATTCGTGACCCGCTCGCTCGCGG
TACTCAGAGTGCAGCCGGGTTGGATCCCCGCTTCGTCGTTCCTGTCGTCGCGGGTGTCACGCGCAACCTGGGG
ATCGTATCGACCACGTCGGCGACGTTCTACTCGCCATACGACCTCGCCCGGAGCTTTGCCACTCTGGATCATCT
GACCCACGGCCGCGTCGGTTGGAATGTTGTGACTTCCAATACGACCGTCGAGGCGCAGAACTTCGGGCTTGCC
CGACACCTCGACCATGACGTGCGATACGACCGTGCCGAGGAATTGCTTGAGGTCGCGTTCAGGCTGTGGGCCA
GTTGGGACGATGGAGCTCTGATCCAGGATAAAGAGGCGGGTGTCTTCGCCGACCCGGACCTGATTCACAGGCT
CGATCATCACGGGGAGAACTTCGATGTTCGGGGCCCGCTGTCGGTTCCCCGCTCACCGCAGGGACGGCCGGT
CATCTTTCAAGCGGGATCATCCACCCGCGGTCGGGATTTTGCTGCGCGCTGGGCAGAAGCGATTTTCGAGATCG
ACCCGACGTCTGTGGGGCGTAAGGCCTACTACGACGACATCAAGTCGCGAGCCTCCGACTTCGGTCGTGATCC
CGACGGCGTCAAGATACTCCCGTCGTTCATTCCGTTTGTGGGTGAGACCGAGTCGATCGCACGGGAAAAGCAG
GCGTTCCACAACGAACTGGCCGATCCGACCGATGGATTGATCACGCTGTCGGTGCACACCGACCATGATTTCTC
CGGCTATGACCTCGACGCTGTGATCGCCGACATCGATGTTCCAGGGACGAAGGGGCTTTTCGAAGTCGCTCGG
AGTCTGAGTGTGAACGAGAACCTGACGCTGCGCGATATCGGAAAGCTGTACGCCCAGGGCGTGTTATTGCCGC
AGTTCGTGGGTACCGCGGCTCAGGTGGCCGACCAGATCGAGGCTGCCGTCGACGGTGGAGAGGCTGATGGGT
TCCTCTTTTCGGCCGGGTATACGCCTGGCGGATTCGAGGAGTTCGCCGATCTCGTCATCCCGGAACTGCAGCG
GCGAGGGCGGTTTCGTACGGAGTACACGGGTTCGACGCTGCGTGAACATCTGGGTCTACCCGCTGATGCGAAT
CTTGTGCCCGTTCCGCGCAAGGCAGTGGGGGCGGCGAAGCTTGCGGCCGCACTCGAGCACCACCACCACCAC
CAC**TGA**GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAG
CATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCG
AATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACAC
TTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGT
CAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGAT

TAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTT
CTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGG
ATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATT
AACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACAT
TCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATAT
CAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATA
GGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGT
CAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCA
TTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATT
CATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAAT
GCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCT

>pET28b1218_T7

GGGGGGAACATTTCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGTCATGAACGTAAACGT
TGTTGGCGGAATCTCTCAATGGGACACTGTTCAGGCCGATGCGTGGTGGTCGGCTGCAGTCGGTCCAGCAACG
AACGGAGAGATCTCGATGTCGGCACATGGGCGTCCAATTCACTTGGGCGGGTTTTTGATTGCAGGGAATGTAAC
CCACAGTCATCCTTCGTGGCGTCATCCACGCAGTGATCCCGGGTTTCTCACACCGGAGTACTACCAGCACCTCG
GTAGGGTTTTCGAACGCGCGAAATTGGACTTCGTCTTCTTTGCCGACAATTCTGCGACTCCTGCCAGCTACCGC
AACGATATTCGTGACCCGCTCGCTCGCGGTACTCAGAGTGCAGCCGGGTTGGATCCCCGCTTCGTCGTTCCTGT
CGTCGCGGGTGTCACGCGCAACCTGGGGATCGTATCGACCACGTCGGCGACGTTCTACTCGCCATACGACCTC
GCCCGGAGCTTTGCCACTCTGGATCATCTGACCCACGGCCGCGTCGGTTGGAATGTTGTGACTTCCAATACGAC
CGTCGAGGCGCAGAACTTCGGGCTTGCCCGACACCTCGACCATGACGTGCGATACGACCGTGCCGAGGAATTG
CTTGAGGTCGCGTTCAGGCTGTGGGCCAGTTGGGACGATGGAGCTCTGATCCAGGATAAAGAGGCGGGTGTCT
TCGCCGACCCGGACCTGATTCACAGGCTCGATCATCACGGGGAGAACTTCGATGTTCGGGGCCCGCTGTCGGT
TCCCCGCTCACCGCAGGGACGGCCGGTCATCTTTCAAGCGGGATCATCCACCCGCGGTCGGGATTTTGCTGCG
CGCTGGGCAGAAGCGATTTTCGAGATCGACCCGACGTCTGTGGGGGCGTAAGGCCTACTACGACGACATCAAG
TCGCGAGCCTCCGACTTCGGTCGTGATCCCGACGGCGTCAAGATACTCCCGTCGTTCATTCCGTTTGTGGGTGA
GACCGAGTCGATCGCACGGGAAAAGCAGGCGTTCCACAACGAACTGGCCGATCCGACCGATGGATTGATCACG
CTGTCGGTGCACAACCGACCATGATTTCTCCGCTATGACCTCGACGCTGTGATCGCGACATCGATGTTCCAGGA
CGAAGGGGCTTTTCGAAGTCGCCTCGGAGTCTGAGTGTGAACGAGAACCTGACGCTGCGCGATATCGAAAGCT
GTACGCCCCAGGCGTGTATGGCGCAGTTCCGTGGTACCGCGGGCTAAGGTGACGACGAATCAGCTGCGTGCAA
CGGGGAAAGGCTGATGGTCTCTTTCGCCGGTATACGCCTTGGTGATTCGAAGAATTGCTGCGATTCTCGT

>pET28b1218_T7T_Reverse_Complement

GGAGAATCGTATTCGACCACCGTCGGGGGAAGTGTCTACTTCGCCATACGACCTCGCCCCGGAAGCTTTGCCA
CCTCTGGATCATCCTGACCCCACGGCCGCGTTCGTTGAAATGTTGTGACTTTCAATACGACCGTCGAGGGCGCA
GAAATTCGGGCTTGCCCGACACCTCGACCATGACGTGCGATACGACCGTGCCGAGGAATTGCTTGAGGTCCGC
GTTCAGGCTGTGGGCCAGTTGGGACGATGGAGCTCTGATCCAGGATAAAGAGGCGGGTGTCTTCGCCGACCCG
GACCTGATTCACAGGCTCGATCATCACGGGGAGAACTTCGATGTTCGGGGCCCGCTGTCGGTTCCCCGCTCAC
CGCAGGGACGGCCGGTCATCTTTCAAGCGGGATCATCCACCCGCGGTCGGGATTTTGCTGCGCGCTGGGCAGA
AGCGATTTTCGAGATCGACCCGACGTCTGTGGGGCGTAAGGCCTACTACGACGACATCAAGTCGCGAGCCTCC
GACTTCGGTCGTGATCCCGACGGCGTCAAGATACTCCCGTCGTTCATTCCGTTTGTGGGTGAGACCGAGTCGAT
CGCACGGGAAAAGCAGGCGTTCCACAACGAACTGGCCGATCCGACCGATGGATTGATCACGCTGTCGGTGCAC
ACCGACCATGATTTCTCCGGCTATGACCTCGACGCTGTGATCGCCGACATCGATGTTCCAGGGACGAAGGGGC
TTTTCGAAGTCGCTCGGAGTCTGAGTGTGAACGAGAACCTGACGCTGCGCGATATCGGAAAGCTGTACGCCCA
GGGCGTGTTATTGCCGCAGTTCGTGGGTACCGCGGCTCAGGTGGCCGACCAGATCGAGGCTGCCGTCGACGG
TGGAGAGGCTGATGGGTTCCTCTTTTCGGCCGGGTATACGCCTGGCGGATTCGAGGAGTTCGCCGATCTCGTC
ATCCCGGAACTGCAGCGGCGAGGGCGGTTTCGTACGGAGTACACGGGTTCGACGCTGCGTGAACATCTGGGTC
TACCCGCTGATGCGAATCTTGTGCCCGTTCCGCGCAAGGCAGTGGGGGCGGCGAAGCTTGCGGCCGCACTCG
AGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAACGAAGCTAGTCTGGCCCGTTCCGC
AAAT

>pET28b1222

CAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGT
CGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGC
GCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGC

CCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAAC
GCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCC
ACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACA
CCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGC
CAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATG
TAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCAC
GCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCA
CCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGG
ATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCG
CCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATAC
CCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATA
GGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTC
GATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTG
TTTAACTTTAAGAAGGAGATATACCATGGTT**ATG**GCTGATCGAGAGCTCCATCTGGGCGTCAATGTCCTCTCGGA
CGGTATGCACCCAGCCGCGTGGCAGTATCCGAGTTCCGATCCGTCGTGGTTCACGGATCCGGCGTACTGGATT
CGTGTTGCGCAGATCGCGGAGCGAGGAACCCTCGACGCGGTCTTCCTCGCCGACAGTCCGTCGTTGTTCCAGC
CGCCCGACCAGCCGCTGAGTGCGCCACCGTTGGCCCTGGACCCGATCGTGTTGTTGTCGACACTGGCATCGGT
GACCACACACATCGGACTCATCGGTACGGTGTCGACCTCGTTCGAGGAGCCGTACAACGTCGCCCGCCGATTC
TCGACGCTGGACCACCTCAGCCGGGGTCGTGTGGCATGGAACGTCGTGACGAGTAGTGATCGGTATGCCTGGA
ACAATTTCGGTGGTGGTGAACAACCCGACCGCGCTACTCGATACGAGCGGGCCGGCGAGTTCATCGAAGTCGT
CCGGGCATTGTGGGATTCGTGGGACGACGACGCAGTTGTCGCCGACAAGTCCACGGGTGCGTTCAGTAAGGTC
GGTGCGATACGACCGATCCGGCATCGCGGTGGGCACTTCTCGGTGGACGGGCCGTTGACTCTACCCAGATCCC
CACAGGGGCATCCGGTGTTGTTTCAGGCAGGCGGTTCCACCGGCGGGTTGGATCTGGCGGCGAAGTACGCCG
ACGGGGTCTTTGCGGCACAGGCCTCGCTCGAGGATGCGCTGTCCAACGCGCAGGAGCTGCGGAGTCGGTTGA
TCGCGCATGGCCGTCCCGCCGAGGCGATCCGAATCATGCCTGGCTTGTCGTTCGTGCTCGGCAGTACGGAGGC
AGAGGCCAGGTCGCGAAACGACGAATTGAACGAGCTCGCCGGGGATCGACGCCTGGCACATCTGGCTGGTCA
ACTCAGCGTCGATGTGGCGGAGCTGAAGTGGGACAAGCCGCTTCCCGGTTGGCTCCTCGAGGGCGCGGCGCC
GATCAGCGGTTCCCAGGGAGCTCGCGACATCGTCGTCAACATCGCTCGGCGGGAGAACCTGACCGTGCGTCAG
CTGCTCGATCGGGTGATCACGTGGCACCGCTTCGTGGTCGGATCGCCTGAACAGATCGCCGATGCCATCGAGG
ACTGGTTCGTTGCGGGCGCTGTCGACGGCTTCAACCTGATGCCGGATGTCTTCCCGTCGGGTCTCGAGTTGTTC
GTCGACCACGTCGTACCGATCCTCCGGGACCGAGGGTTGTTCCGGCGGGAGTACACATCGACGACATTGCGTG
GGCATCTGGGCCTCGAGCGCACCCCAGACCGGCCGTCGTCGGGTTCGATCCGCCGGACCGGTAAGCTTGCGG
CCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGC
TGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGA
AAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGT
GGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTC
TCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTA
CGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTT
TCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTAT
CTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAA
AATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAAC
CCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCA
TCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGA
GAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATC
AATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCC
GGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAA
TCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAA
GGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGA
ATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGT

>pET28b1222_T7

TTAACGTACGTTCCCTCTAGAATATTTTGTTTACTTTAAGAAGGAGATATACCATGGTTATGGCTGATCGAGAGCT
CCATCTGGGCGTCAATGTCCTCTCGGACGGTATGCACCCAGCCGCGTGGCAGTATCCGAGTTCCGATCCGTCG
TGGTTCACGGATCCGGCGTACTGGATTCGTGTTGCGCAGATCGCGGAGCGAGGAACCCTCGACGCGGTCTTCC
TCGCCGACAGTCCGTCGTTGTTCCAGCCGCCCGACCAGCCGCTGAGTGCGCCACCGTTGGCCCTGGACCCGAT
CGTGTTGTTGTCGACACTGGCATCGGTGACCACACACATCGGACTCATCGGTACGGTGTCGACCTCGTTCGAGG
AGCCGTACAACGTCGCCCGCCGATTCTCGACGCTGGACCACCTCAGCCGGGGTCGTGTGGCATGGAACGTCGT

GACGAGTAGTGATCGGTATGCCTGGAACAATTTCGGTGGTGGTGAACAACCCGACCGCGCTACTCGATACGAG
CGGGCCGGCGAGTTCATCGAAGTCGTCCGGGCATTGTGGGATTCGTGGGACGACGACGCAGTTGTCGCCGAC
AAGTCCACGGGTGCGTTCAGTAAGGTCGGTGCGATACGACCGATCCGGCATCGCGGTGGGCACTTCTCGGTGG
ACGGGCCGTTGACTCTACCCAGATCCCCACAGGGGCATCCGGTGTTGTTTCAGGCAGGCGGTTCCACCGGCGG
GTTGGATCTGGCGGCGAAGTACGCCGACGGGGTCTTTGCGGCACAGGCCTCGCTCGAGGATGCGCTGTCCAA
CGCGCAGGAGCTGCGGAGTCGGTTGATCGCGCATGGCCGTCCCGCCGAGGCGATCCGAATCATGCCTGGCTT
GTCGTTCGTGCTCGGCAGTACGGAGGCAGAGGCCAGGTCGCGAAACGACGAATTGAACGAGCTCGCCGGGGA
TCGACGCCTGGCACATCTGGCTGGTCAACTCAGCGTCGATGTGGCGGAGCTGAAGTGGGACAAGCCGCTTCCT
CGGTTGGCTCCTCGAGGGCGCGGCGCCGATCAGCGGTTCCCAGGAGCTCGCGACATCGTCGTCAACATCGCT
CGGCGGGAGAACCTGACCGTGCGTCAGCTGCTCGATCGGGTGATCACGTGGCACCGCTTCGTGGTCGGATCG
CCTGAACAGAATCGCGATGCCATCGAGGACTGGATCGTGCGGCGCTGTCGACGGCTTCACTGAATGCCGGATG
TCTCCCGTCGGTCTCGAGTGTCGTCGACACGTCGTACGATCTCAGAACCGAGAGGTCCGGCCGGAGATACCAT
CGACGACCTGGCCTGGCATCTGGACTCGAGGCACTTCAGAGAACTCG

