The Properties of the Fatty Aldehyde Decarbonylase from *Synechocystis* PCC6803

Submitted by

Robert James Kalibala

to the University of Exeter as a thesis for the degree of Masters by Research in Biological Science, (March 2012)

I certify that all materials in this thesis which is not my work has been identified and that no materials has been submitted and approved for the award of a degree by this or any other University.

Signature……

Robert James Kalibala

ABSTRACT

Alkanes dominate the constituents of gasoline, diesel, and jet fuel and are naturally produced by diverse species; saturated and unsaturated fatty acids are converted to alkanes and alkenes respectively by the enzyme aldehyde decarbonylase (AD). Here we describe the over-expression, purification, data collected and X-ray crystal structure solved for the AD protein from *Synechocystis* PCC6803.

This report describes the optimisation of over-expression, protein purification and characterization and crystallisation of the *Synechocystis* cyanobacterial AD enzyme (SynADC) has been carried out. The optimisation of protein expression has been carried out using the pET160, pET22b and pCold$^{TM\ II}$. Expression of soluble protein was obtained with all vectors. The initial Lumio^{TM} tag on pET160 prevented the protein from crystallising; the pCold$^{TM\ II}$ vector with a small His-tag was used for high soluble protein over-expression. The purification of the SynADC was optimized and the enzyme was characterised biochemically, SynADC was found to be a dimer of 29 kDa molecular weight. Metal contents were investigated using ICP-MS, SynADC protein was found to contain; Zn, Fe, Ni and Mn metals in a ratio (2.37, 1.16, 0.137, and 0.032) mg/l respectively.

The enzyme has been assayed using a series of ferredoxin assays of (C_8, C_{10}, C_{12}, C_{13}, C_{16} and C_{18}) and activity has been determined using C_{13} aldehyde and C_{18} aldehyde.

The enzyme has been successfully crystallised with four different ligands (valeric acid, Hexanoic acid, C_4 and C_8) using the microbatch method and metal soaking, this has allowed the X-ray structure to be determined. Based on this structure predication of electron transfer mechanism, a mutagenesis experiment has been carried out with the change of Asp143 to Asn, Leu and Ala. The enzyme has been assayed using PMS. Experiments to determine potential proteins, which could interact with SynADC, have been carried out. Positive results have been obtained using SDS-PAGE however, more protein is required for mass spectrometric determination.

This project was part of a larger study to clone and solve the structure of the *Synechocystis* Cyanobacterial AD in order to understand its substrate specificity and mechanism. Work carried out in collaboration with others is clearly mentioned in this thesis.

Acknowledgements

Firstly, I would like to thank Prof. Nicholas Smirnoff and Prof. Jenny Littlechild for giving me the opportunity to undertake this project and for the continuous help and guidance throughout. Thank you for giving me the freedom to develop the project in directions that interested challenged me to the limit.

Secondly, I would like to thank Dr. Misha Isupov for carrying out the data processing and structure refinement, Dr Christoph Edner for his supervision in the lab and for getting me back on track with my practical work when I was at my wits end and for reminding me what it felt like for practical work to run as it should, Mr. Kevin for help me with ICP – MS experiments and Dr Hannah Florence for help me with Mass spectrometry experiments.

Thirdly, one of the biggest thank you has to go to the “Mezz” and my beloved colleagues of “Biocats”. You know who you are and your help, patience, friendship and morale support. I only hope and wish I could manage to do for you all what you did for me.

And finally to my friends and family: George for support and distracting me when my brain was ready to implode, as Dr. Tomas, Dr. Paul and Prof. Jenny Littlechild for believing in me always, for your unconditional support and encouragement you’re truly made a huge difference in my life thank you ever so much.

ABBREVIATIONS

$^{\circ}C$	Degree centigrade
A	Amps
Å	Angstrom (10^{-10}m)
A_{280}	Absorbance at 280nm
A_{600}	Absorbance at 600 nm
APS	Ammonium persulphate
BLAST	Basic local alignment search tool
DMSO	dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic Acid
g	Acceleration due to gravity
g	grams
GF	Gel filtration
hr	Hour
IPTG	Isopropyl ß-D-galactopyranoside
K	Kelvin
kDa	Kilo Dalton
mg	milligrams
min	Minute
ml	milliliter
MW	Molecular weight
nm	nanometer
NMR	Nuclear Magnetic resonance
No.	Number
OD	Optical density
PAGE	Poly acrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein date bank
PMSF	phenylmethylsulfonyl fluoride
PI	Isoelectric point
ppt	Precipitate
rtm	Room Temperature
s	Second

SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
TEMED	N, N, N, N- tetramethylethylene diamide
Tris	Tris [hydroxymethyl] aminomethane
UV	Ultra violet
V	Volts
v/v	volume per volume
V_0	Initial velocity
V_{max}	Maximum velocity
Vol.	Volume
w/v	Weight to volume

Organism abbreviations

E. coli	*Escherichia coli*

CONTENTS

CHAPTER 3: PROTEIN EXPRESSION AND PURIFICATION

CHAPTER 6: PROTEIN CRYSTALLISATION

CHAPTER 8: CONCLUDING COMMENTS AND FUTURE WORK

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1:

INTRODUCTION

1.1 Hydrocarbons

Hydrocarbons are extremely important energy resources; they can be produced chemically although the processes involved are expensive and not environmental friendly either. Many plants, insects and microbes naturally produce small quantities of alkanes and alkenes, the long-chain carbon and hydrocarbon molecules that are major components of diesel, petrol (gasoline) and jet fuel. Efforts to make the transition from fossil fuels to renewable alternatives have focused on the conversion of renewable biomass to “drop in” compatible fuels and chemicals. Biotechnology companies (such as LS9, based in South San Francisco, California) have emerged using microbes to produce renewable biofuels. Pinpointing the biochemical pathways that organisms (plants, insects and microbes) use to do this is critical. A paper published in Science (Schirmer *et al*., 2010), reported that they have identified the genes responsible and expressed these genes in the bacterium *Escherichia coli*, fed it glucose, and showed that it directly secreted diesel–like fuel – a mixture of alkanes and alkenes with chain lengths from 13 to 17 carbon atoms. This is in agreement with what is required for diesel, although further tweaking of the product is still required to generate the hydrocarbons used in petrol and jet fuel, which is at high demand from companies.

1.1.2 Historical background of microbial hydrocarbons

Hydrocarbons are the most stable group of naturally occurring compounds (Ladygina *et al*., 2006). They are thought to retain much of their original architecture over a very long period of time. Hydrocarbon biomarkers are used to determine the age of ancient bacteria, archaea, and eukaryotes (Summons *et al*., 1999, Brocks *et al*., 2003). The oldest fossils known are of bacterial and algal origin, some more than 3 billion years old. The first studies carried out on the isolation of the hydrocarbons-like

substances from the cells of marine bacteria and algae were conducted to elucidate the role of microorganisms in the genesis of petroleum and formation of organic substances in marine sediments.

Extensive investigation of microbial hydrocarbons has been greatly enhanced by the development of new analytical techniques. After 60 years of research, a lot of information has appeared on the interacellular hydrocarbons of different systematic groups of microorganisms and mechanisms of the hydrocarbon biosynthesis. In contrast to higher organisms, microorganisms can be cultivated in reactors that allow the industrial production of hydrocarbons to be developed. In recent years, microbial synthesis of extracellular aliphatic and volatile non-methane hydrocarbons has attracted considerable interest in view of the development of effective and environmentally safe methods for biofuel production (Fukuda *et al*., 1987, Belyaeva *et al*., 1995 and Bagaeva, 1998).

1.1.3 The role of hydrocarbons in organisms

In higher plants, alkenes are mainly involved in the synthesis of the epicuticular wax layer (Ladygina *et al*., 2006), which primarily function to reduce water loss (evaporation) through the epidermis; further more, this outer layer has a major function in the plant interactions with herbivorous insects and the plant pathogenic fungi. The composition of pollen wax is cited to be an important factor for the proper pollen-pistil interactions (Aarts *et al*., 1995). In some species *Botryococus*, hydrocarbons serve as the storage intermediates for the synthesis of epoxides and other lipids.