>pET28b1222_T7T_Reverse_Complement

GGACTAGCTTCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGC
AAGCTTACCGGTCCGGCGGATCGAACCCGACGACGGCCGGTCTGGGGTGCGCTCGAGGCCCAGATGCCCACG
CAATGTCGTCGATGTGTACTCCCGCCGGAACAACCCTCGGTCCCGGAGGATCGGTACGACGTGGTCGACGAAC
AACTCGAGACCCGACGGGAAGACATCCGGCATCAGGTTGAAGCCGTCGACAGCGCCCGCAACGAACCAGTCCT
CGATGGCATCGGCGATCTGTTCAGGCGATCCGACCACGAAGCGGTGCCACGTGATCACCCGATCGAGCAGCTG
ACGCACGGTCAGGTTCTCCCGCCGAGCGATGTTGACGACGATGTCGCGAGCTCCCTGGGAACCGCTGATCGGC
GCCGCGCCCTCGAGGAGCCAACCGGGAAGCGGCTTGTCCCACTTCAGCTCCGCCACATCGACGCTGAGTTGAC
CAGCCAGATGTGCCAGGCGTCGATCCCCGGCGAGCTCGTTCAATTCGTCGTTTCGCGACCTGGCCTCTGCCTC
CGTACTGCCGAGCACGAACGACAAGCCAGGCATGATTCGGATCGCCTCGGCGGGACGGCCATGCGCGATCAA
CCGACTCCGCAGCTCCTGCGCGTTGGACAGCGCATCCTCGAGCGAGGCCTGTGCCGCAAAGACCCCGTCGGC
GTACTTCGCCGCCAGATCCAACCCGCCGGTGGAACCGCCTGCCTGAAACAACACCGGATGCCCCTGTGGGGAT
CTGGGTAGAGTCAACGGCCCGTCCACCGAGAAGTGCCCACCGCGATGCCGGATCGGTCGTATCGCACCGACCT
TACTGAACGCACCCGTGGACTTGTCGGCGACAACTGCGTCGTCGTCCCACGAATCCCACAATGCCCGGACGAC
TTCGATGAACTCGCCGGCCCGCTCGTATCGAGTAGCGCGGTCGGGTTGTTCACCACCACCGAAATTGTTCCAGG
CATACCGATCACTACTCGTCACGACGTTTCCATGCCACACGACCCCGGCTGAGTGGTCCAGCGTCGAGAATCG
GCGGGCGACGTTGTACGGCTCCTCGAACGAGGTCGACACCGTACCGATGAGTTCCGATGTGTGTGGTCACGAT
GCCAGTGTCGACAACACACGATCGGTTCAGGGTCAACGTGGCGCCACTCAGCCGCTGTCGGCGCTGAACACGA
CGAACTGTTCGGCAGAGACGCGTCAGGTTCTCGCTCCGCGATCTGGCCATCCGAATCCAGTACGCTGATCCGTT
ACCAGACGATCGGTCCTCGGAATCTGCAGCGCCTGAGTCATACGTTCCGGAGAGGACGT

>pK18mobsacB1218AB

CTTCTGTTTCTATCAGGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAAAA
GGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTC
AGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAA
AAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTC
AGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCA
CCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGG
GTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCC
CAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCC
GAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCA
GGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGC
TCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGC
CTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGA
TACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACG
CAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCG
GGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCG
GCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACATGATTACGAATTCG
AGCTCGGTACCCGGGGATCCCGCGACCCTGCCCAACCTGCGTGACCTCGGTGGCTGGCAGGCCGGCGGCGG
CAAGGCGGTCCGTCCGGGCGTGCTGTTCCGCTCGACCGACTTCTCGTCGCTCGCCGACGCCGACGTCGCCCC
GTTCGAATCACTCGGCGTCCGCACCATCTACGACCTGCGCTCGACCGCCGAGCGGGATGCGCTGCCCGATCCG
ACGCTGCCCGACGTCACCGACATCCATCTCGACGTGCTCGCCGACGCCGCGATGGCGGTCCCGGCCAACCTC

GGCAAGATCCTGGCCGACCCGAAGACGGTCGCGATGGCCAGCAAGGAACTGTCCGGCGGCAAGGCCAAGGAA
CTGATCGCCCAGACCTATCAGGGCATCGTCTCGCTGCCCAGCGCACTGCGCTCCTACCAGTCCTTCTACCGCG
GTCTGCTCGGCGACCACCCGCACCCGTCGCTGTTCCACTGCACCACCGGCAAGGACCGGACCGGCTGGGCCG
CCGCGTCGTTCCTGACCCTGATGGGCGTCTCCGCCGACGACGTCTTCCACGACTACCTGCTGACCAATGACCG
TCTGATCCCGGCACTCAAGCCGATCTTCGACGAATTCGCCGAGGCCGGTGGCGATCCCGACATCCTGCTGCCG
GTGCTCGGTGTGGACCGCGCCTACCTCGAGACCGCGCTCACCGAGGTCGACACCCGTTACGGCGGTATCGAG
GGCTACTTCACCCAGGGGCTGGGTATCGACGAGGCCGCGCAGAACAAGCTGCGTGATCTGTACCTGGTGGCTC
GCTAGAATCCTGAATCGCCGCGGTTGCATCATGATTCGATGAGAAATCGGATCATGATGCAACCGCTTTTTGGTG
ACTCGGGCCGGCTATGACAGGCGAATCCGGCCCGTCTTTTGCCGTGCATGAACGTAAACGTTGTTGGCGGAAT
CTCTCAATGGGACACTGTTCAGGCCGATGCGTGGTGGTCGGCTGCAGTCGGTCCAGCAACGAACGGAGAGATC
TCGTCTAGATCGTCCTCGATCAACTGACCAAGACATTTACCCATCATGGACAATCGGCGACGGTGCTGAACAGT
ATCGATCTTGAGGTCGATACCGGCACGGTCTTCGCTGTCGTAGGTCCCAGCGGTGCCGGAAAAACCACGCTCG
CACGGTGTATCAATCTGTTGGAGCAGCCGACTTCGGGCCGCGTCGTAGTCGGGGGTCAGGAGTTGACGGGCCT
GAAGGAATCCCAGTTGAGGTCTGCTCGGCGCAGGATTGGAACGGTCTTTCAGGCGTCGAGCGTACTGTCTCGG
AGAACTGCGGCCGGCAACGTGGCACTCCCTCTGGAGTGCCTGGGCGTGACTCCGGCGGAGACCAAGGCACGG
GTATCCGAACTGCTTGATCGCGTGGGTCTTTCGCATCGGTCTGACCACTATCCACATCAGCTGAGCGGGGGCCA
GCGTCAACGAGTGGGTATTGCCCGGGCGCTGGCATTGCGTCCGTCCGTGCTGCTCGCCGACGAGGCGACATC
GGGTTTGGACCCGGAGACGACCGCATCGATCGTCGATCTGCTCGACGAGTTGCGTCGCGACCTCGATCTGACC
ATCCTCGCCATCACTCACGACATGAGCTTCGTTCGGAATCTCGCCGACAGCGTCGCGCGGCTCGATCATGGCAA
GATCGTCGAACAAGGCGACATCGTTGATCTTCTGACCGATCCGGTATCCGAGCTGGGCAGAGGTCTGCTCCCG
AGCGTCGATGTCCCAGCGGGCCAACCCGATAGGCAGTTGTGGCGGGTTCTCTATCGGGACGCGCATGCCGCA
CGGGACTGGATCGAGCGGTTGAGCCGCGCGATTCGCGTTCCGGTGGAGCTGCAGTCGGCGGCAGTCGAAGTG
GTCAATGGAGTGCAGATCGGGCAGGCGGTCGTGTCGGTTGCTGCGCGCAACGGTCTCGACATCGCCGGACTTC
TGTCGGAGTGGGGCCTGGAAGCCATCGAGGTCTCGGATGCTAACTCGGAGAAGAAGGTCGAAAGCTTGGCACT
GGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCC
CTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGG
CGAATGGCGATAAGCTAGCTTCACGCTGCCGCAAGCACTCAGGGCGCAAGGGCTGCTAAAGGAAGCGGAACAC
GTAGAAAGCCAGTCCGCAGAAACGGTGCTGACCCCGGATGAATGTCAGCTACTGGGCTATCTGGACAAGGGAA
AACGCAAGCGCAAAGAGAAAGCAGGTAGCTTGCAGTGGGCTTACATGGCGATAGCTAGACTGGGCGGTTTTAT
GGACAGCAAGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTG
GATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGGATGAGGATCG
TTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGA
CTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTT
TTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTCCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCA
CGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGA
AGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGC
GGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACG
TACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAA
CTGTTCGCCAGGCTCAAGGCGCGGATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGC
CGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTAT
CAGGACATAGCGTTGGCTACCC

>pK18mobsacB1218AB_M13R

ATGGGCCTTGGGTACCACGGGATCCCGCGACCCTGCCCAACCTGCGTGACCTCGGTGGCTGGCAGGCCGGCG
GCGGCAAGGCGGTCCGTCCGGGCGTGCTGTTCCGCTCGACCGACTTCTCGTCGCTCGCCGACGCCGACGTCG
CCCCGTTCGAATCACTCGGCGTCCGCACCATCTACGACCTGCGCTCGACCGCCGAGCGGGATGCGCTGCCCGA
TCCGACGCTGCCCGACGTCACCGACATCCATCTCGACGTGCTCGCCGACGCCGCGATGGCGGTCCCGGCCAA
CCTCGGCAAGATCCTGGCCGACCCGAAGACGGTCGCGATGGCCAGCAAGGAACTGTCCGGCGGCAAGGCCAA
GGAACTGATCGCCCAGACCTATCAGGGCATCGTCTCGCTGCCCAGCGCACTGCGCTCCTACCAGTCCTTCTACC
GCGGTCTGCTCGGCGACCACCCGCACCCGTCGCTGTTCCACTGCACCACCGGCAAGGACCGGACCGGCTGGG
CCGCCGCGTCGTTCCTGACCCTGATGGGCGTCTCCGCCGACGACGTCTTCCACGACTACCTGCTGACCAATGA
CCGTCTGATCCCGGCACTCAAGCCGATCTTCGACGAATTCGCCGAGGCCGGTGGCGATCCCGACATCCTGCTG
CCGGTGCTCGGTGTGGACCGCGCCTACCTCGAGACCGCGCTCACCGAGGTCGACACCCGTTACGGCGGTATC
GAGGGCTACTTCACCCAAGGGCTGGGTATCGACGAGGCCGCGCAGAACAAGCTGCGTGATCTGTACCTGGTGG
CTCGCTAGAATCCTGAATCGCCGCGGTTGCATCATGATTCGATGAGAAATCGGATCATGATGCAACCGCTTTTTG
GTGACTCGGGCCGGCTATGACAGGCGATCCGGCCCGTCTTTTGCCGTGCATGAACGTAAACGTTGTTGGCGGA
ATCTCTCAATGGACACTGTTTCAGGCCGATGCGTGGTGGTCGGCTGCAGTCGGTCAACCAACGAACCGAGAGAT

CTCGTCTAGACGTCTAGATCGCCCTCGATCACCTGACCAGACATTTACCCATCATGGACAATCGGCGACTGCGC
TGAACAGATTCGATCTTGAGGTCGATACCAGCACGGTCTTTCTCTGTCGTAGGTCCCAGCGGTGCTCGAAAAAA
CATCGCTCGCACGGTGTATTCCAATCTGTTGGAACCACCCGACTTCCGGCCGCCGTCCTTATCGGGGTTCAGAA
CTTGACAGGCCCGAAGAAATTC

>pK18mobsacB1218AB_21M13_reverse_complement

GGCGGATTCTAGGACGAGTCGCCGAGCCGTGGCGATTCCGACATCTGCTGCCGTGCTCGGTGTGGAATCGCGC
CTACCTCGAGACGCTCACGAGGTCGAACACCGTACCGACGTATCGAGGGCTACTTCACCCAGGGGCCTGGTAT
CGACGAGCCGCGCAGACAAGCTGCGTGATCTGTACTGTGCTCGCTAGAATCCTGAATCGCCGCGGTTGCATCA
TGATTCGATGAGAAATCGGATCATGATGCAACCGCTTTTTGGTGACTCGGGCCGGCTATGACAGGCGAATCCGG
CCCGTCTTTTGCCGTGCATGAACGTAAACGTTGTTGGCGGAATCTCTCAATGGGACACTGTTCAGGCCGATGCG
TGGTGGTCGGCTGCAGTCGGTCCAGCAACGAACGGAGAGATCTCGTCTAGACGTCTAGATCGTCCTCGATCAA
CTGACCAAGACATTTACCCATCATGGACAATCGGCGACGGTGCTGAACAGTATCGATCTTGAGGTCGATACCGG
CACGGTCTTCGCTGTCGTAGGTCCCAGCGGTGCCGGAAAAACCACGCTCGCACGGTGTATCAATCTGTTGGAG
CAGCCGACTTCGGGCCGCGTCGTAGTCGGGGGTCAGGAGTTGACGGGCCTGAAGGAATCCCAGTTGAGGTCT
GCTCGGCGCAGGATTGGAACGGTCTTTCAGGCGTCGAGCGTACTGTCTCGGAGAACTGCGGCCGGCAACGTG
GCACTCCCTCTGGAGTGCCTGGGCGTGACTCCGGCGGAGACCAAGGCACGGGTATCCGAACTGCTTGATCGC
GTGGGTCTTTCGCATCGGTCTGACCACTATCCACATCAGCTGAGCGGGGGCCAGCGTCAACGAGTGGGTATTG
CCCGGGCGCTGGCATTGCGTCCGTCCGTGCTGCTCGCCGACGAGGCGACATCGGGTTTGGACCCGGAGACGA
CCGCATCGATCGTCGATCTGCTCGACGAGTTGCGTCGCGACCTCGATCTGACCATCCTCGCCATCACTCACGAC
ATGAGCTTCGTTCGGAATCTCGCCGACAGCGTCGCGCGGCTCGATCATGGCAAGATCGTCGAACAAGGCGACA
TCGTTGATCTTCTGACCGATCCGGTATCCGAGCTGGGCAGAGGTCTGCTCCCGAGCGTCGATGTCCCAGCGGG
CCAACCCGATAGGCAGTTGTGGCGGGTTCTCTATCGGGACGCGCATGCCGCACGGGACTGGATCGAGCGGTT
GAGCCGCGCGATTCGCGTTCCGGTGGAGCTGCAGTCGGCGGCAGTCGAAGTGGTCAATGGAGTGCAGATCGG
GCAGGCGGTCGTGTCGGTTGCTGCGCGCAACGGTCTCGACATCGCCGGACTTCTGTCGGAGTGGGGCCTGGA
AGCCATCGAGGTCTCGGATGCTAACTCGAGAGAAGTA

>pK18mobsacB1218AB_18-750

GGAGGGACGTCGCGATGGCCAGCAGGAACTGTCCGGCGGCAAGGCCAAGGAACTGATCGCCCAGACCTATCA
GGGCATCGTCTCGCTGCCCAGCGCACTGCGCTCCTACCAGTCCTTCTACCGCGGTCTGCTCGGCGACCACCCG
CACCCGTCGCTGTTCCACTGCACCACCGGCAAGGACCGGACCGGCTGGGCCGCCGCGTCGTTCCTGACCCTG
ATGGGCGTCTCCGCCGACGACGTCTTCCACGACTACCTGCTGACCAATGACCGTCTGATCCCGGCACTCAAGC
CGATCTTCGACGAATTCGCCGAGGCCGGTGGCGATCCCGACATCCTGCTGCCGGTGCTCGGTGTGGACCGCG
CCTACCTCGAGACCGCGCTCACCGAGGTCGACACCCGTTACGGCGGTATCGAGGGCTACTTCACCCAGGGGCT
GGGTATCGACGAGGCCGCGCAGAACAAGCTGCGTGATCTGTACCTGGTGGCTCGCTAGAATCCTGAATCGCCG
CGGTTGCATCATGATTCGATGAGAAATCGGATCATGATGCAACCGCTTTTTGGTGACTCGGGCCGGCTATGACA
GGCGAATCCGGCCCGTCTTTTGCCGTGCATGAACGTAAACGTTGTTGGCGGAATCTCTCAATGGGACACTGTTC
AGGCCGATGCGTGGTGGTCGGCTGCAGTCGGTCCAGCAACGAACGGAGAGATCTCGTCTAGACGTCTAGATCG
TCCTCGATCAACTGACCAAGACATTTACCCATCATGGACAATCGGCGACGGTGCTGAACAGTATCGATCTTGAG
GTCGATACCGGCACGGTCTTCGCTGTCGTAGGTCCCAGCGGTGCCGGAAAAACCACGCTCGCACGGTGTATCA
ATCTGTTGGAGCAGCCGACTTCGGGCCGCGTCGTAGTCGGGGGTCAGGAGTTGACGGGCCTGAAAGGAATCC
CAGTTGAGGTCTGCTCGGCGCAGGATTGGAAACGGTCTTTCAGGCGTCGAGCGTACTGTCTCGGAGAACTGCG
GCCGGCAACGTGGCACTCCCTCTGGAGTGCCTGGGCGTGACTCCGGCGGAGACCAAGGCACGGGTATCCGAA
CTGCTTGATCGCGTGGGTCTTTCGCATCGGTCTGACCACTATCCACATCAGCTGAGCGGGGGCCAGCGTCAAC
GAGTGGGTATTGCCCGGGCGCTGGCATTGCGTCGTCGTGCTGCTCGCCGACGAGGCGACATCGGTTGACTCG
GAGACGACCGCATCGATCGTCGATCTGCTCGACGAGTGCGTCGCGACCTCGATCTGACCATCCTCGCCATCACT
CACGAACATGACTCGTCGAATCTCGCCGACGGCGTCGCGCGGTCGATCATGCAGGATCGTCGAACAGGCACAT
CGTTGAATCTTCTGGAAACCGATTTCGGTAATTCCGAGAACGCGTG

>1218A fragment

GGATCCCGCGACCCTGCCCAACCTGCGTGACCTCGGTGGCTGGCAGGCCGGCGGCGGCAAGGCGGTCCGTC
CGGGCGTGCTGTTCCGCTCGACCGACTTCTCGTCGCTCGCCGACGCCGACGTCGCCCCGTTCGAATCACTCGG
CGTCCGCACCATCTACGACCTGCGCTCGACCGCCGAGCGGGATGCGCTGCCCGATCCGACGCTGCCCGACGT
CACCGACATCCATCTCGACGTGCTCGCCGACGCCGCGATGGCGGTCCCGGCCAACCTCGGCAAGATCCTGGC
CGACCCGAAGACGGTCGCGATGGCCAGCAAGGAACTGTCCGGCGGCAAGGCCAAGGAACTGATCGCCCAGAC
CTATCAGGGCATCGTCTCGCTGCCCAGCGCACTGCGCTCCTACCAGTCCTTCTACCGCGGTCTGCTCGGCGAC

CACCCGCACCCGTCGCTGTTCCACTGCACCACCGGCAAGGACCGGACCGGCTGGGCCGCCGCGTCGTTCCTG
ACCCTGATGGGCGTCTCCGCCGACGACGTCTTCCACGACTACCTGCTGACCAATGACCGTCTGATCCCGGCACT
CAAGCCGATCTTCGACGAATTCGCCGAGGCCGGTGGCGATCCCGACATCCTGCTGCCGGTGCTCGGTGTGGAC
CGCGCCTACCTCGAGACCGCGCTCACCGAGGTCGACACCCGTTACGGCGGTATCGAGGGCTACTTCACCCAGG
GGCTGGGTATCGACGAGGCCGCGCAGAACAAGCTGCGTGATCTGTACCTGGTGGCTCGCTAGAATCCTGAATC
GCCGCGGTTGCATCATGATTCGATGAGAAATCGGATCATGATGCAACCGCTTTTTGGTGACTCGGGCCGGCTAT
GACAGGCGAATCCGGCCCGTCTTTTGCCGTGCATGAACGTAAACGTTGTTGGCGGAATCTCTCAATGGGACACT
GTTCAGGCCGATGCGTGGTGGTCGGCTGCAGTCGGTCCAGCAACGAACGGAGAGATCTCG

>1218B fragment

TCGTCCTCGATCAACTGACCAAGACATTTACCCATCATGGACAATCGGCGACGGTGCTGAACAGTATCGATCTTG
AGGTCGATACCGGCACGGTCTTCGCTGTCGTAGGTCCCAGCGGTGCCGGAAAAACCACGCTCGCACGGTGTAT
CAATCTGTTGGAGCAGCCGACTTCGGGCCGCGTCGTAGTCGGGGGTCAGGAGTTGACGGGCCTGAAGGAATC
CCAGTTGAGGTCTGCTCGGCGCAGGATTGGAACGGTCTTTCAGGCGTCGAGCGTACTGTCTCGGAGAACTGCG
GCCGGCAACGTGGCACTCCCTCTGGAGTGCCTGGGCGTGACTCCGGCGGAGACCAAGGCACGGGTATCCGAA
CTGCTTGATCGCGTGGGTCTTTCGCATCGGTCTGACCACTATCCACATCAGCTGAGCGGGGGCCAGCGTCAAC
GAGTGGGTATTGCCCGGGCGCTGGCATTGCGTCCGTCCGTGCTGCTCGCCGACGAGGCGACATCGGGTTTGG
ACCCGGAGACGACCGCATCGATCGTCGATCTGCTCGACGAGTTGCGTCGCGACCTCGATCTGACCATCCTCGC
CATCACTCACGACATGAGCTTCGTTCGGAATCTCGCCGACAGCGTCGCGCGGCTCGATCATGGCAAGATCGTC
GAACAAGGCGACATCGTTGATCTTCTGACCGATCCGGTATCCGAGCTGGGCAGAGGTCTGCTCCCGAGCGTCG
ATGTCCCAGCGGGCCAACCCGATAGGCAGTTGTGGCGGGTTCTCTATCGGGACGCGCATGCCGCACGGGACT
GGATCGAGCGGTTGAGCCGCGCGATTCGCGTTCCGGTGGAGCTGCAGTCGGCGGCAGTCGAAGTGGTCAATG
GAGTGCAGATCGGGCAGGCGGTCGTGTCGGTTGCTGCGCGCAACGGTCTCGACATCGCCGGACTTCTGTCGG
AGTGGGGCCTGGAAGCCATCGAGGTCTCGGATGCTAACTCGGAGAAGAAGGTCGAAAGCTT

>pET23d1222AB

TGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAG
CTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGA
TGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCT
GATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCC
GCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTC
TCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCAT
CAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGC
GTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTG
TAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGA
TGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGA
AAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCC
TGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAAC
CGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATC
GGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCAT
GCGCACCCGTGGCCAGGACCCAACGCTGCCCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGA
GACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGCTAGCATGACTGGT
GGACAGCAAATGGGTCGGATCCGAATTCCACCGGCGATTTCTTCGCCCGCCGCGCCAAGCGTGGCTGACCCCT
CCTCTTGCATGTCAGACAGGAACACATCAGGTCATGAGTACCCAGACATCCCCTGAAAACGATCAGTCCGAGCA
GGGGTTCACGATTGTCAAGCGAAAGCGTTGGCCATTCATCGTCGGTGCGGTGGTTGTCATCGGCGCGATCGTG
GCCGCCTTCGTCTACAGTCGCGTGGGCGGCTCGGATTCGGCGGTGACCGAGTCCGGGGCAACTCTGAAAGTTG
CTTATCTCGAGTCGAATCCGGCTGAGAAGGCGGTGATCGACTTCATCAGGGACAACGTCGCGGCCGACTACGA
TATCAAGGTCGAGGGAACTGGTATCGCTGACAGTACACAGATCAATCGTGCGATCAGCGAAGGTGAGCTGGCC
GGAACGATCTTTCAACACGAACACTGGCTGGGCCAGGTCCTCGATGCCAATCCCGACTTCAAAGAACAGGCGG
GAACGGGTCCGTTCTTCCACTGGATCTTCGGTATCTGGTCGGACAAATACTCCTCACCGCAAGATATTCCGGAA
AATGCGCGGGTCTCGCTCTTGGCTGACCCCGCAAACCAGGCGCAAGGTCTGTGGTACCTCGAACAGGCCGGG
CTCATCAAGCTTCGCGCGGGCGCTGACGTTGCCACGTTGACACCCAAGGACATCGTCGAGAACCCCAAGAATC
TGCAGTTCACGCTACTCGACTTCGGTGCGCAGCCACGAGCGCTCAGCGAACTCGATGCCGTCGTCGGTTACGC
CGAGTCATTTGTGGCCGCCGGTATCTCGGAGGACAAGCTGATCTTCTCGCCACCGTCACCGAACCAGTTCGCAT
CTGTGTTGACCGTCGGCAGCAATTACGTGGAGTCGGACAACATCCAGAACCTGATCAAGGCGTTCAAGGATCCG
CGTGTACAGAAATTCATCGCCACCGATCCGGAGACGAAGAAGTTGATCCTGCCGGCCGATCCGACAGCGGCGA
ACATCTAGAGTCGCTCGTCCGGACGCGTCTGTCTCCGGGCCCGGCTCGGACTACCGGAACACGATGCACGTGG
TGGTGCCGTGCGCGACCAGTTTGCCGTCGGCGGAGAACACCTTGCCCTCGGCGGTCGCGGTACGGCCACCCA
CGTGGATCACGGTGCCGACACCGGTGAGTTCGCCGGCGTCGAGCGCCACCGACCGGATGTAGTTCACCTTGA

GTTCCAGGGTCGTGTAGCCGACGCCGGCGGGCAACGTCGTGTGTACCGCGCAGCCCATCACCGAGTCGAGCA
AGGTCGCACAGATCCCGCCGTGCACAGTCCCGAGCGGATTGGAGAAGTCGGGCTTCGGGGTCACCACGAAGC
GCACCTCGCCCTCCTCGATACTCGCCGGGCGCATCCCCAGCAGCCGGCCGATGCCGGGCTGATCGTGATGGG
GGGCGGCCTGCCATGCCCGTAGCAGCTCCAGGCCGGACATCTGGGTGGGATCACCGATGGGTTCGGTCGAGG
TCGTCATGGTTATGTATGGTTGCATAGAACCGCCTGATTATGAAGGTCTACATAGGAGGTGGCGGTGATGTCCG
ACGACGACTCCGGCGACATCGGCACGTCGCGTGTGGGGGACGACGGTCCGCGCGAACGGATGATCGTGCACG
CTGCCGACCTGATCGGCCGTGACGGTGTGGCGGCCACCTCGATCGGGGACGTCATCTCCGCGAGCGGGGCGC
CCCGCGGGTCCATCTACCACCATTTCCCCGGCGGGAAGACGCAGCTGGTCACCGAGGCCGTCCGCTACGCCG
GTGACTTCATCACCCGACGCATCGGAGGTCAGCATCCCGGCTCGCCGTCGGAGGCGGTCTTCGGCATCGGCGA
CGTCTGGCGTCGCATGCTGGTCAACACCGACTATCAGTTCGGGTGCCCCGTCCTGGCCGGTGGACTGTCCCGG
CGCGCCGAACCCGAGGTAGCCGACGAGTCGCAGCGCATCTTCGGCGACTGGTTGCGGCTGATCGCGACGAGC
TCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCC
GAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGT
CTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATT
AAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTC
GCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGG
TTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCG
CCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGA
ACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAA
TGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGG
GAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACC
CTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTT
TTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGG
GTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGT
TTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAA
CTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGAT
GGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACA
ACGATCGGAG