Although the full understanding the role of hydrocarbons are yet to be established in microorganisms, they are found to arrange at the surface walls of fungal spores, which appears to fulfill the protective function. This is important in plant disease development because the spores become more resistant to desiccation. The functions of intracellular hydrocarbons of microorganisms remain unclear because of their low content, although they might be thought to play a role in the carbon and energy supply in microorganisms.

1.2 Intracellular hydrocarbons of microorganisms

When we look at the intracellular hydrocarbons of microorganisms, the first report on the production of C_{10} – C_{25} aliphatic hydrocarbons by sulfate-reducing bacteria grown in the seawater-containing media supplemented with organic or fatty acids was made in 1944. Stone and Zobell continued the research and in 1952, when they isolated hydrocarbon fractions from the marine bacteria *Serratia marinorubrum* and *Vibrio ponticus* grown on seawater peptone media. The development of gas-liquid chromatography initiated the extensive studies of microbial hydrocarbons and helped to overcome the experimental problem of the exclusion of adventitious hydrocarbon contaminations. This application demanded defined media and reagents of high purity. By using labeled growth substrates, the hydrocarbon biosynthesis can be studied in more detail for various groups of microorganisms including bacteria, yeast and fungi. However, the hydrocarbon content of different systematic groups of microorganisms varies over a wide range. Gram-positive bacteria of the genus *Clostridium*, whose growth is based on the fermentative process, produces interacellular hydrocarbons from C_{11} to C_{35} with the predominance of middle-chain *n*-alkanes (C_{18} – C_{27}) or long- chain alkanes (C_{25} – C_{35}) needed for growth. In contrast, Gram-negative anaerobic bacteria *Desulfovibrio desulfuricans* allows growth by using other metabolic mechanisms based on the anaerobic sulfate respiration (sulfate reduction). Nevertheless, they also produce C_{11} – C_{35} hydrocarbons with the predominance of C_{25} – C*35 n*-alkanes. The hydrocarbon production in *Micrococcus* and *Sercina* (aerobic Gram-positive bacteria) is the most thoroughly investigated. They mainly produce monounsaturated (C_{23} – C_{30}) hydrocarbons with a double bond located near the middle of the chain (Stone and Zobell, 1952).

Like any other microorganisms mentioned in a review on microbial synthesis of hydrocarbons (Ladygina *et al*., 2006), yeasts are able to synthesize a wide range of hydrocarbons from C_{10} to C_{34}, which include not only n-alkanes but also unsaturated and branched components. Ladygina *et al*., 2006 high lightened that hydrocarbon synthesis of yeasts depend considerably on the growth conditions, which provide a way for physiological regulation of the process. For example, results suggest that anaerobic conditions promote the increased production of intracellular hydrocarbon in *Sacharomyces oviformis.*

In fungi, the hydrocarbons tend to accumulate in mycelia and spores but differ significantly in amount and composition. A bulk of information concerning microbial hydrocarbons demonstrates that the ability to synthesize hydrocarbons is widespread among microorganisms. Therefore, certain systematic groups of microorganisms are characterized by specific composition of the hydrocarbon fractions, for example, cyanobacteria are found to be unique in their ability to produce 7- and 8-methylheptadecanes; photosynthetic bacteria are distinguished by the synthesis of cyclic hydrocarbon, whereas in fungi, long-chain hydrocarbons are predominant.

1.3 Extracellular hydrocarbons of microorganisms

1.3.1 Long chain hydrocarbons

Microbial production of extracellular long-chain hydrocarbons receives considerable interest in connection with their role in petroleum formation and the possibility of their industrial application as modern energy resources. Long chain hydrocarbons are present in many different types of organisms including bacteria, fungi, algae, higher plants and insects. They are thought to be the major components of insect cuticular lipids, where they serve a critical function of preventing desiccation, the cuticular hydrocarbons of the female housefly serve as components of the sex pheromone (Kolattukudy *et al*., 1976). There are various types of hydrocarbons reported to be found in insects (Jackson and Blomquist 1976), normal methyl-branched and unsaturated components; often range in chain length from C_{12} to C_{50}. Processes for the production of extracellular hydrocarbons that have been attempted and involve the cultivation of sulphated-reducing bacteria *Desulfovibrio desulfuricans* in medium containing mineral salts and calcium lactate under anaerobic conditions in an atmosphere of CO_2 + H_2 in (1:20) ratio (Ladygina *et al*., 2006). The earlier attempts by Jones, 1969, and Tornabene *et al*., 1970 for the extensive production of these extracellular hydrocarbons using bacterium *Vibrio furnissii* revealed that up to 50% of hydrocarbons of the dry biomass can be generated. The hydrocarbons produced were similar to that of kerosene and light oil therefore, *V. furnissii* is a promising producer of renewable fuel.

1.4 Hydrocarbon synthesis pathways in organisms

Biosynthesis of straight-chain hydrocarbons

There many ways of how straight-chain hydrocarbon can be produced in microorganisms but only two routes for the biosynthesis of straight-chain hydrocarbons which are widely understood: the “elongation-decarboxylation” and “head–to-head condensation” pathways. The best known of them, is the “elongation-decarboxylation” pathway. Long-chain fatty acids are formed through the continuous addition of a C_2 unit delivered from malonyl-CoA with the subsequent decarboxylation. Cyanobacterium *Nostoc muscorum* and *Escherichia coli* are suggested to have this mechanism (elongation-decarboxylation) for their hydrocarbon synthesis (Albro and Dittmer 1970; Naccarato *et al*., 1974).

The other widely considered pathway for the alkane biosynthesis by microorganisms is the involvement of head-to-head condensation of two fatty acids with the subsequent decarboxylation of one of them. The mechanism involves the reaction of decarboxylation; which involves a direct loss of the carboxyl carbon of a fatty acid; the lost carboxyl carbon group comes from the electron withdrawing adjacent to the α-carbon. The elimination of CO_2 from carboxylic acids requires high energy and, therefore, β-substituent helps to stabilize the negative charge generated by CO_2 released.

There are most likely other pathways, in other organisms yet to be found, of how direct straight-chain hydrocarbons are made. Insects for example, can use Cytochrome P450 enzyme to make alkanes, through no one has yet looked into the genes involved in this pathway. Although it will be hard to find pathways that improve on the two widely known pathways (the elongation-decarboxylation and head-to-head condensation), and then transform these pathways to an industrial organism like *E. coli* it is worth trying.

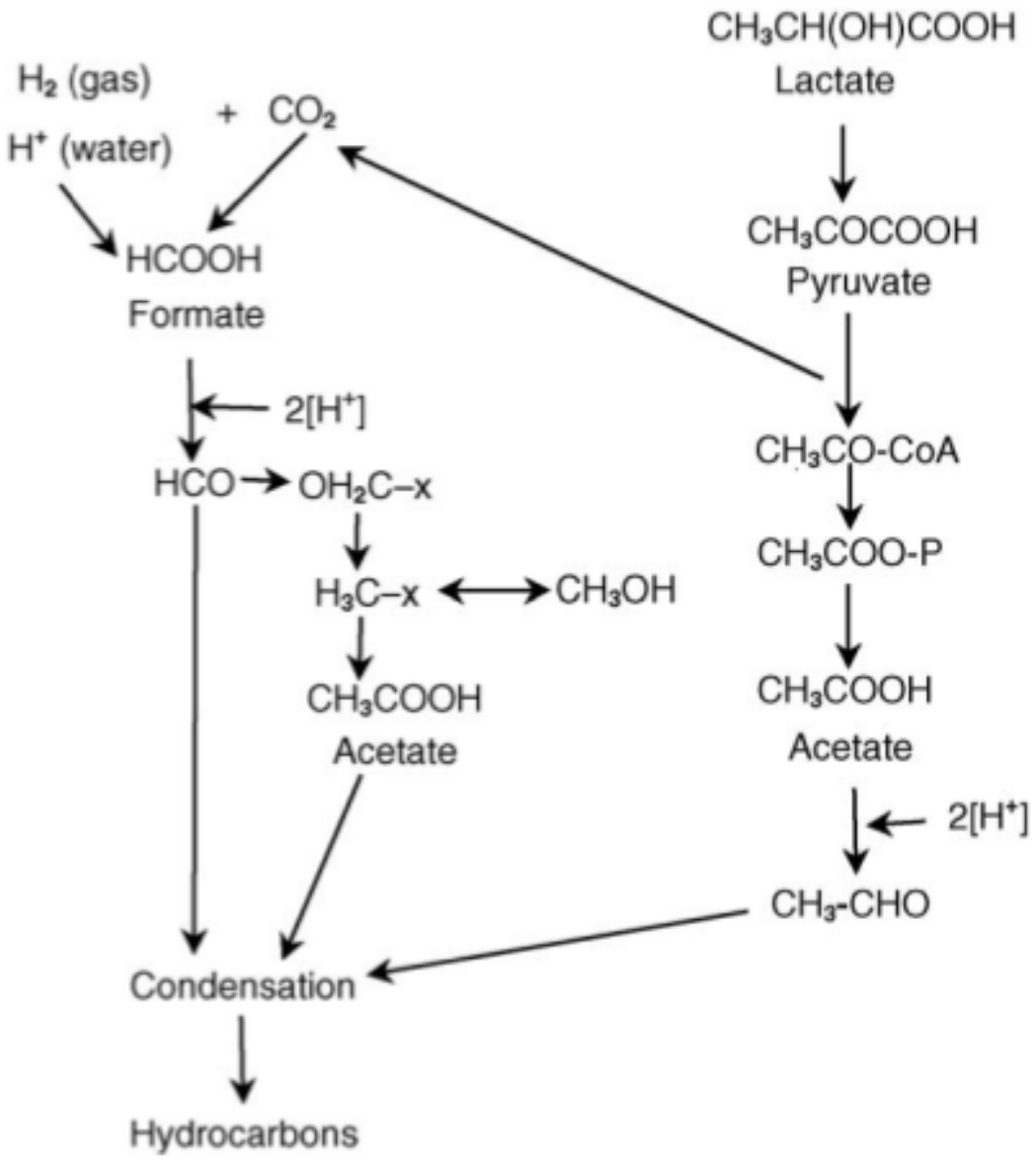


Figure 1.0 Pathway for the hydrocarbon biosynthesis by sulfate-reducing bacteria (Adapted from Bagaeva, 1998). Hydrocarbon synthesis by sulfate-reducing bacteria involves formation of acetate and formate from CO_2 with the subsequent reduction of these acids to aldehydes, which in turn, undergo aldol condensation with the chain elongation to produce hydrocarbon.

1.5 Cyanobacteria aldehyde decarbonylase

The search for new biofuels has generated increased interest in biochemical pathways that produce hydrocarbons. Hydrocarbons are simple molecules however

the biosynthesis of these molecules that lack any chemical functional groups is surprisingly challenging (Buist, 2007). Biochemical reactions that remove functionality, such as decarboxylation, dehydration and reduction of double bonds, invariably rely on the presence of adjacent functional groups to stabilize unfavorable transition states. Enzymes involved in hydrocarbon biosynthesis are therefore of interest both for applications in biofuels production and because of the unusual and chemically difficult reactions they catalyze. One enzyme that has attracted particular interest is aldehyde decarbonylase (AD), which catalyzes the decarbonylation of long-chain fatty aldehydes, to the corresponding alkanes (Cheesbrough and Kolattukudy, 1984; Dennis and Kolattukudy, 1992 and Schirmer *et al*., 2010). Evidence suggests that a two-step pathway for alkane biosynthesis, consisting of (1) reduction of an acyl-carrier –protein / coenzyme –A-linked fatty acyl thioester to the corresponding aldehyde and thiol and (2) C1–C2 bond of the fatty aldehyde is added to yield the alkane which is conserved across many species. However, the most consistent reports are from the cyanobacteria **(**Schirmer *et al*., 2010) and natural habitats dominated by cyanobacteria.

Alkane biosynthesis from the cyanobacteria pathway consists of an acyl-acyl carrier protein reductase and an aldehyde decarbonylase (Figure 1.3), which together are thought to convert intermediates of fatty acid metabolism to alkane and alkenes. Recently Douglas and his coworker (Douglas *et al*., 2011) have discovered a new pathway through which saturated fatty acids are being converted to alkanes (and unsaturated fatty acids to alkenes) in cyanobacteria. This entails scission of the C1-C2 bond of a fatty aldehyde intermediate by the enzyme (AD), a ferritin-like protein with a dinuclear metal cofactor of unknown composition.

Cyanobacteria are phylogenetically homogenous, with more than 50 sequenced genomes publically available. Very recently, Schirmer and co-workers identified pairs of genes encoding orthologous sets of fatty acyl-ACP reductases and aldehyde decarbonylase (AD) from ten species of cyanobacteria (Schirmer *et al*., 2010). The AD orthologue shows an incredible conversion of octadecanal, R-CHO (R=n-$C_{17}H_{35}$), to heptadecane (R-H) *in vitro.*

AD was originally identified in the studies on the biosynthesis of hydrocarbon waxes by higher plants and algae (Kunst and Samuels, 2003). They have integral membrane proteins, which converts long-chain aldehydes, derived from fatty acids, to alkanes and carbon monoxide. The enzyme requires divalent metal ions for activity, (Dennis and Kolattukudy, 1992; Schneider-Balhaddad and Kolattukudy, 2000) however; the identity of the metal remains unclear. It was speculated that the pea enzyme is active with Co, Cu, or Zn (Schneider-Balhaddad and Kolattukudy, 2000). The AD found in cyanobacteria (cAD) is a soluble version. A crystal structure for cAD from *Prochlorococcus marinus* MIT9313 had been solved as part of a structural proteomics project, although no function had been assigned (Joint Center of Structural Genomics).

> "*The structure revealed that cAD is a member of the non-hemedinuclear iron oxygenase family of enzymes exemplified by methane monoxygenase, type 1 ribonucleotide reductase, and ferritin*. *In these enzymes, the diiron center is contained within an antiparallel four-*α*-helix bundle in which two histidines and four carboxylates either from aspartate or glutamate supply the protein ligands to the metal ions"* (Das *et al*., 2011)

The dinuclear iron center of cAD was found to superimpose very closely on those of other enzymes; however in the structure the carboxylate of a bound long-chain fatty acid bridges the two iron atoms, displacing one of the glutamate ligands (Das *et al*., 2011) (Figure 1.1).

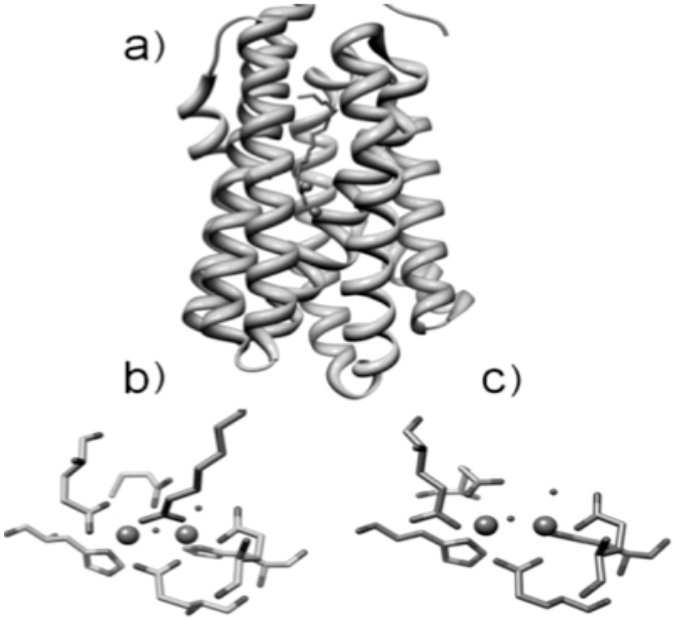
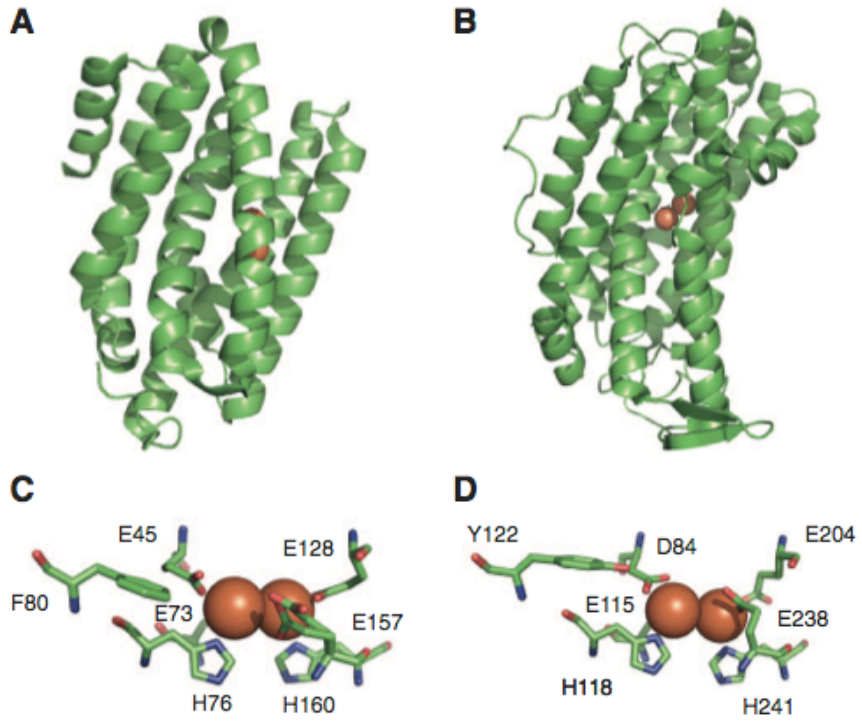