>pET23d1222AB_T7

ACTTCTCTAGATATTTTGTTTACTTTAAGAAGGAGATATACCATGGCTAGCATGACTGGTGGACAGCAAATGGGT
CGGATCCGAATTCCACCGGCGATTTCTTCGCCCGCCGCGCCAAGCGTGGCTGACCCCTCCTCTTGCATGTCAG
ACAGGAACACATCAGGTCATGAGTACCCAGACATCCCCTGAAAACGATCAGTCCGAGCAGGGGTTCACGATTGT
CAAGCGAAAGCGTTGGCCATTCATCGTCGGTGCGGTGGTTGTCATCGGCGCGATCGTGGCCGCCTTCGTCTAC
AGTCGCGTGGGCGGCTCGGATTCGGCGGTGACCGAGTCCGGGGCAACTCTGAAAGTTGCTTATCTCGAGTCGA
ATCCGGCTGAGAAGGCGGTGATCGACTTCATCAGGGACAACGTCGCGGCCGACTACGATATCAAGGTCGAGGG
AACTGGTATCGCTGACAGTACACAGATCAATCGTGCGATCAGCGAAGGTGAGCTGGCCGGAACGATCTTTCAAC
ACGAACACTGGCTGGGCCAGGTCCTCGATGCCAATCCCGACTTCAAAGAACAGGCGGGAACGGGTCCGTTCTT
CCACTGGATCTTCGGTATCTGGTCGGACAAATACTCCTCACCGCAAGATATTCCGGAAAATGCGCGGGTCTCGC
TCTTGGCTGACCCCGCAAACCAGGCGCAAGGTCTGTGGTACCTCGAACAGGCCGGGCTCATCAAGCTTCGCGC
GGGCGCTGACGTTGCCACGTTGACACCCAAGGACATCGTCGAGAACCCCAAGAATCTGCAGTTCACGCTACTC
GACTTCGGTGCGCAGCCACGAGCGCTCAGCGAACTCGATGCCGTCGTCGGTTACGCCGAGTCATTTGTGGCCG
CCGGTATCTCGGAGGACAAGCTGATCTTCTCGCCACCGTCACCGAACCAGTTCGCATCTGTGTTGACCGTCGGC
AGCAATTACGTGGAGTCGGACAACATCCAGAACCTGATCAAGGCGTTCAAGGATCCGCGTGTACAGAAATTCAT
CGCCACCGATCCGGAGACGAAGAAGTTGATCCTGCCGGCTCGATCCGACAGCGCGAACATCTAAGAGTCGCTC
GTCCGGACGCGTCTGTCTCCGGGCCCGGCTCGGACTACTGAACACGATGCACGTGGTGGTGCCGTGCGCGAC
TAGGTTTGCCGTCGGCGGAAAATCACTTGCCCTCGCGGTCGCGTTACGGGCACCCACGTGATCACGGTGCGAA
CACCGGGTAGTTCGCCGGCTTCAGCGCCACCGAACGGATGTAGTCCAACTGATTCCAGATTCGTTAGCCGAAG
CTGCGGCAACGTCGTGTTACGGCGCAATCAATACGAGTTCGAGCAAGGTTCGCACAGAATCTGCAGTCAGTTCC
CGATCGGAATGGGAGAAGATCTGGCCTA

>pET23d1222AB_22-750

CCGGGCCTGGAAGGCGGTGATCGACTTCATCAGGGACAACGTCGCGGCCGACTACGATATCAAGGTCGAGGG
ACCTGGTATCGCTGACAGTACACAGATCAATCGTGCGATCAGCGAAGGTGAGCTGGCCGGAACGATCTTTCAAC
ACGAACACTGGCTGGGCCAGGTCCTCGATGCCAATCCCGACTTCAAAGAACAGGCGGGAACGGGTCCGTTCTT
CCACTGGATCTTCGGTATCTGGTCGGACAAATACTCCTCACCGCAAGATATTCCGGAAAATGCGCGGGTCTCGC
TCTTGGCTGACCCCGCAAACCAGGCGCAAGGTCTGTGGTACCTCGAACAGGCCGGGCTCATCAAGCTTCGCGC
GGGCGCTGACGTTGCCACGTTGACACCCAAGGACATCGTCGAGAACCCCAAGAATCTGCAGTTCACGCTACTC

GACTTCGGTGCGCAGCCACGAGCGCTCAGCGAACTCGATGCCGTCGTCGGTTACGCCGAGTCATTTGTGGCCG
CCGGTATCTCGGAGGACAAGCTGATCTTCTCGCCACCGTCACCGAACCAGTTCGCATCTGTGTTGACCGTCGGC
AGCAATTACGTGGAGTCGGACAACATCCAGAACCTGATCAAGGCGTTCAAGGATCCGCGTGTACAGAAATTCAT
CGCCACCGATCCGGAGACGAAGAAGTTGATCCTGCCGGCCGATCCGACAGCGGCGAACATCTAGAGTCGCTCG
TCCGGACGCGTCTGTCTCCGGGCCCGGCTCGGACTACCGGAACACGATGCACGTGGTGGTGCCGTGCGCGAC
CAGTTTGCCGTCGGCGGAGAACACCTTGCCCTCGGCGGTCGCGGTACGGCCACCCACGTGGATCACGGTGCC
GACACCGGTGAGTTCGCCGGCGTCGAGCGCCACCGACCGGATGTAGTTCACCTTGAGTTCCAGGGTCGTGTAG
CCGACGCCGGCGGGCAACGTCGTGTGTACCGCGCAGCCCATCACCGAAGTCGAGCAAGTCGCACAGATCCGC
CGTGCCACAGTCCCGAGCGGAATGGAAAAGTCGGCTTCGGGGTCACCACGAACGACCTCGCCTCCTCGATACT
CGCCGGGGCATCTCTACAGCGGCGATGCGGCTGATCTGGATGGGGGCGGCTGCATGCCCTACACTCAGCGGA
CTCTGGTGGATCACCGAATGGTCGGTCAGTCGTCATGGTAGTATGGTGCATAACCCTGAATTAAAAGTCTAATAG
AAGGTGGCGTGAGTTCCAAACATCGCACTCGACTCTCGTGGAACAGGCTCCGGAAGGATACTGAACTTGCACGT
TGCGTTACGTGCGACTATGGACTCCTCTGAGG

>pET23d_T7T_reverse complement

CCGGATCCGACAAGCGGCGAACATCAAATTTGTTCTTCGGACGCGTTTATAATCGGGCCGACTTCGGACTACCG
GACCCCGATGCACCTGGTGGTGCGTGCGCGGACCAGTTGCTGTGGGCGGAGAACACCTTGCCCTCGGCGGTC
GCGATACAGCCCACCCACGTGGATCACGGTGCCGACACCGTTGAGTTCGCCGGCGTCGAAGCGCCACCGACC
GGATGTAGTTCACCCTTGAGTTCCAGGGTCGTGTAGCCGACGCCGGCGGGCAACGTCGTGTGTACCGCGCAGC
CCATCACCGAGTCGAGCAAAGGTCGCACAGATCCCGCCGTGCACAGTCCCGAGCGGATTGGAGAAAGTCGGG
CTTCGGGGTCACCACGAAGCGCACCTCGCCCTCCTCGATACTCGCCCGGGCGCATCCCCAGCAGCCGGCCGA
TGCCGGGCTGATCGTGATGGGGGGCGGCCTGCCATGCCCGTAGCAGCTCCAGGCCGGACATCTGGGTGGGAT
CACCGATGGGTTCGGTCGAGGTCGTCATGGTTATGTATGGTTGCATAGAACCGCCTGATTATGAAGGTCTACATA
GGAGGTGGCGGTGATGTCCGACGACGACTCCGGCGACATCGGCACGTCGCGTGTGGGGGACGACGGTCCGC
GCGAACGGATGATCGTGCACGCTGCCGACCTGATCGGCCGTGACGGTGTGGCGGCCACCTCGATCGGGGACG
TCATCTCCGCGAGCGGGGCGCCCCGCGGGTCCATCTACCACCATTTCCCCGGCGGGAAGACGCAGCTGGTCA
CCGAGGCCGTCCGCTACGCCGGTGACTTCATCACCCGACGCATCGGAGGTCAGCATCCCGGCTCGCCGTCGG
AGGCGGTCTTCGGCATCGGCGACGTCTGGCGTCGCATGCTGGTCAACACCGACTATCAGTTCGGGTGCCCCGT
CCTGGCCGGTGGACTGTCCCGGCGCGCCGAACCCGAGGTAGCCGACGAGTCGCAGCGCATCTTCGGCGACTG
GTTGCGGCTGATCGCGACGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGA
GATCCGGCTGCTAACAAAGCCCGAAAGAAAAGGCGAAGGTC

>pK18mobsacB1222AB

GTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAGGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGG
AGTTCTTCGCCCACCCCAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGA
GTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGT
AATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCT
TTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCA
CCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG
GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTA
TGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGA
GAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACT
TGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTAC
GGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTAT
TACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGA
AGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGAC
AGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCA
GGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAG
CTATGACATGATTACGAATTCCACCGGCGATTTCTTCGCCCGCCGCGCCAAGCGTGGCTGACCCCTCCTCTTGC
ATGTCAGACAGGAACACATCAGGTCATGAGTACCCAGACATCCCCTGAAAACGATCAGTCCGAGCAGGGGTTCA
CGATTGTCAAGCGAAAGCGTTGGCCATTCATCGTCGGTGCGGTGGTTGTCATCGGCGCGATCGTGGCCGCCTT
CGTCTACAGTCGCGTGGGCGGCTCGGATTCGGCGGTGACCGAGTCCGGGGCAACTCTGAAAGTTGCTTATCTC
GAGTCGAATCCGGCTGAGAAGGCGGTGATCGACTTCATCAGGGACAACGTCGCGGCCGACTACGATATCAAGG
TCGAGGGAACTGGTATCGCTGACAGTACACAGATCAATCGTGCGATCAGCGAAGGTGAGCTGGCCGGAACGAT
CTTTCAACACGAACACTGGCTGGGCCAGGTCCTCGATGCCAATCCCGACTTCAAAGAACAGGCGGGAACGGGT

CCGTTCTTCCACTGGATCTTCGGTATCTGGTCGGACAAATACTCCTCACCGCAAGATATTCCGGAAAATGCGCG
GGTCTCGCTCTTGGCTGACCCCGCAAACCAGGCGCAAGGTCTGTGGTACCTCGAACAGGCCGGGCTCATCAAG
CTTCGCGCGGGCGCTGACGTTGCCACGTTGACACCCAAGGACATCGTCGAGAACCCCAAGAATCTGCAGTTCA
CGCTACTCGACTTCGGTGCGCAGCCACGAGCGCTCAGCGAACTCGATGCCGTCGTCGGTTACGCCGAGTCATT
TGTGGCCGCCGGTATCTCGGAGGACAAGCTGATCTTCTCGCCACCGTCACCGAACCAGTTCGCATCTGTGTTGA
CCGTCGGCAGCAATTACGTGGAGTCGGACAACATCCAGAACCTGATCAAGGCGTTCAAGGATCCGCGTGTACA
GAAATTCATCGCCACCGATCCGGAGACGAAGAAGTTGATCCTGCCGGCCGATCCGACAGCGGCGAACATCTAG
AGTCGCTCGTCCGGACGCGTCTGTCTCCGGGCCCGGCTCGGACTACCGGAACACGATGCACGTGGTGGTGCC
GTGCGCGACCAGTTTGCCGTCGGCGGAGAACACCTTGCCCTCGGCGGTCGCGGTACGGCCACCCACGTGGAT
CACGGTGCCGACACCGGTGAGTTCGCCGGCGTCGAGCGCCACCGACCGGATGTAGTTCACCTTGAGTTCCAGG
GTCGTGTAGCCGACGCCGGCGGGCAACGTCGTGTGTACCGCGCAGCCCATCACCGAGTCGAGCAAGGTCGCA
CAGATCCCGCCGTGCACAGTCCCGAGCGGATTGGAGAAGTCGGGCTTCGGGGTCACCACGAAGCGCACCTCG
CCCTCCTCGATACTCGCCGGGCGCATCCCCAGCAGCCGGCCGATGCCGGGCTGATCGTGATGGGGGGCGGCC
TGCCATGCCCGTAGCAGCTCCAGGCCGGACATCTGGGTGGGATCACCGATGGGTTCGGTCGAGGTCGTCATGG
TTATGTATGGTTGCATAGAACCGCCTGATTATGAAGGTCTACATAGGAGGTGGCGGTGATGTCCGACGACGACT
CCGGCGACATCGGCACGTCGCGTGTGGGGGACGACGGTCCGCGCGAACGGATGATCGTGCACGCTGCCGACC
TGATCGGCCGTGACGGTGTGGCGGCCACCTCGATCGGGGACGTCATCTCCGCGAGCGGGGCGCCCCGCGGG
TCCATCTACCACCATTTCCCCGGCGGGAAGACGCAGCTGGTCACCGAGGCCGTCCGCTACGCCGGTGACTTCA
TCACCCGACGCATCGGAGGTCAGCATCCCGGCTCGCCGTCGGAGGCGGTCTTCGGCATCGGCGACGTCTGGC
GTCGCATGCTGGTCAACACCGACTATCAGTTCGGGTGCCCCGTCCTGGCCGGTGGACTGTCCCGGCGCGCCGA
ACCCGAGGTAGCCGACGAGTCGCAGCGCATCTTCGGCGACTGGTTGCGGCTGATCGCGACGAGCTCCGTCGA
CCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAA
CTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTC
CCAACAGTTGCGCAGCCTGAATGGCGAATGGCGATAAGCTAGCTTCACGCTGCCGCAAGCACTCAGGGCGCAA
GGGCTGCTAAAGGAAGCGGAACACGTAGAAAGCCAGTCCGCAGAAACGGTGCTGACCCCGGATGAATGTCAGC
TACTGGGCTATCTGGACAAGGGAAAACGCAAGCGCAAAGAGAAAGCAGGTAGCTTGCAGTGGGCTTACATGGC
GATAGCTAGACTGGGCGGTTTTATGGACAGCAAGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGGT
TGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCT
GATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTG
GGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTG
TCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTCCAAGACGAGG
CAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGG
GAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAA
AGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAG
CGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGA
GCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGGATGCCCGACGGCGAGGATCTCGT
CGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTG
GCCGGCTGGGTGTGGCGGACCGC

>pK18mobsacB1222AB_M13R

ACCGTTACGTCCGTATTTCTTCCCCCGCCGCGCCAAGCGTGGCTGACCCTCCTCTTGCATGTCAGACAGGAACA
CATCAGGAGATGAGTACCCAGACATCCCCTGAAAACGATCAGTCCGATCAGGGGTTCACGATTGTCAAGCGAAA
GCGTTGGCCATTCATCGTCGGTGCGGTGGTTGTCATCGGCGCGATCGTGGCCGCCTTCATCTACAGTCGCGTG
GGCGGCTCGGATTCGGCGGTGACCGAGTCCGGGGCAACTCTGAAAGTTGCTTATCTCGAGTCGAATCCGGCTG
AGAAGGCGGTGATCGACTTCATCAGGGACAACGTCGCGGCCGACTACGATATCAAGGTCGAGGGAACTGGTAT
CGCTGACAGTACACAAATCAATCGTGCGATCAGCGAAGGTGAGCTGGCCGGAACGATCTTTCAACACGAACACT
GGCTGGGCCAGGTCCTCGATGCCAATCCCGACTTCAAAGAACAGGCGGGAACGGGTCCGTTCTTCCACTGGAT
CTTCGGTATCTGGTCGGACAAATACTCCTCACCGCAAGATATTCCGGAAAATGCGCGGGTCTCGCTCTTGGCTG
ACCCCGCAAACCAGGCGCAAGGTCTGTGGTACCTCGAACAGGCCGGGCTCAT