Figure 1.1 Structure of cAD from *P. marinus* (PDB 2OC5A). a) Ribbon diagram showing position of iron atoms and co-crystallized fatty acid. b) Structure of the diiron cluster showing protein ligands and co-crystallized fatty acid, which displaces one Glu ligand. c) Comparison with the diiron cluster of methane monooxygenase from *Methylococcus capsulatus* (PDB 1XVC). (Das *et al*., 2011)

Cyanobacterial ADs are identified to be members of the ferritin-like or ribonucleotide reductase-like family of nonheme diiron enzymes (Stubbe and Gelasco, 1998; Das *et al*., 2011).



10

Figure1.2 Comparison of the similar three-dimensional structure of the *Prochlorococcus* AD and ribonucleotide reductase R2 from *E. coli*. (**A**) AD from *P. marinus* MIT9313 from PDB file 2OC5. (**B**) Ribonucleotide reductase R2 from *E. coli* from PDB file1RIB. (**C**) Active site of the *P. marinus* MIT9313 AD. (**D**) Active site of the *E. coli* ribonucleotide reductase R2 Amino acid residues that coordinate the metal centers shown. The structure shows that two iron atoms are involved Schirmer *et al*., (2010).

Ferritin-like, diiron-carboxylate proteins participate in a range of functions including iron regulation, mono-oxygenation, and reactive radical production. These proteins are characterized by the fact that they catalyze dioxygen-dependent oxidation-hydroxylation reactions within di-iron centers; one exception is manganese catalase, which catalyses peroxide-dependent oxidation-reduction within a dimanganese center (Schirmer *et al*., 2010). Di-iron-carboxylate proteins are said to be further characterized by the presence of duplicate metal ligands, glutamates and histidines (ExxH) and two additional glutamates within a four-helix bundle (Schirmer *et al*., 2010) Outside of these conserved residues there is little obvious homology. Members include bacterioferritin, ferritin, rubrerythrin, aromatic and alkene monooxygenase hydroxylases (AAMH), ribonucleotide reductase R2 (RNRR2), acyl-ACP-desaturases

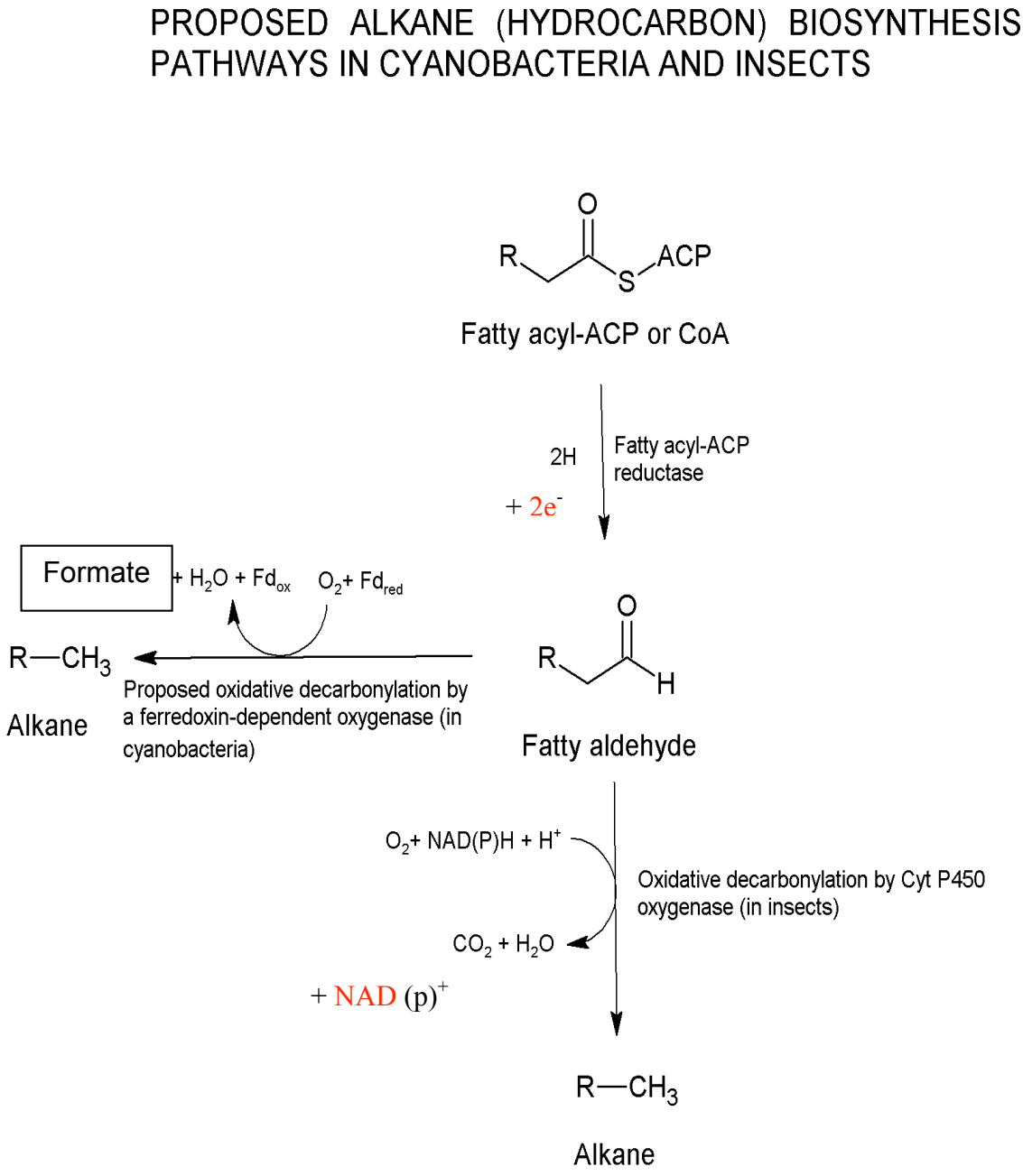
(Acyl_ACP_Desat), manganese (Mn) catalases, demethoxyubiquinone hydroxylases (DMQH), DNA protecting proteins (DPS), and ubiquinol oxidases (AOX), and the aerobic cyclase system, Fe-containing subunit (ACSF) (Schirmer, *et al*., 2010).

1.6 P450 monooxygenases

Cytochrome P450 monooxygenases (CYP) are heme protein-dependent mixed function oxidase systems that utilize NADPH and or NADH to reductively cleave atmospheric dioxygen to produce a functionalized organic substrate and a molecule of water. The substrates of CYP enzymes include; metabolic intermediates such as lipids and steroidal hormones, as well as other toxic chemicals such as drugs. These enzymes have been identified in all domains of life i.e., present in animals, plants, fungi, protists, bacteria, archaea, and even viruses (Guengerich 2008). In eukaryotic cells, P450s occur predominantly in microsomal membranes after synthesis by cytosolic ribosomes and co-translational transport into the endoplasmic reticular system while as in animals CYPs are found in numerous mammalian tissues and cell types (Schuler and Werck-Reichhart 2003).

$$RH + O_2 + NADPH + H^+ \rightarrow ROH + H_2O + NADP^+$$

(Common reaction catalyzed by cytochrome P450 monooxygenases)



ACP = acyl carrier protein. CoA = coenzyme A.
Fd_{red} = reduced ferredoxin. Fd_{ox} = oxidised ferredoxin

Figure 1.3 Proposed Microbial Biosynthesis of Alkanes (Schirmer *et al.*, 2010)

1.7 History of enzymology

The field of enzymology began in the 1800s. The first recorded phenomenon was by Joseph Gay-Lussac who observed that sugar decomposition by yeast produced ethanol and carbon dioxide during alcoholic fermentation. After attempts to mimic the effect chemically failed, causing Louis Pasteur to conclude that fermentation could only occur in living cells. It was assumed that cells were endowed with a "vital force" that enabled fermentation to proceed, but Justus Liebig argued that chemical substances or "ferments" were responsible for the biological processes. In 1878 Fredrich Wilheim Kuhne made-up the term "enzyme" in an attempt to emphasize that it was something in yeast causing fermentation, as opposed to yeast itself. Fifteen years later in 1893 Eduard and Hans Buchner made cell-free liquid extracts of microorganisms, which proved the theory of Kuhne and Liebig – they were using a yeast extract for pharmaceutical studies and added thick sugar syrup to stop any bacterial action. The sugar was to act as a preservative, but it had the opposite effect and a gas was produced. The sugar had fermented, producing carbon dioxide and alcohol, in the same way as if the whole yeast cell had been present. This finally proved that specific components of cells are responsible for catalysis, and the term enzyme has since been used for all protein capable of catalyzing a specific reaction.

> *"…the most striking characteristics of enzymes are their catalytic power and specificity by utilizing the full repertoire of intermolecular forces, enzymes bring substrates together in an optimal orientation, the prelude to making and breaking chemical bonds by selecting stabilising transition states an enzyme can determine which one several chemical reactions occurs ." STRYER, 1995*

1.7.1 Enzymes and biotechnology

Once the potential of enzymes was realised their properties were soon utilised for the benefit of mankind. Enzymes were already being used to produce beer and cider and soon greater possibilities were seen. Genentech the first biotechnology company was founded in 1976 and this was mother of the "biotechnology" era. This was followed by the greatest break-through – the development of polymerase chain reaction (PCR) by Kary Mullis in the 1980s (Mullis *et al*, 1986; Mullis and Faloona, 1987) this

transformed science, the field of biotechnology suddenly grew as cloning was a less time consuming procedure and became a more commonplace technique. Along with the theory of PCR, there was birth of *Taq* polymerase enzyme, which could withstand all of the temperatures used during a PCR cycle without being denatured, made cloning by PCR a cost effective method. Protein could now be produced in the yields required for industrial use and in easily maintained organisms such as yeast and *E. coli*.

1.7.2 Enzyme classification

The Enzyme Commission Classification System was brought in to classify the macromolecules according to the reactions that they are known to catalyse. Each group of enzymes therefore falls into an "EC" class that categorises it with other enzymes performing the same role, for example EC 1.1.1.1 symbolises an alcohol dehydrogenase and includes both prokaryotic and eukaryotic proteins, which metabolise alcohols, aldehydes and ketones. This allows us to be able to identify and compare structural motifs and catalytic amino acids of any enzyme classes. There are 6 main classes of enzymes, oxidoreductase (EC1), transferases (EC2) hydrolases (EC3), lyases (EC4), isomerases (EC5) and ligases (EC6).

1.8 Aim and objectives

The aim of this project was to carry out studies with another cAD from *Synechocystis* species. This enzyme has already been cloned by Dr. Christoph Edner in the laboratory of Prof. Nick Smirnoff at Exeter University.

Aims of the master's thesis were to: -

- Carry out bioinformatics studies with *Synechocystis* cAD to look at its homology with related proteins.
- Optimise the over-expression and protein purification of this enzyme.
- Biochemically characterise the enzyme and determination of metal content.

- Pull down assays to determine other proteins, which interact with *Synechocystis* cAD.
- Carry out crystallisation trials with the *Synechocystis* cAD protein.
- Co-crystallisation and soaking with metals and substrates.
- Determining the X-ray structure with Dr. Misha Isupov in the laboratory of Prof. J. Littlechild.
- Initial site-directed mutagenesis experiments based on the X-ray structure to establish mechanism of electron transfer.

Work carried out by others in Prof. Nick Smirnoff and Prof. J. Littlechild laboratory is indicated in the text as appropriate.

CHAPTER 2:

PROTEIN SEQUENCE ANALYSIS

2.1 Introduction

The amino acid sequence of the *Synechocystis* cAD enzyme protein sequence is 231 (sll0208) amino acids long (including the start methionine residue) and has a theoretical molecular weight of 26177.7 Da. Knowledge of this primary structure can be utilised to calculate parameters of the protein and search databases for homologous protein sequences. Every comparison results in a score and larger scores indicate a higher degree of similarity. The program BLAST (Basic Local Alignment Search Tool) (Altschul *et al*., 1990 and 1997) is an approximation algorithm method to find the highest scoring locally optimal alignments between a query sequence and a database (Barton, 1996).

The identification of homology or similar proteins that have previously studied can enable predictions to be made for the likely characteristics of the target protein based upon those observed in the similar protein. These predictions are likely to hold true if the sequence identity between the two proteins is 40% or higher, although lower identity proteins may also share similar characteristics.

Sequence and structural analysis of *Synechocystis* AD compared to other known protein in nature was carried using several computer programs, for both primary and tertiary structures comparisons. This was done in order to gain a better understanding of the cAD and the features involved in substrate recognition, catalytic mechanism and stability. ADs differ from organism to organism. Insect alkanal (fatty aldehyde) decarbonylase differs from the cyanobacterial enzyme because it is a cytochrome P450 monooxygenase. In this case not much similarity is expected to that of the cADs (Figure 2.4).

2.2 Materials and Methods

2.2.1 Primary sequence analysis

Analysis was carried out using ClustalW (Higgins *et al*., 1994) on the amino acid sequences available. Sequences were obtained from the protein data bank (Berman *et al.*, 2000) and entered into the Bioedit programme (Hall, 1999) that aligns data using the program ClustalW (Higgins *et al*., 1994). This provides both a graphic output and quantitative values for sequence identities.

A phylogenetic tree was also calculated using the web-based program GeneBee (Brodsky *et al*., 1992; Brodsky *et el*., 1995), which calculates the relationships between the AD by building probability phylogenetic trees, based on the matrix of pairs distances between the sequences. The NCBI database was used.

2.2.2 Conserved domains if sll0208 is a Fe-S protein

BLAST and psi-BLAST (at NCBI) was base to look for proteins similar to sll0208, then aligned with *P. marinus* MIT9313 to determine if the cysteine (C) and histidine (H) residues that could bind Fe are conserved in the cyanobacterial proteins.

2.3 Results

2.3.1 Aldehyde decarbonylases in databases

BLASTp results confirm 130 proteins present in databases, which are related to the hypothetical protein sll0208 [*Synechocystis sp*. PCC 6803]. The sll0208 [*Synechocystis sp*. PCC 6803] protein is more phylogenetically related to hypothetical protein cce_0778 [*Cyanothece sp*. ATCC 51142] and last related to *Laccaria bicolor* S238N-H82 dehydrogenase protein from *Bacillus clausii* KSM-K1 (Figure 2.5).