>pK18mobsacB1222AB_21M13_reverse_complement

GTACCTCCTCGGACACATGGCTCACGAGCTCACTCAGTCCGCCTCTAGGCGTCGTACGTGTACCGTTGACCCCA
AGGAACATCGTGCGAGAAACCCCAAGAATTGGCAAGTCACGCTATCGAATTTCGTGGCCAGCCCACGAGCGCT
CAGCGAACTCGGATGCCGTCGTCGTTACGCCGAGTCATTTGTGCGGCCCGGTATTTCGGAGGAACAAAGCTGA
TCTTCTCGCCCACCGTCACCCGATCAGTTCGCAATCTCTTGTGAACGGTCGCCAGCAATTACGTGGAGTCCGAA
CAACATCCAGAACCTGATCCAAAGCGGTTCAAGGATCCGGCGTGTACAAGAAATCCATCCGCCCACCGATCCGG

AGAACGAAGAAGTTGATCCTGCCGGCCGATCCGACAGCGGGCGAACATCTAGAGTCGCTCGTCCGGACGCGTC
TGTCTCCCGGGCCCGGCTCGGACTACCGGAACACGATGCACGTGGTGGTGCCGTGCGCGAACCAGTTTGCCGT
CGGCGGAGAACACCTTGCCCTCGGCGGTCGCGGTACGGCCACCCACGTGGATCACGGTGCCGACACCGGTGA
GTTCGCCGGCGTCGAGCGCCACCGACCGGATGTAGTTCACCTTGAGTTCCAGGGTCGTGTAGCCGACGCCGGC
GGGCAACGTCGTGTGTACCGCGCAGCCCATCACCGAGTCGAGCAAGGTCGCACAGATCCCGCCGTGCACAGT
CCCGAGCGGATTGGAGAAGTCGGGCTTCGGGGTCACCACGAAGCGCACCTCGCCCTCCTCGATACTCGCCGG
GCGCATCCCCAGCAGCCGGCCGATGCCGGGCTGATCGTGATGGGGGGCGGCCTGCCATGCCCGTAGCAGCTC
CAGGCCGGACATCTGGGTGGGATCACCGATGGGTTCGGTCGAGGTCGTCATGGTTATGTATGGTTGCATAGAA
CCGCCTGATTATGAAGGTCTACATAGGAGGTGGCGGTGATGTCCGACGACGACTCCGGCGACATCGGCACGTC
GCGTGTGGGGGACGACGGTCCGCGCGAACGGATGATCGTGCACGCTGCCGACCTGATCGGCCGTGACGGTGT
GGCGGCCACCTCGATCGGGGACGTCATCTCCGCGAGCGGGGCGCCCCGCGGGTCCATCTACCACCATTTCCC
CGGCGGGAAGACGCAGCTGGTCACCGAGGCCGTCCGCTACGCCGGTGACTTCATCACCCGACGCATCGGAGG
TCAGCATCCCGGCTCGCCGTCGGAGGCGGTCTTCGGCATCGGCGACGTCTGGCGTCGCATGCTGGTCAACACC
GACTATCAGTTCGGGTGCCCCGTCCTGGCCGGTGGACTGTCCCGGCGCGCCGAACCCGAGGTAGCCGACGAG
TCGCAGCGCATCTTCGGCGACTGGTTGCGGCTGATCGCGACGAGCTCCGTCGACTGCCGACCCC

>pK18mobsacB1222AB_22-750

GCCGGAAAGCGGTGATCGACTTCATCAGGGACAACGTCGCGGCCGACTACGATATCAAGGTCGAGGGAACTGG
TATCGCTGACAGTACACAGATCAATCGTGCGATCAGCGAAGGTGAGCTGGCCGGAACGATCTTTCAACACGAAC
ACTGGCTGGGCCAGGTCCTCGATGCCAATCCCGACTTCAAAGAACAGGCGGGAACGGGTCCGTTCTTCCACTG
GATCTTCGGTATCTGGTCGGACAAATACTCCTCACCGCAAGATATTCCGGAAAATGCGCGGGTCTCGCTCTTGG
CTGACCCCGCAAACCAGGCGCAAGGTCTGTGGTACCTCGAACAGGCCGGGCTCATCAAGCTTCGCGCGGGCG
CTGACGTTGTCACGTTGACACCCAAGGACATCGTCGAGAACCCCAAGAATCTGCAGTTCACGCTACTCGACTTC
GGTGCGCAGCCACGAGCGCTCAGCGAACTCGATGCCGTCGTCGGTTACGCCGAGTCATTTGTGGCCGCCGGTA
TCTCGGAGGACAAGCTGATCTTCTCGCCACCGTCACCGAACCAGTTCGCATCTGTGTTGACCGTCGGCAGCAAT
TACGTGGAGTCGGACAACATCCAGAACCTGATCAAGGCGTTCAAGGATCCGCGTGTACAGAAATTCATCGCCAC
CGATCCGGAGACGAAGAAGTTGATCCTGCCGGCCGATCCGACAGCGGCGAACATCTAGAGTCGCTCGTCCGGA
CGCGTCTGTCTCCGGGCCCGGCTCGGACTACCGGAACACGATGCACGTGGTGGTGCCGTGCGCGACCAGTTT
GCCGTCGGCGGAGAACACCTTGCCCTCGGCGGTCGCGGTACGGCCACCCACGTGGATCACGGTGCCGACACC
GGTGAGTTCGCCGGCGTCGAGCGCCACCGACCGGATGTAGTTCACCTTGAGTTCCAGGGTCGTGTAGCCGACG
CCGGCGGGCAACGTCGTGTGTACCGCGCAGCCCATCACCGAGTCGAGCAAGGTCGCACAGATCCCGCCGTGC
ACAGTCCCGAGCGGATTGGAGAAGTCGGGCTTCGGGGTCACCACGAAGCGCACCTCGCCCTCCTCGATACTCG
CCGGGCGCATCCTCAGCAGCCGGCCGATGCCGGCTGATCGTGATGGGGGGCGGCCTGCCATGCCCGTAGCAG
CTCCAGGCCGACATCTGGTGGATCACCGATGGGTTCGTCGAGTCGTCATGTTATGTATGGTTGCATAGACGCTG
ATTATGAGTCTTACATAGGAGGTGGGCGGTGATGTCGACGACGACTCGCGAACTTCGCACGTCCCGTGTGGGG
GACGACGTTCCGCCGAACGATGATCGGCAGCCTTGCAACTTGATCGCCGGAACGGTGTGGCGCCACCTCGATC
GGACGTCATCTCCGACGCCTCTCGCGGGTCCAATCAACATCAT

>1222A fragment

GAATTCCACCGGCGATTTCTTCGCCCGCCGCGCCAAGCGTGGCTGACCCCTCCTCTTGCATGTCAGACAGGAA
CACATCAGGTCATGAGTACCCAGACATCCCCTGAAAACGATCAGTCCGAGCAGGGGTTCACGATTGTCAAGCGA
AAGCGTTGGCCATTCATCGTCGGTGCGGTGGTTGTCATCGGCGCGATCGTGGCCGCCTTCGTCTACAGTCGCG
TGGGCGGCTCGGATTCGGCGGTGACCGAGTCCGGGGCAACTCTGAAAGTTGCTTATCTCGAGTCGAATCCGGC
TGAGAAGGCGGTGATCGACTTCATCAGGGACAACGTCGCGGCCGACTACGATATCAAGGTCGAGGGAACTGGT
ATCGCTGACAGTACACAGATCAATCGTGCGATCAGCGAAGGTGAGCTGGCCGGAACGATCTTTCAACACGAACA
CTGGCTGGGCCAGGTCCTCGATGCCAATCCCGACTTCAAAGAACAGGCGGGAACGGGTCCGTTCTTCCACTGG
ATCTTCGGTATCTGGTCGGACAAATACTCCTCACCGCAAGATATTCCGGAAAATGCGCGGGTCTCGCTCTTGGC
TGACCCCGCAAACCAGGCGCAAGGTCTGTGGTACCTCGAACAGGCCGGGCTCATCAAGCTTCGCGCGGGCGC
TGACGTTGCCACGTTGACACCCAAGGACATCGTCGAGAACCCCAAGAATCTGCAGTTCACGCTACTCGACTTCG
GTGCGCAGCCACGAGCGCTCAGCGAACTCGATGCCGTCGTCGGTTACGCCGAGTCATTTGTGGCCGCCGGTAT
CTCGGAGGACAAGCTGATCTTCTCGCCACCGTCACCGAACCAGTTCGCATCTGTGTTGACCGTCGGCAGCAATT
ACGTGGAGTCGGACAACATCCAGAACCTGATCAAGGCGTTCAAGGATCCGCGTGTACAGAAATTCATCGCCACC
GATCCGGAGACGAAGAAGTTGATCCTGCCGGCCGATCCGACAGCGGCGAACA

>1222B fragment

GTCGCTCGTCCGGACGCGTCTGTCTCCGGGCCCGGCTCGGACTACCGGAACACGATGCACGTGGTGGTGCCG
TGCGCGACCAGTTTGCCGTCGGCGGAGAACACCTTGCCCTCGGCGGTCGCGGTACGGCCACCCACGTGGATC
ACGGTGCCGACACCGGTGAGTTCGCCGGCGTCGAGCGCCACCGACCGGATGTAGTTCACCTTGAGTTCCAGG
GTCGTGTAGCCGACGCCGGCGGGCAACGTCGTGTGTACCGCGCAGCCCATCACCGAGTCGAGCAAGGTCGCA
CAGATCCCGCCGTGCACAGTCCCGAGCGGATTGGAGAAGTCGGGCTTCGGGGTCACCACGAAGCGCACCTCG
CCCTCCTCGATACTCGCCGGGCGCATCCCCAGCAGCCGGCCGATGCCGGGCTGATCGTGATGGGGGGCGGCC
TGCCATGCCCGTAGCAGCTCCAGGCCGGACATCTGGGTGGGATCACCGATGGGTTCGGTCGAGGTCGTCATGG
TTATGTATGGTTGCATAGAACCGCCTGATTATGAAGGTCTACATAGGAGGTGGCGGTGATGTCCGACGACGACT
CCGGCGACATCGGCACGTCGCGTGTGGGGGACGACGGTCCGCGCGAACGGATGATCGTGCACGCTGCCGACC
TGATCGGCCGTGACGGTGTGGCGGCCACCTCGATCGGGGACGTCATCTCCGCGAGCGGGGCGCCCCGCGGG
TCCATCTACCACCATTTCCCCGGCGGGAAGACGCAGCTGGTCACCGAGGCCGTCCGCTACGCCGGTGACTTCA
TCACCCGACGCATCGGAGGTCAGCATCCCGGCTCGCCGTCGGAGGCGGTCTTCGGCATCGGCGACGTCTGGC
GTCGCATGCTGGTCAACACCGACTATCAGTTCGGGTGCCCCGTCCTGGCCGGTGGACTGTCCCGGCGCGCCGA
ACCCGAGGTAGCCGACGAGTCGCAGCGCATCTTCGGCGACTGGTTGCGGCTGATCGCGACGAGCTC

## 7.8 Vectors maps of plasmids in this study

Below are the vector maps produced during this study. Vector maps were produced using the GenSmart Design (GeneScript, Piscataway, NJ) DNA construct design program. Annotation were made by the GenSmart Design software and *E. coli* or *Gordonia* NB4-1Y specific DNA sequences were annotated separately.

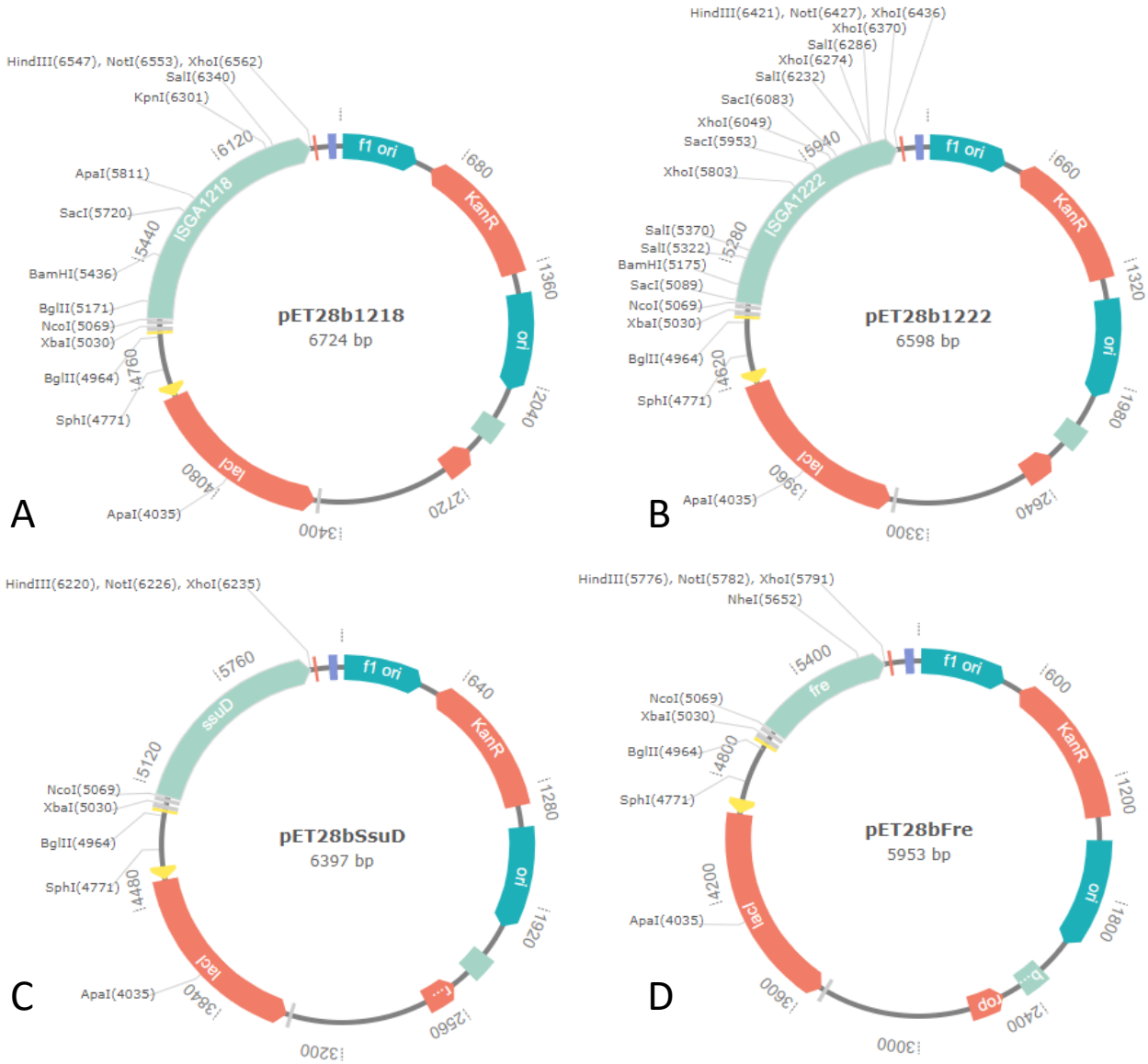


Figure S19. Vectors maps of pET28b1218 (A), pET28b1222 (B), pET28bSsuD (C) and pET28bFre (D). Vector backbones are the pET28b plasmid (Merck, Dramstadt, Germany).

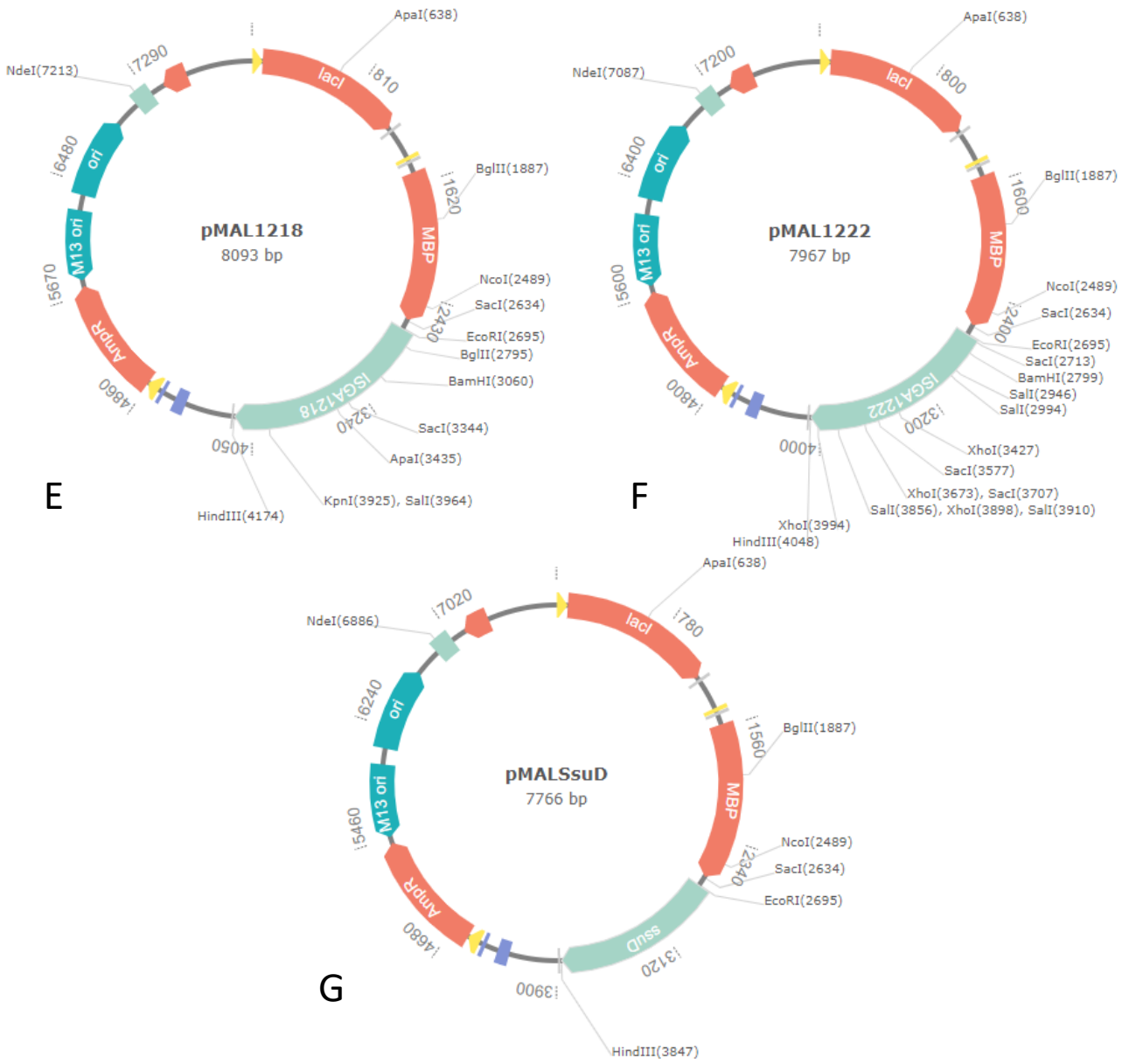


Figure S20. Vectors maps of pMAL1218 (E), pMAL1222 (F) and pMALSsuD (G). Vector backbones are the pMAL-c2 plasmid (New England BioLabs, Ipswich, MA)

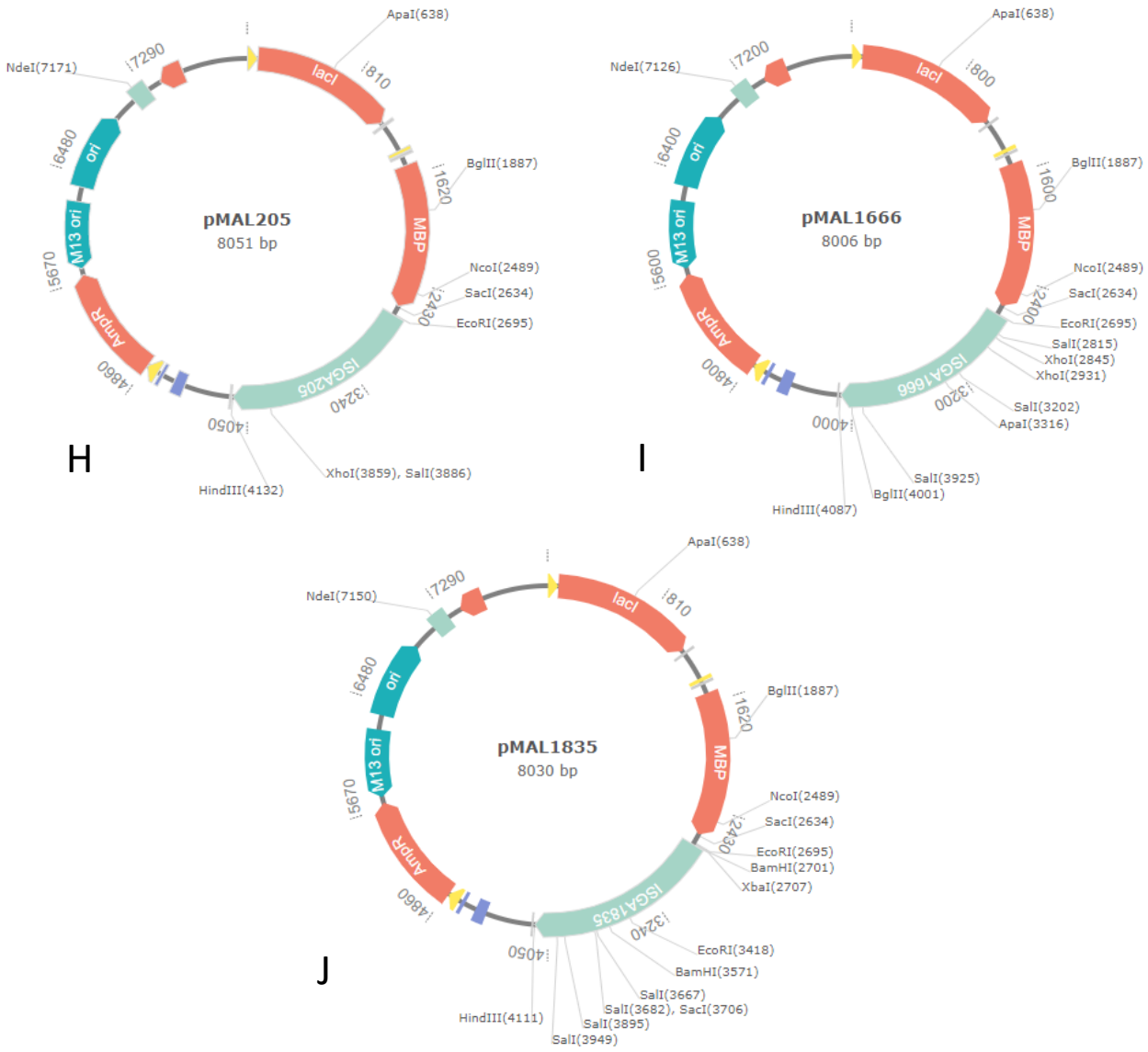


Figure S21. Vectors maps of pMAL205 (H), pMAL1666 (I), pMAL1835 (J). Vector backbones are the pMAL-c2 plasmid (New England BioLabs, Ipswich, MA).

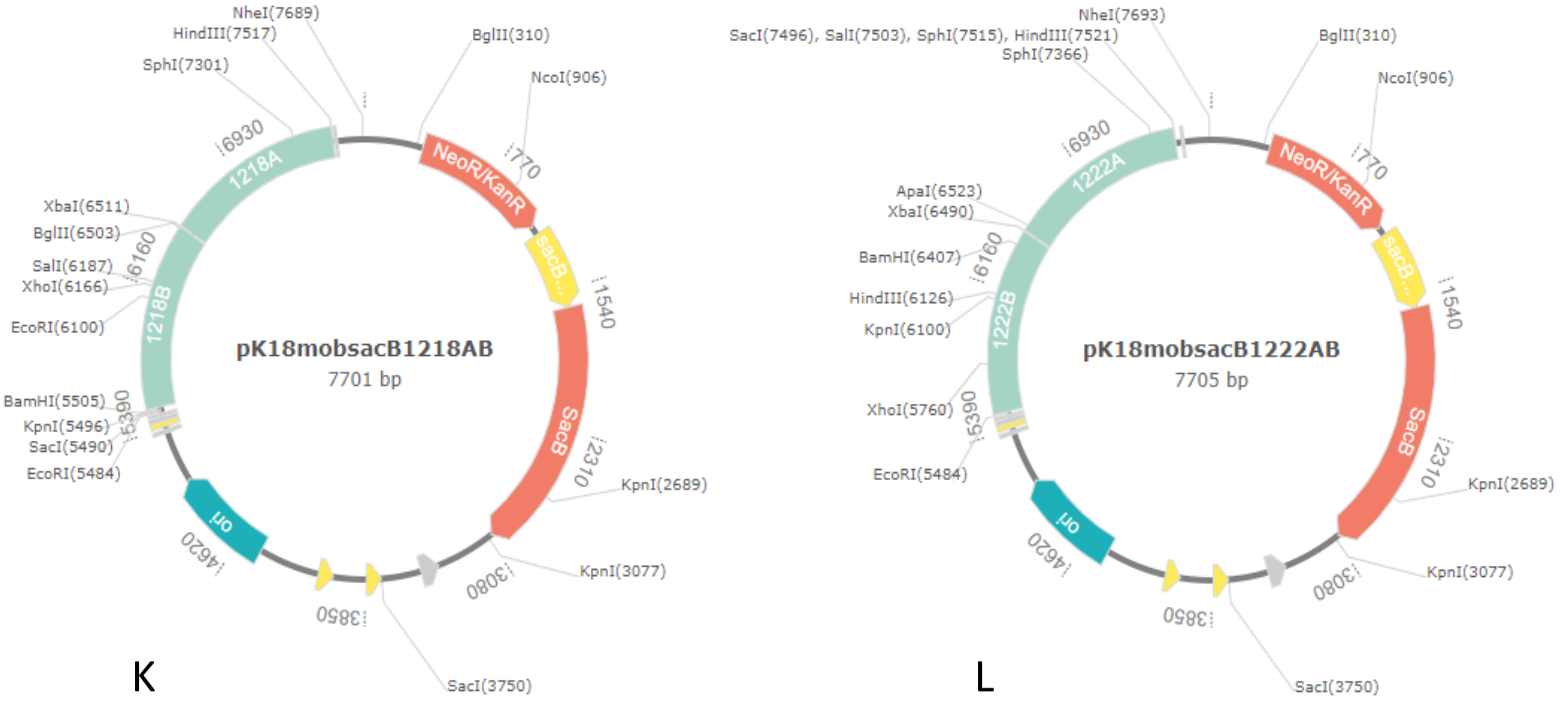


Figure S22. Vectors maps of pK18mobsacB1218AB (K) and pK18mobsacB1222AB (L). Vector backbones are the pK18mobsacB plasmid (Schafer *et al*. 1994).

## 7.9 Amino acid sequence of *Gordonia* NB4-1Y and Class C monooxygenases

Below are the amino acids sequences used to place candidate *Gordonia* NB4-1Y enzymes among annotated Class C monooxygenases. Amino acid sequences were converted to FASTA format and labeled with the following paradigm: protein_genus_species_strain. For example, the long-chain alkane monooxygenase form *Geobacillus thermodinitrificans* NG80-2 would be: LadA_*Geobacillus_thermodinitifians*_NG80-2.