Sequence alignment of fatty aldehyde decarbonylase (Sll0208) with the AD from *P. marinus* high-lightened a number of conserved sequences present (Figure 2.1).

```
gi|33863499|ref|NP_895059.1|_h      MPTLEMPVAAVLDSTVGSSEALPDFTSDRYKDAYSRINAIVIEGEQEAHD
Sll0208_LOCUS_BAA10217_231aa_[      ------------MPELAVRTEFDYSSEIYKDAYSRINAIVIEGEQEAYS
                                                                .  :  .:   *::*:
******************:.

gi|33863499|ref|NP_895059.1|_h      NYIAIGTLLPDHVEELKRLAKMEMRHKKGFTACGKNLGVEADMDFAREFF
Sll0208_LOCUS_BAA10217_231aa_[      NYLQMAELLPEDKEELTRLAKMENRHKKGFQACGNNLQVNPDMPYAQEFF
                                    **: :. ***:. ***.****** ****** ***:** *:.** :*:***

gi|33863499|ref|NP_895059.1|_h      APLRDNFQTALGQGKTPTCLLIQALLIEAFAISAYHTYIPVSDPFARKIT
Sll0208_LOCUS_BAA10217_231aa_[      AGLHGNFQHAFSEGKVVTCLLIQALIIEAFAIAAYNIYIPVADDFARKIT
                                    * *:.*** *:.:**. ********:******:**: ****:* ******

gi|33863499|ref|NP_895059.1|_h      EGVVKDEYTHLNYGEAWLKANLESCREELLEANRENLPLIRRMLDQVAGD
Sll0208_LOCUS_BAA10217_231aa_[      EGVVKDEYTHLNYGEEWLKANFATAKEELEQANKENLPLVWKMLNQVQGD
                                    *************** *****: :.:*** :**:*****: :**:** **

gi|33863499|ref|NP_895059.1|_h      AAVLQMDKEDLIEDFLIAYQESLTEIGFNTREITRMAAAALVS-
Sll0208_LOCUS_BAA10217_231aa_[      AKVLGMEKEALVEDFMISYGEALSNIGFSTREIMRMSSYGLAGV
                                    * ** *:** *:***:*:* *:*::***.**** **:: .*..
```

Figure 2.1. Sequence alignment of fatty aldehyde decarbonylase from *Synechocystis* sp. PCC7942_or f1593 with the AD from *P. marinus* NCBI accession number NP_895059, showing conserved sequences**.** The alignment was created using ClustalW (Thompson *et al*., 1994).

2.3.2 Conserved domains

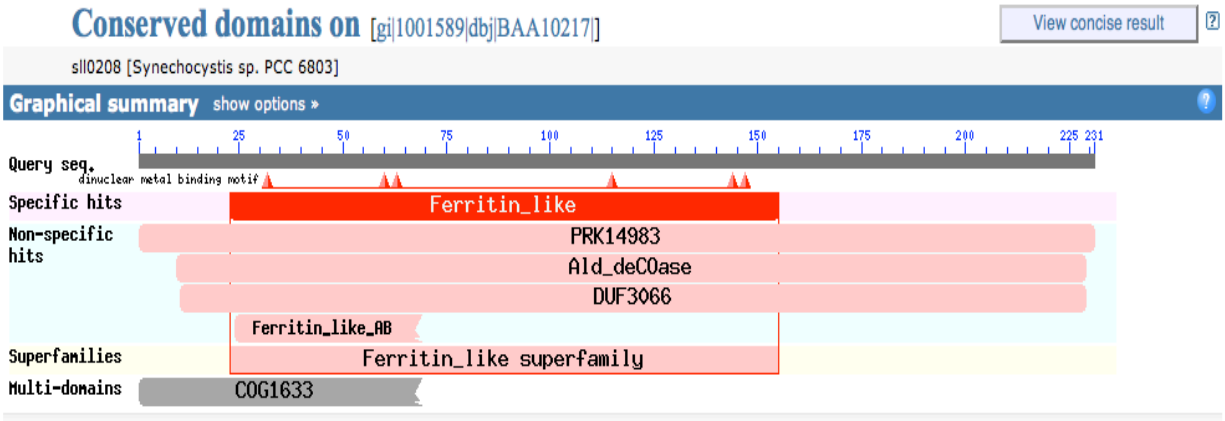


Figure 2.2 Showing conserved domains of sll0208

From the figure 2.2 above, seven domains are conserved in the sll0208 protein; namely the ferritin_like, PRK14983, Ald_deCOase, DUF3066, Ferritin_like_AB, Ferritin_like superfamily and C0G1633.

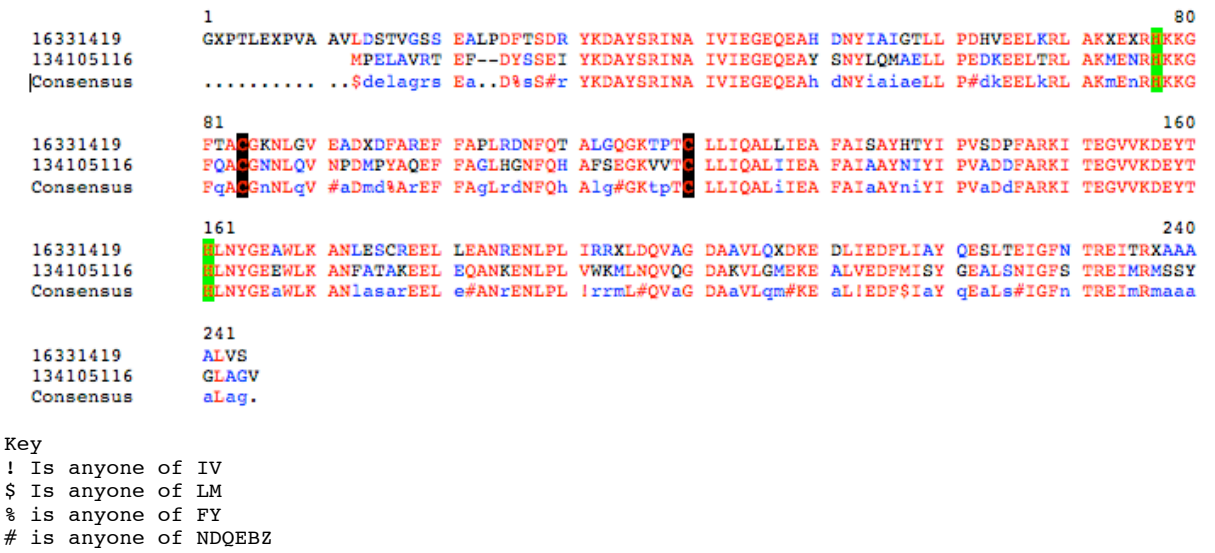


Figure 2.3 The alignment between *Synechocystis* AD (16331419) with fatty aldehyde decarbonylase from *P. marinus* MIT9313 (134105116). Both the cysteine (C) and histidine (H) residues Conserved are highlighted in black and green colour respectively.

SynADC protein is a Fe-S protein, a non heme diiron clusters residing within ferritin-like, four-helix-bundle protein with both cysteine (C) and histidine (H) residues conserved (Figure 2.3) as they are speculated to be responsible for the binding of Fe in the cyanobacterial protein. Cyanobacteria are responsible for ~ 50% of all photosynthesis on earth. Cyanobacteria contain cytoplasm structures and Carboxysomes, which are a large organized unit, part of a carbon concentrating mechanism capable to concentrate rubisco plus Carbonic anhydrase. The enzyme carbonic anhydrase catalyzes the production of CO_2 from bicarbonate.

Iron-sulfur proteins are proteins characterized by the presence of iron-sulfur clusters containing sulfide-linked di-, tri-, and tetrairon centers in variable oxidation states. Iron-sulfur clusters are present in a selection of metalloproteins, such as the ferredoxins, as well as NADH dehydrogenase, hydrogenases Coenzyme Q -

cytochrome c reductase, Succinate coenzyme Q reductase and nitrogenase (Lippard, 1994). Iron-sulfur clusters are best known for their role in the oxidation-reduction reactions of mitochondrial electron transport. Both Complex I and Complex II of oxidative phosphorylation have multiple Fe-S clusters. They have many other functions including catalysis as illustrated by aconitase, generation of radicals as illustrated by SAM-dependent enzymes, and as sulfur donors in the biosynthesis of lipoic acid and biotin. When a photon lands on photosystem 1, the electron is excited to enter the non-sulfur clusters and then the electron exits by turning into ferredoxin.