>ISGA1218
MNVNVVGGISQWDTVQADAWWSAAVGPATNGEISMSAHGRPIHLGGFLIAGNVTHSHPSWRHPRSDPGFLTPEYY
QHLGRVFERAKLDFVFFADNSATPASYRNDIRDPLARGTQSAAGLDPRFVVPVVAGVTRNLGIVSTTSATFYSPYDLA
RSFATLDHLTHGRVGWNVVTSNTTVEAQNFGLARHLDHDVRYDRAEELLEVAFRLWASWDDGALIQDKEAGVFADP
DLIHRLDHHGENFDVRGPLSVPRSPQGRPVIFQAGSSTRGRDFAARWAEAIFEIDPTSVGRKAYYDDIKSRASDFGRD
PDGVKILPSFIPFVGETESIAREKQAFHNELADPTDGLITLSVHTDHDFSGYDLDAVIADIDVPGTKGLFEVARSLSVNE
NLTLRDIGKLYAQGVLLPQFVGTAAQVADQIEAAVDGGEADGFLFSAGYTPGGFEEFADLVIPELQRRGRFRTEYTGS
TLREHLGLPADANLVPVPRKAVGAA

>ISG1222
MADRELHLGVNVLSDGMHPAAWQYPSSDPSWFTDPAYWIRVAQIAERGTLDAVFLADSPSLFQPPDQPLSAPPLAL
DPIVLLSTLASVTTHIGLIGTVSTSFEEPYNVARRFSTLDHLSRGRVAWNVVTSSDRYAWNNFGGGEQPDRATRYERA
GEFIEVVRALWDSWDDDAVVADKSTGAFSKVGAIRPIRHRGGHFSVDGPLTLPRSPQGHPVLFQAGGSTGGLDLAAK
YADGVFAAQASLEDALSNAQELRSRLIAHGRPAEAIRIMPGLSFVLGSTEAEARSRNDELNELAGDRRLAHLAGQLSV
DVAELKWDKPLPGWLLEGAAPISGSQGARDIVVNIARRENLTVRQLLDRVITWHRFVVGSPEQIADAIEDWFVAGAVD
GFNLMPDVFPSGLELFVDHVVPILRDRGLFRREYTSTTLRGHLGLERTPDRPSSGSIRRTG

>ISGA205
MPTTPHRPLILNATDMATANHIAFGLWRLADPDKPDYTTLRFWTDLAIELEQSGFDALFLTDALGQLDTYTASADPALR
TATQTPLDDPLLAVSAMAAVTEQLGFAVTVSATYEHPYLLARKFTTLDHLTDGRIGWNIVTSQLDSAARNLGLERQIPH
DERYERAEEFLTVAYKLWEGSWDEGAVLRDRGTGVYADPSRVHAIGHHGRYFSVPGAALSEPSRQRTPVLYQAGTS
PRGSLFAARHAEIVFVAGHEPDVLRRNIDRIRVLAREQGREPDDIKFVASALVITDETDAGAEAKLRRYQDAYSIEGALT
HFSAITGIDWSEYDIDAPLSYIETDSNRSILASLTTDAPPGSVWTLRRLLAPARGVSYADAVVGSGTTVADRLEKLADET
GVDGFNLSAAVAHESYRDIADHVIPVLRDRGRIRRPASATASLREKLFETTEPHIGSRHPASRYRNAFTGLPSAAPRPV
AAS

>ISGA1666
MTVARPRMIFNAFNMFTVSHHDQGMWAWPGSRQREYNSVDYWVDVARLLERGHFDTLFFADVLAPYDTFGDSSER
AISSGMQFPVNDPGTLIPALAHATDDLGFVLTQNILQEPPYAFARKMSSLDHLTRGRIAWNIVTTFLPGAGRNLGFAGL
PDHAERYARADDFVDVVYKLWEASWEDDAVIADAATGRYNDPAKIHRIDHTGPYYDVVGPHLCEPSPQRTPFLVQAG
VSARGRDFAGRNAEALFINALSPQEAAPVVADVRAAAARHGRDPASVVLFGILGFVVGSTEAEAKRLQEEITDFQSID
AHLAKQSVFLGYDFGQLDPNEPIGEIAKRPEGKEGVVGQLIAMSPNDRFTIGELVRWYGNLRVVGTPEQIADHIEAWQ
DAGVGGMNVQYVVSPGTFEDFVDHVAPELERRGIMQDRYRPGTLREKIFPGNGPYLPEAHPARGHRRAAFGVTV

>ISGA1835
MSRPPIYNGFLHLTPNHHSHGFWRTPEGAVQYGYSKLDPYVDVVQTLERGLFDTLFIADVVGVYDLDFGDGTTTIRA
GSQFPEPDPVTIVSALGHATEHLGIAVTSNIIQSHPFTFARQLSSLDHFTDGRVAWNIVTSYLSNGFRNYGYDSIVGHD
ERYAWAQEYADVTYKLWEHSWQDGAVIHDPATNRFFDPDKIRTIDHVGPRYQVQGPHIVEPSPQRTPVLFQAGNSS
AGREFAVNNAEVTFLPSQTPATAREDIAVLDALAREKGRNPASLKKIVTLSTVIGSTEEEAKRKQQYFRDNIDFEALQA
FWSGGSGVDLTSVDPETPLAELAQRAQLGDHVRSIFRAAAQSQDEPESVSWRDYLLAQGLLPGRFAGTPEQIADHV

AEWVESGVDGFNVVPITTLGWWDEWVDHVVPVLQDRGLAQREYHQGTLRNKLFRSGDALDPTHRGRQIRLADVIGA
KG

>ISGA08960
MNSPTTRRADTDGPAHFHWFLPTSGDGREVIGGLQSAGVLGTASTIRPPDLDYLALVAKTAERLGFESVLTPTGTWC
HDAWLTTAALIRETSRLTFLVAFRPGLITPTLAAQQAATFAEFSGGRLALNIVCGGDAEEQRRFGDRLTKEQRYARAG
EFLTIVRQAWTGTPFDFTGEYYDVSGAVVAHPPVPAPPVFFGGASEPAREVAASSVDTYLTWTEPPGKVAALIADVR
ARAARHGRTLSFGIRAHVISRDTSEEAWAEARRLVDRMDPALIALARERLLQSESEGQRRQLDLNADLDRLEVHPGL
WAGYGLVRPGAGTAFVGSHAEVAALIAEYRAIGVDHFILSGQPHIEEAFWFAEGVVPLVRAAERAAAGPAAVVSGRE
R

>SsuD_*Escherichia_coli*
MSLNMFWFLPTHGDGHYLGTEEGSRPVDHGYLQQIAQAADRLGYTGVLIPTGRSCEDAWLVAASMIPVTQRLKFLVA
LRPSVTSPTVAARQAATLDRLSNGRALFNLVTGSDPQELAGDGVFLDHSERYEASAEFTQVWRRLLQRETVDFNGK
HIHVRGAKLLFPAIQQPYPPLYFGGSSDVAQELAAEQVDLYLTWGEPPELVKEKIEQVRAKAAAHGRKIRFGIRLHVIV
RETNDEAWQAAERLISHLDDETIAKAQAAFARTDSVGQQRMAALHNGKRDNLEISPNLWAGVGLVRGGAGTALVGD
GPTVAARINEYAALGIDSFVLSGYPHLEEAYRVGELLFPLLDVAIPEIPQPQPLNPQGEAVANDFIPRKVAQS

>SsuD_*Bacillus_subtilis*
MEILWFIPTHGDARYLGSESDGRTADHLYFKQVAQAADRLGYTGVLLPTGRSCEDPWLTASALAGETKDLKFLVAVR
PGLMQPSLAARMTSTLDRISDGRLLINVVAGGDPYELAGDGLFISHDERYEATDEFLTVWRRLLQGETVSYEGKHIKV
ENSNLLFPPQQEPHPPIYFGGSSQAGIEAAAKHTDVYLTWGEPPEQVKEKIERVKKQAAKEGRSVRFGIRLHVIARET
EQEAWEAAERLISHLDDDTIAKAQAALSRYDSSGQQRMAVLHQGDRTKLEISPNLWAGIGLVRGGAGTALVGDPQTI
ADRIAEYQALGIESFIFSGYPHLEEAYYFAELVFPLLPFENDRTRKLQNKRGEAVGNTYFVKEKNA

>SsuD_*Pseudomonas_putida*
MSLNIFWFLPTHGDGKYLGTSEGARAVDHGYLQQIAQAADRLGFGGVLIPTGRSCEDSWLVAASLIPVTQRLKFLVAL
RPGIISPTVAARQAATLDRLSNGRALFNLVTGGDPDELAGDGLHLNHQERYEASVEFTRIWRKVLEGEVVDYDGKHIQ
VKGAKLLYPPIQQPRPPLYFGGSSEAAQDLAAEQVELYLTWGEPPSAVAEKIAQVREKAAAQGREVRFGIRLHVIVRE
TNEEAWAAADKLISHLDDDTIARAQASLARFDSVGQQRMAALHNGNRDKLEVSPNLWAGVGLVRGGAGTALVGDGP
TVAARVKEYAELGIDTFIFSGYPHLEESYRVAELLFPHLDVQRPEQAKTSGYVSPFGEMVANDILPKSVAQS

>DszA_*Rhodococcus*_sp._IGTS8
MTQQRQMHLAGFFSAGNVTHAHGAWRHTDASNDFLSGKYYQHIARTLERGKFDLLFLPDGLAVEDSYGDNLDTGVG
LGGQGAVALEPASVVATMAAVTEHLGLGATISATYYPPYHVARVFATLDQLSGGRVSWNVVTSLNDAEARNFGINQH
LEHDARYDRADEFLEAVKKLWNSWDEDALVLDKAAGVFADPAKVHYVDHHGEWLNVRGPLQVPRSPQGEPVILQA
GLSPRGRRFAGKWAEAVFSLAPNLEVMQATYQGIKAEVDAAGRDPDQTKIFTAVMPVLGESQAVAQERLEYLNSLVH
PEVGLSTLSSHTGINLAAYPLDTPIKDILRDLQDRNVPTQLHMFAAATHSEELTLAEMGRRYGTNVGFVPQWAGTGEQ
IADELIRHFEGGAADGFIISPAFLPGSYDEFVDQVVPVLQDRGYFRTEYQGNTLRDHLGLRVPQLQGQPS

>DszB_*Rhodococcus*_sp._IGTS8
MTSRVDPANPGSELDSAIRDTLTYSNCPVPNALLTASESGFLDAAGIELDVLSGQQGTVHFTYDQPAYTRFGGEIPPL
LSEGLRAPGRTRLLGITPLLGRQGFFVRDDSPITAAADLAGRRIGVSASAIRILRGQLGDYLELDPWRQTLVALGSWEA
RALLHTLEHGELGVDDVELVPISSPGVDVPAEQLEESATVKGADLFPDVARGQAAVLASGDVDALYSWLPWAGELQA
TGARPVVDLGLDERNAYASVWTVSSGLVRQRPGLVQRLVDAAVDAGLWARDHSDAVTSLHAANLGVSTGAVGQGF
GADFQQRLVPRLDHDALALLERTQQFLLTNNLLQEPVALDQWAAPEFLNNSLNRHR

>NtaA_*Aminobacter_aminovorans*
MGANKQMNLGFLFQISGVHYGGWRYPSAQPHRATDIQYYAEIVRTAERGKLDFCFLADSIAAYEGSADQQDRSKDAL
MAAEPKRLLEPFTLLAALAMVTEHIGLVTTATTTYNEPYTMARLFASLDHITNGRAGWNVVTSANLAEAHNFGRDGHV
EHGDRYARAEEFINVVFKLWDSIEDGAYLRDKLAGRYGLSEKIHFINHIGEHFKVRGPLNVPRPPQGHPVIVQAGSSH
PGKELAARTAEVVFTAQQTLADGKAFYSDVKGRMAKYGRSSENLKVLPGVVVYVAETESEAKAKYETVSNLVPPDFG
LFMLSDLLGEIDLKQFDIDGPLPEDLPEAKGSQSRREVIINLARRENLTIRQLYQRVSGASGHRSIWGTPKQIADQFEQ
WVYEEAADGFNILPPYLPESMNDFVNFVVPELQRRGIFRTEYEGSTLRDHLGLARPKNSVAKPS

>LuxA_*Vibrio_harveyi*
MKFGNFLLTYQPPELSQTEVMKRLVNLGKASEGCGFDTVWLLEHHFTEFGLLGNPYVAAAHLLGATETLNVGTAAIVL
PTAHPVRQAEDVNLLDQMSKGRFRFGICRGLYDKDFRVFGTDMDNSRALMDCWYDLMKEGFNEGYIAADNEHIKFP
KIQLNPSAYTQGGAPVYVVAESASTTEWAAERGLPMILSWIINTHEKKAQLDLYNEVATEHGYDVTKIDHCLSYITSVD
HDSNRAKDICRNFLGHWYDSYVNATKIFDDSDQTKGYDFNKGQWRDFVLKGHKDTNRRIDYSYEINPVGTPEECIAII
QQDIDATGIDNICCGFEANGSEEEIIASMKLFQSDVMPYLKEKQ

>LadA_*Geobacillus_thermodenitrificans*_NG80-2
MTKKIHINAFEMNCVGHIAHGLWRHPENQRHRYTDLNYWTELAQLLEKGKFDALFLADVVGIYDVYRQSRDTAVREA
VQIPVNDPLMLISAMAYVTKHLAFAVTFSTTYEHPYGHARRMSTLDHLTKGRIAWNVVTSHLPSADKNFGIKKILEHDE
RYDLADEYLEVCYKLWEGSWEDNAVIRDIENNIYTDPSKVHEINHSGKYFEVPGPHLCEPSPQRTPVIYQAGMSERG
REFAAKHAECVFLGGKDVETLKFFVDDIRKRAKKYGRNPDHIKMFAGICVIVGKTHDEAMEKLNSFQKYWSLEGHLAH
YGGGTGYDLSKYSSNDYIGSISVGEIINNMSKLDGKWFKLSVGTPKKVADEMQYLVEEAGIDGFNLVQYVSPGTFVDF
IELVVPELQKRGLYRVDYEEGTYREKLFGKGNYRLPDDHIAARYRNISSNV

>SnaA_*Streptomyces_pristinaespiralis*
MTAPRRRITLAGIIDGPGGHVAAWRHPATKADAQLDFEFHRDNARTLERGLFDAVFIADIVAVWGTRLDSLCRTSRTE
HFEPLTLLAAYAAVTEHIGLCATATTTYNEPAHIAARFASLDHLSGGRAGWNVVTSAAPWESANFGFPEHLEHGKRYE
RAEEFIDVVKKLWDSDGRPVDHRGTHFEAPGPLGIARPPQGRPVIIQAGSSPVGREFAARHAEVIFTRHNRLSDAQDF
YGDLKARVARHGRDPEKVLVWPTLAPIVAATDTEAKQRLQELQDLTHDHVALRTLQDHLGDVDLSAYPIDGPVPDIPY
TNQSQSTTERLIGLARRENLSIRELALRLMGDIVVGTPEQLADHMESWFTGRGADGFNIDFPYLPGSADDFVDHVVPE
LQRRGLYRSGYEGTTLRANLGIDAPRKAGAAA