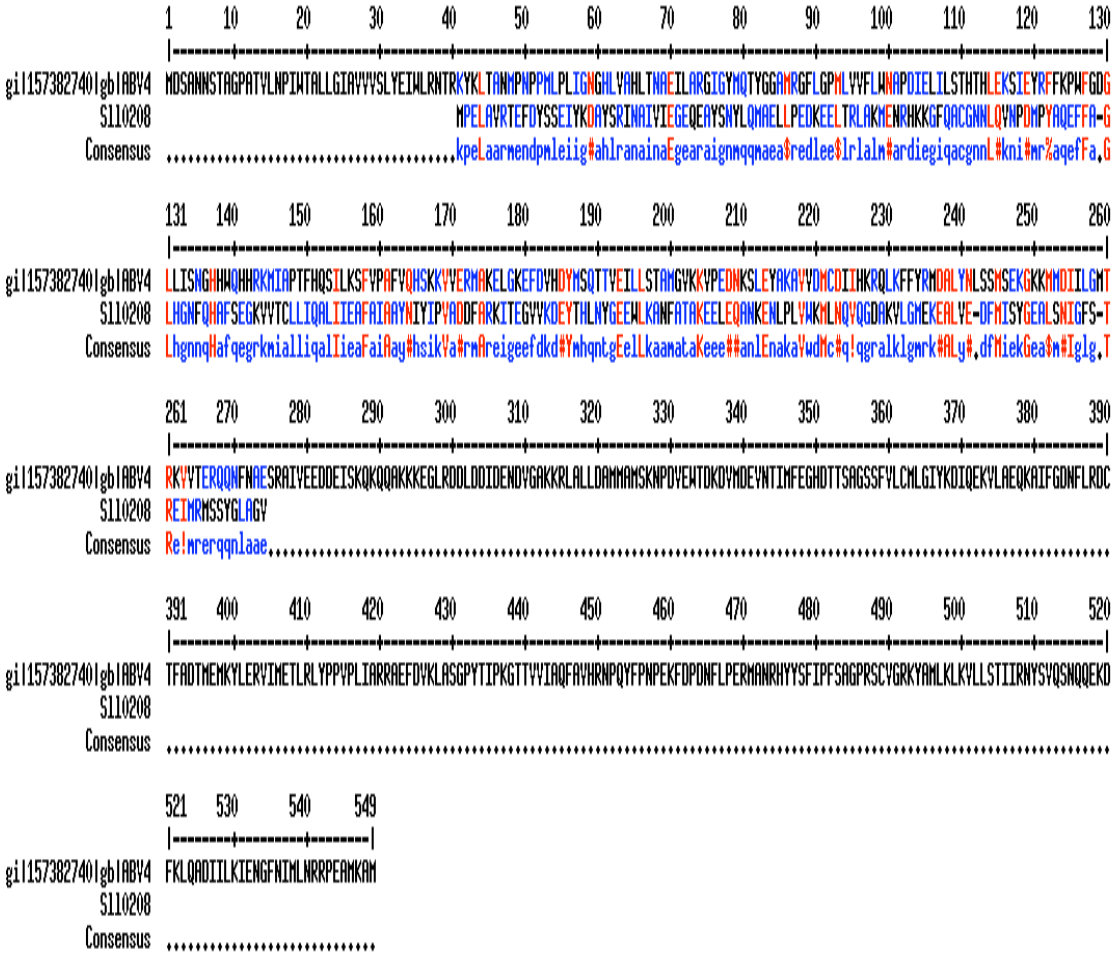


Figure 2.4. The alignment between insect alkanal (fatty aldehyde) decarbonylase (gil157382740) and *Synechocystis* AD (sll0208). Red marks show the amino acids, which are conserved both in *Synechocystis* and insects and blue marks show inserted/ mutated amino acids.

The alignment shows its clear that *Synechocystis* AD shares conserved features common with insect alkanal (fatty aldehyde) decarbonylase.