>DmoA_*Hypomicrobium_sulfonivorans*
MKKRIVLNAFDMTCVSHQSAGTWRHPSSQAARYNDLEYWTNMAMELERGCFDCLFIADVVGVYDVYRGSAEMALR
DADQVPVNDPFGAISAMAAVTEHVGFGVTAAITFEQPYLLARRLSTLDHLTKGRVAWNVVSSYLNSAALNIGMDQQLA
HDERYEMADEYMEVMYKLWEGSWEDDAVKRDKKSGVFTDGSKVHPINHQGKYYKVPGFHICEPSPQRTPVIFQAG
ASGRGSKFAASNAEGMFILTTSVEQARQITTDIRNQAEAAGRSRDSIKIFMLLTVITGDSDEAAEAKYQEYLSYANPEG
MLALYGGWTGIDFAKLDPDEPLQAMENDSLRTTLESLTHGENAKKWTVRDVIRERCIGGLGPVLVGGPQKVADELER
WVDEGGVDGFNLAYAVTPGSVTDFIDYIVPELRKRGRAQDSYKPGSLRRKLIGTNDGRVESTHPAAQYRDAYVGKES
VADRTQPSPFANAKAPVAE

>CamE36_*Pseudomonas_putida*
MAMETGLIFHPYMRPGRSARQTFDWGIKSAVQADSVGIDSMMISEHASQIWENIPNPELLIAAAALQTKNIKFAPMAHL
LPHQHPAKLATMIGWLSQILEGRYFLGIGAGAYPQASYMHGIRNAGQSNTATGGEETKNLNDMVRESLFIMEKIWKRE
PFFHEGKYWDAGYPEELEGEEGDEQHKLADFSPWGGKAPEIAVTGFSYNSPSMRLAGERNFKPVSIFSGLDALKRH
WEVYSEAAIEAGHTPDRSRHAVSHTVFCADTDKEAKRLVMEGPIGYCFERYLIPIWRRFGMMDGYAKDAGIDPVDAD
LEFLVDNVFLVGSPDTVTEKINALFEATGGWGTLQVEAHDYYDDPAPWFQSLELISKEVAPKILLPKR

>CamP_*Pesudomonas_putida*
MKCGFFHTPYNLPTRTARQMFDWSLKLAQVCDEAGFADFMIGEHSTLAWENIPCPEIIIGAAAPLTKNIRFAPMAHLLP
YHNPATLAIQIGWLSQILEGRYFLGVAPGGHHTDAILHGFEGIGPLQEQMFESLELMEKIWAREPFMEKGKFFQAGFP
GPDTMPEYDVEIADNSPWGGRESMEVAVTGLTKNSSSLKWAGERNYSPISFFGGHEVMRSHYDTWAAAMQSKGFT
PERSRFRVTRDIFIADTDAEAKKRAKASGLGKSWEHYLFPIYKKFNLFPGIIADAGLDIDPSQVDMDFLAEHVWLCGSP
ETVKGKIERMMERSGGCGQIVVCSHDNIDNPEPYFESLQRLASEVLPKVRMG

>EmoA_*Chelativorans* sp. BNC1
MRKRRMYLVSWLNSSGVLPNSWNEGRGNRARIFDLENYIRSAEIARRGRIDAFFLADQPQLTPNPKVRPEYPFDPIVL
AAAITGRVPDIGGIVTASTSFSLPYTLARQIASVNLLSGGRIGWNAVTTANPAVAANYGAAIATHDNRYERAEEFLEVVH
GLWNSWKFPWDEAIGPNPNPFGEVMPINHEGKYFKVAGPLNVPLPPYGPPVVVQAGGSDQGKRLASRFGEIIYAFLG
SKPAGRRFVAEARAAARAQGRPEGSTLVLPSFVPLIGSTEAEVKRLVAEYEAGLDPAEQRIEALSKQLGIDLERINVDQ
VLQEKDFNLPKESATPIGILKSMVDVALDEKLSLRQLALRMRLIAGTPDQVADRLIDWWQDEAADGFVINAPLLPDALEI
FVDQVVPILQSRGVFPRSYTESTLRERLGLPRNPLG

>RutA_*Escherichia_coli*
MQDAAPRLTFTLRDEERLMMKIGVFVPIGNNGWLISTHAPQYMPTFELNKAIVQKAEHYHFDFALSMIKLRGFGGKTE
FWDHNLESFTLMAGLAAVTSRIQIYATAATLTLPPAIVARMAATIDSISGGRFGVNLVTGWQKPEYEQMGIWPGDDYF
SRRYDYLTEYVQVLRDLWGTGKSDFKGDFFTMNDCRVSPQPSVPMKVICAGQSDAGMAFSARYADFNFCFGKGVN
TPTAFAPTAARMKQAAEQTGRDVGSYVLFMVIADETDDAARAKWEHYKAGADEEALSWLTEQSQKDTRSGTDTNVR
QMADPTSAVNINMGTLVGSYASVARMLDEVASVPGAEGVLLTFDDFLSGIETFGERIQPLMQCRAHLPALTQEVA

## 7.10 Protein sequencing results

In order to confirm the production of MBP1218 and MBP1222, the peptide profile of samples tentatively containing MBP1218 and MBP1222 were searched against proteins in the *Gordonia* NB4-1Y and *E. coli* genomes by the University of Guelph Mass Spectrometry Facility. The peptide profile of MBP1218 against the *E. coli* genome revealed 219 protein matches, 2 of which are maltose-binding periplasmic protein (P0AEX9 and P0AEY0) with post-translational modification (PTM), and against the *Gordonia* NB4-1Y genome, 2 matches to ISGA 1218 with PTM. The peptide profile of MBP1222 against the *E. coli* genome revealed 207 protein matches with 2 being to MBP (P0AEX9 and P0AEY0) with PTM and against the *Gordonia* NB4-1Y genome, 4 matches, 2 being ISGA 1222. Peptide coverage of MBP1218 and 1222 is 44 and 60%, respectively, and distributed across the entire protein.

Table S13. Proteins identified of LC-MS of digested semi-purified MBP1218

| Protein Group | Protein ID | Accession | -10lgP | Coverage (%) | Coverage (%) Sample 1 | Intensity Sample 1 | #Peptides | #Unique | #Spec Sample 1 | PTM | Avg. Mass |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 11030 | WP_053777105.1 | 265.64 | 44 | 44 | 2.05E7 | 88 | 86 | 172 | Y | 49949 |
| **1** | 11031 | EMP10004.2 | 265.64 | 44 | 44 | 2.05E7 | 88 | 86 | 172 | Y | 49949 |

Table S14. Proteins identified by LC-MS of digested semi-purified MBP1222

| Protein Group | Protein ID | Accession | -10lgP | Coverage (%) | Coverage (%) Sample 2 | Intensity Sample 2 | #Peptides | #Unique | #Spec Sample 2 | PTM | Avg. Mass |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 15830 | EMP10005.1 | 276.58 | 60 | 60 | 3.63E7 | 113 | 112 | 221 | Y | 48928 |
| **1** | 15831 | WP_020794796.1 | 276.58 | 60 | 60 | 3.63E7 | 113 | 112 | 221 | Y | 48928 |
| **2** | 11036 | WP_020793331.1 | 69.15 | 3 | 3 | 9.74E4 | 3 | 2 | 5 | N | 52095 |
| **2** | 11037 | EMP11479.1 | 69.15 | 3 | 3 | 9.74E4 | 3 | 2 | 5 | N | 52095 |


Figure S23. LC-MS identified peptide coverage of ISGA 1218. Blue bars represent peptides identified by LC-MS and are aligned with the amino acid sequence of ISGA 1218.

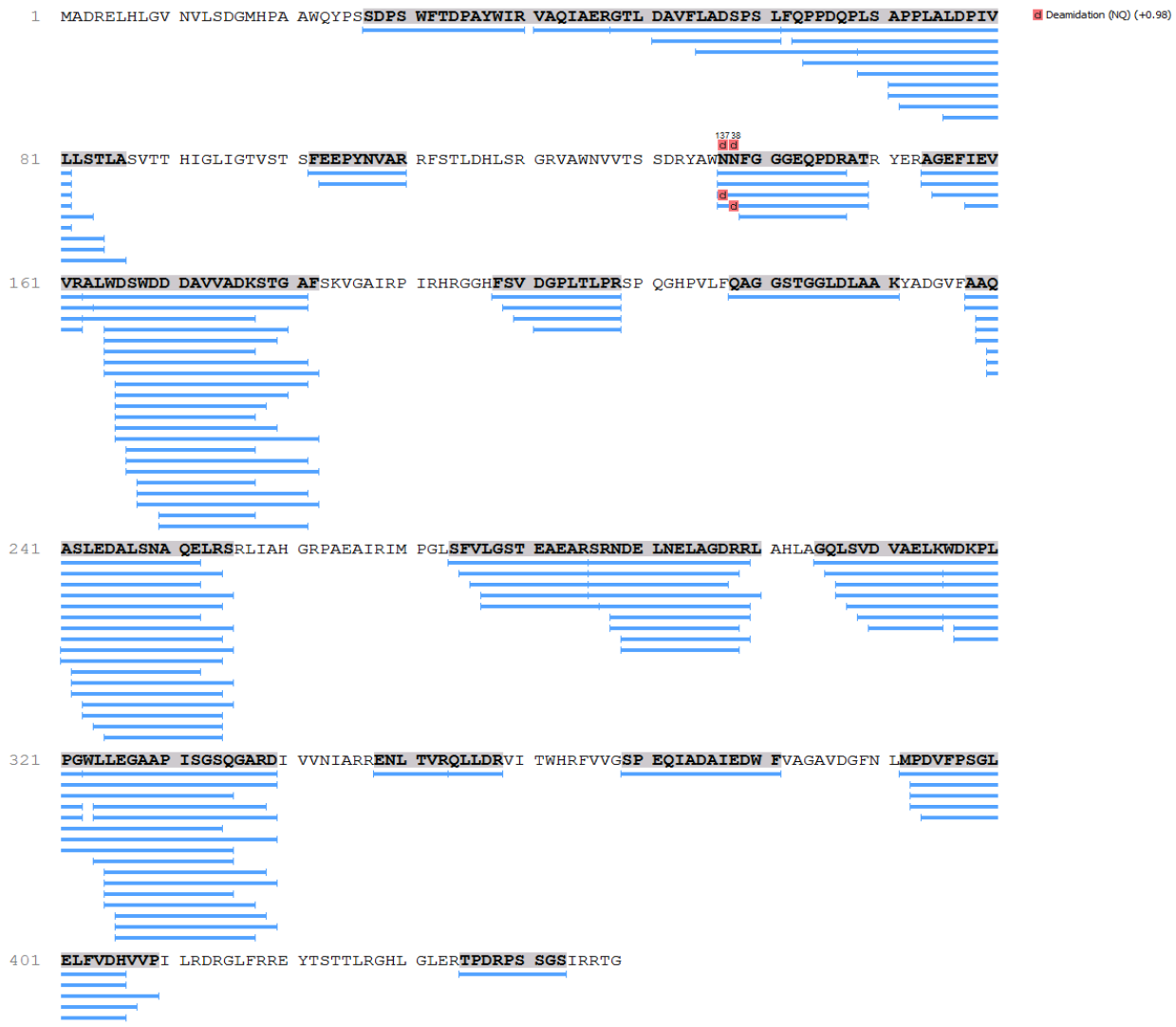


Figure S24. LC-MS identified peptide coverage of ISGA 1222. Blue bars represent peptides identified by LC-MS and are aligned with the amino acid sequence of ISGA 1222.

## 7.11 Photoreduction of flavin

In order to produce $FMNH_2$ *in vitro* without Fre for the monooxygenases in this study, FMN was photoreduced anoxically in the presence of 40-fold EDTA. In order to produce anoxic conditions, nitrogen was bubbled by needle in an airtight flask sealed with a butyl-rubber septa and aluminum crimp seal. A second needle was added to maintain equal pressure within and without the flask. While bubbling, the solution was exposed to a 15-watt, 120-volt light source. Once the solution changed from a bright yellow to a pale yellow (around 5 minutes), the solution was considered photoreduced. The protein and substrate were added by piercing the septa with a clean needle and adding the enzyme substrate mix directly to the solution while maintaining light. The $FMNH_2$/substrate/enzyme mix was incubated for 30 minutes before introducing oxygen by either vacuum or removing the butyl-rubber septa. Gaseous compounds were trapped on a 600 mg C18 cartridge (Grace Davison Discovery Science, Il).

## 7.12 Counter selection of pK18mobsacB in *Gordonia* NB4-1Y

If a single recombinant *Gordonia* NB4-1Y colony was isolated, the following would have been done to remove the genome integrate plasmid.

Isogenic colonies of transconjugant *Gordonia* would be inoculated in LB or NB without antibiotics and grown for 4 days at 30°C with shaking. Cultures would be diluted in sterile water at 1:10, 1:100 and 1:1000, plated on LB or NB agar with 10% wt/vol sucrose and grown for 1-4 days at 25°C. Isolated colonies would be picked with a sterile toothpick and struck, sequentially, on two plates: LB or NB agar with 50 µg/mL of kanamycin followed by LB or NB agar without antibiotics. Colonies which could not grow on kanamycin would be picked from the LB or NB agar without selection and further confirmed with 18-250/750 or 22-250/750 primers. Colonies which produced a 1000 bp product would be considered double recombinants and had ISGA 1218 or 1222 deleted.

## 7.13 Sulfur limiting growth assay: *Gordonia* NB4-1Y

If *Gordonia* NB4-1Y mutants were produced, the following would have been followed to determine if the mutants can grow on select sulfur sources.

Growth medium to be tested would be M9 minimal medium supplemented with 200-400 µM $MgSO_4$, 6:2 FTSA or no added sulfur. Isogenic colonies of wild-type and mutant *Gordonia* NB4-1Y would be inoculated in M9 minimal medium supplemented with 400 µM $MgSO_4$ and grown to saturation at 30°C. The following day, cells would be harvested and washed twice with water and resuspended in 1-2 mL of water. In four separate clean sterile 125-mL flasks, *Gordonia* NB4-1Y strains would be added to a final $OD_{660}$ of 0.05 to 100 mL of M9 minimal medium supplemented with 400 µM of $MgSO_4$, octane sulfonate, 6:2 FTSA or no added sulfur. The flasks would be mixed, and 5 mL aliquots would be dispensed to 50-mL Kimax culture tubes (Kimble-Chase, TN). Culture tubes would be incubated at 30°C with shaking and sacrificed at 0, 24, 48, 72 and 96 hours by taking 2.1 mL aliquots. Aliquots would be divided to make three growth measurements, 1 mL for $OD_{660}$, 1 mL for protein content quantification by Qubit Protein Assay Kit following sonication and 0.1 mL for serial dilution and plate count on LB or NB agar.