Proteins similar to s110208

```
NP_442147      1   --------MPELAVRT-EFDYSSEIYKDAYSRINAIVIEGEQEAYSNYLQMAELLPEDKEELTRLAKMENRHKKGFQACG  71
YP_001802195   1   --------MQELALRS-ELDFNSETYKDAYSRINAIVIEGEQEAYQNYLDMAQLLPEDEAELIRLSKMENRHKKGFQACG  71
YP_001660323   1   --------MPELAVPL-ELDFTSETYKSAYSRINAIVIEGEYEANSNYIQLADILTDNKEELHRLAKMENRHMKGFQACG  71
ZP_00514700    1   --------MQELAVRS-ELDFNSETYKDAYSRINAIVIEGEQEAYENYIDMGELLPGDKDELIRLSKMENRHKKGFQACG  71
CAO90780       1   --------MPELAVPL-ELDFTSETYKSAYSRINAIVIEGEYEANSNYIQLADILTDNKEELHRLAKMENRHMKGFQACG  71
ZP_01728578    1   --------MQELALRS-ELDFNSETYKDAYSRINAIVIEGEQEAHQNYIDMAQLLPEDEAELIRLSKMENRHKKGFQACG  71
YP_002370707   1   --------MQELVQRS-ELDFTNPTYKDAYSRINAIVIEGEQEAHQNYIDMAQLLPEHQEELIRLSKMENRHKKGFEACG  71
NP_489323      1   --------MQQVAADL-EIDFKSEKYKDAYSRINAIVIEGEQEAYENYIQLSQLLPDDKEDLIRLSKMESRHKKGFEACG  71
ZP_03273549    1   --------MPQLETIT-ELDFQNETYKDAYSRINAIVIEGEQEASDNYIKLGEMLPEEREELIRLSKMEKRHKKGFQACG  71
YP_323043      1   --------MQQVAADL-EIDFKSEKYKDAYSRINAIVIEGEQEAYENYIQLSQLLPDDKEDLIRLSKMESRHKKGFEACG  71
BAI93031       1   --------MPQLETIA-ELDFQNETYKDAYSRINAIVIEGEQEAYDNYIKLGEMLPEEREELIRLSKMEKRHMKGFQACG  71
ZP_06380209    1   --------MPQLETIA-ELDFQNETYKDAYSRINAIVIEGEQEASDNYIKLGEMLPEEREELIRLSKMEKRHMKGFQACG  71
ZP_05027177    1   MQTGENLLMQQLTVSQ-ELDFNSETYKDAYSRINAIVIEGEQEAHQNYIQLAELLPDQKDELTSLAKMENRHKKGFQACG  79
YP_003421663   1   --------MQELALRS-ELDFNSETYKDAYSRINAIVIEGEQEAYQNYLDMVHMLPKNKDELVRLSKMENRHKTGFQACG  71
ZP_01628096    1   --------MQQLAAEL-KIDFQSEKYKDAYSRINAIVIEGEQEAHDNYITLGEMLPELKDELIRLSKMESRHKKGFEACG  71
YP_001865325   1   --------MQQLTDQSKELDFKSETYKDAYSRINAIVIEGEQEAHENYITLAQLLPESHDELIRLSKMESRHKKGFEACG  72
YP_002481151   1   --------MPQVQSPS-AIDFYSETYQDAYSRIDAIVIEGEQEAHDNYLKLTELLPDCQEDLVRLAKMEARHKKGFEACG  71
YP_003722152   1   --------MQQLVEEIEKIDFQSEKYKDAYSRINAIVIEGEQEAHENYITLAKLLPESKEELMRLSKMESRHKKGFEACG  72
ZP_07109384    1   --------MQQLEASP-AIDFETATYKDAYSRINAIVIEGEQEAYDNYIRLGEMLPDQKDVLIALSKMENRHMKGFQACG  71
ZP_05038068    1   --------MQTLEVSP-AMDFQSETYKDAYSRINAIVIEGELEANNNYKQLSEHLGDFKDDLLKLARMENRHMKGFQACG  71

NP_442147     72   NNLQVNPDMPYAQEFFAGLHGNFQHAFSEGKVVTCLLIQALIIEAFAIAAYNIYIPVADDFARKITEGVVKDEYTHLNYG  151
YP_001802195  72   KNLNVTPDMDYAQQFFAELHGNFQKAKAEGKIVTCLLIQSLIIEAFAIAAYNIYIPVADPFARKITEGVVKDEYTHLNFG  151
YP_001660323  72   QNLKITPDMDYAREFFSSLHNNFQIAYAEGKVVTCLLIQSLIIEAFAIAAYNIYIPVADPFARKITEGVVKDEYLHLNFG  151
ZP_00514700   72   KNLKVTPDMDYAERFFSQLHGNFQTAKAEGKIVTCLLIQSLIIEAFAIAAYNIYIPVADPFARKITENVVKDEYSHLNFG  151
CAO90780      72   QNLQITPDMEYAKEFFSSLHNNFQIAYAEGKVVTCLLIQSLIIEAFAIAAYNIYIPVADPFARKITESVVKDEYLHLNFG  151
ZP_01728578   72   KNLDVTPDMDYAQQFFSQLHNNFQTAKAEGKIVTCLLIQSLIIEAFAIAAYNIYIPVADPFARKITEGVVKDEYTHLNFG  151
YP_002370707  72   NNLSVTPDMQYAQEFFSSLHGNFQKAKAEGKIVTCLLIQSLIIEAFAIAAYNIYIPVADPFARKITEGVVKDEYTHLNFG  151
NP_489323     72   RNLQVSPDMEFAKEFFAGLHGNFQKAAAEGKIVTCLLIQSLIIECFAIAAYNIYIPVADDFARKITEGVVKDEYSHLNFG  151
ZP_03273549   72   RNLEVTPDMDFGREFFAKLHGNFQKAAAEGKLVTCLLIQSLIIESFAIAAYNIYIPVADPFARKITEGVVKDEYEHLNFG  151
YP_323043     72   RNLQVSPDIEFAKEFFAGLHGNFQKAAAEGKVVTCLLIQSLIIECFAIAAYNIYIPVADDFARKITEGVVKDEYSHLNFG  151
BAI93031      72   RNLEVSPDMDFGREFFAQLHGNFQKAAAEGKLVTCLLIQSLIIESFAIAAYNIYIPVADPFARKITEGVVKDEYEHLNFG  151
ZP_06380209   72   RNLEVSPDMDFGREFFAQLHGNFQKAAAEGKLVTCLLIQSLIIESFAIAAYNIYIPVADPFARKITEGVVKDEYEHLNFG  151
ZP_05027177   80   RNLSVTADMEFAKEYFSDLHQNFQTAAASGNIVTCLLIQSLIIECFAIAAYNIYIPVADPFARKITEGVVKDEYMHLNFG  159
YP_003421663  72   KNLNVIPDMQYAKEFFSQLHENFQIAKNEKKVVTCLLIQALIIEAFAIAAYNIYIPVADPFARKITENVVKDEYKHLNFG  151
ZP_01628096   72   RNLSVKPDMPFAQKFFSGLHENFQKAAAEGQVVTCLLIQSLIIECFAIAAYNIYIPVADDFARKITEGVVKDEYSHLNFG  151
YP_001865325  73   RNLAVTPDLQFAKEFFSGLHQNFQTAAAEGKVVTCLLIQSLIIECFAIAAYNIYIPVADDFARKITEGVVKEEYSHLNFG  152
YP_002481151  72   RNLKVTPDMEFAQQFFADLHNNFQKAAAANKIATCLVIQALIIECFAIAAYNIYIPVADDFARKITENVVKDEYTHLNFG  151
YP_003722152  73   RNLQVTPDMQFAKEFFSGLHQNFQTAAAAGNVVTCLLIQSLIIECFAIAAYNIYIPVADDFARKITEGVVKEEYSHLNFG  152
ZP_07109384   72   RNLKVTPDMEFAKDFFAALHSNFQEAAAAGKVVTCLVIQALIIECFAIAAYNIYIPVADDFARKITEGVVKDEYNHLNFG  151
ZP_05038068   72   KNLSVNPDMPFAKEFFAQLHDNFQTALAEGKIVTCLLIQSLIIETFAISAYNIYIPVADDFARKITEGVVKDEYMHLNFG  151

NP_442147     152  EEWLKANFATAKEELEQANKENLPLVWKMLNQVQGDAKVLGMEKEALVEDFMISYGEALSNIGFSTREIMRMSSYGLAGV  231
YP_001802195  152  EVWLKEHFEASKAELEDANKENLPLVWQMLNQVEKDAEVLGMEKEALVEDFMISYGEALSNIGFSTREIMKMSAYGLRAA  231
YP_001660323  152  EEWLKANFETAKEELEAANRANLPIVWRMLNQVENDARVLGMEKEALVEDFMISYGEALSNIGFSTRDIMRMSAYGLTAV  231
ZP_00514700   152  EVWLKENFEASKAELEQANKENLPIVWQMLNEVEDDAEILGMEKEALVEDFMISYGEALGNIGFSTREIMKMSAHGLAAV  231
CAO90780      152  EEWLKANFETAKEELEAANRANLPIVWRMLNQVEDDARVLAMEKEALVEDFMISYGEALNNIGFSTRDIMRMSAYGLTAV  231
ZP_01728578   152  EIWLKEHFEASKAELEEANKKNLPIVWQMLNQVEKDAEVLGMEKEALVEDFMISYGEALSNIGFSTREIMKMSSHGLSAA  231
YP_002370707  152  EVWLQEHFEESKAELEEANKANLPIVWEMLNQVEGDAKVLGMEKEALVEDFMISYGEALSNIGFSTRDIMRMSSHGLVAA  231
NP_489323     152  EVWLQKNFAQSKAELEEANRHNLPIVWKMLNQVADDAAVLAMEKEALVEDFMIQYGEALSNIGFTTRDIMRMSAYGLTAA  231
ZP_03273549   152  EEWLKAHFEESKAELEEANRQNLPLVWKMLNQVEKDASILGMEKEALIEDFMIAYGEALSNIGFTTRDIMRMSAYGLAGV  231
YP_323043     152  EVWLQKNFAQSKAELEEANRHNLPIVWKMLNQVADDAAVLAMEKEALVEDFMIQYGEALSNIGFTTRDIMRMSAYGLTAA  231
BAI93031      152  EEWLKAHFEESKAELEEANRQNLPLVWKMLNQVEKDASILGMEKEALIEDFMIAYGEALSNIGFTTRDIMRMSAYGLAGV  231
ZP_06380209   152  EEWLKAHFEESKAELEEANRQNLPLVWKMLNQVEKDASILGMEKEALIEDFMIAYGEALSNIGFTTRDIMRMSAYGLAGV  231
ZP_05027177   160  EEWLKENFEASKTELEQANKQNLPLVWRMLNQVEKDAHILGMEKDALVEDFMIAYGEALSNIGFTTRDIMRMSAYGLTAA  239
YP_003421663  152  EVWLGENFESSKIELEEANKTNLPIVWKMLNEVEQDASILGMEKEALVEDFMISYGEALGNIGFSTREIMRMSSHGLRAS  231
ZP_01628096   152  EVWLKENFAQSKAELEAANRQNLPIVWKMLNEVENDAHVLAMEKEALVEDFMIQYGETLSNIGFTTRDIMKMSAYGLTAA  231
YP_001865325  153  EVWLKEHFAESKAELELANRQNLPIVWKMLNQVEGDAHTMAMEKDALVEDFMIQYGEALSNIGFSTRDIMRLSAYGLIGA  232
YP_002481151  152  EEWLKANFDSQREEVEAANRENLPIVWRMLNQVETDAHVLGMEKEALVESFMIQYGEALENIGFSTREIMRMSVYGLSAA  231
```

```
YP_003722152   153   EVWLKEHFAESKAELDDANRQNLPIVWQMLNQVADDARVLAMEKEALVEDFMIQYGEALSNIGFTTRDIIRLSAYGLATV   232
ZP_07109384    152   EEWLKAHFEESKAEVDTANRQNLPIVWRMLNQVEDDARVLGMEKDALVEDFMIAYGEALSNIGFTTRDIMRMSAYGLTAA   231
ZP_05038068    152   EEWLKANFEASKAELETANRANLPLIWKMLNQVEEDAAVLGMEKDALIEDFMITYGEALANIGFSARDVMRLSAQGLAAV   231
```

Figure 2.5 Sequence alignments for the 20 fatty aldehyde decarbonylase proteins found on NCBI database present in nature.

2.3.3 Evolution background of decarbonylase enzyme in nature

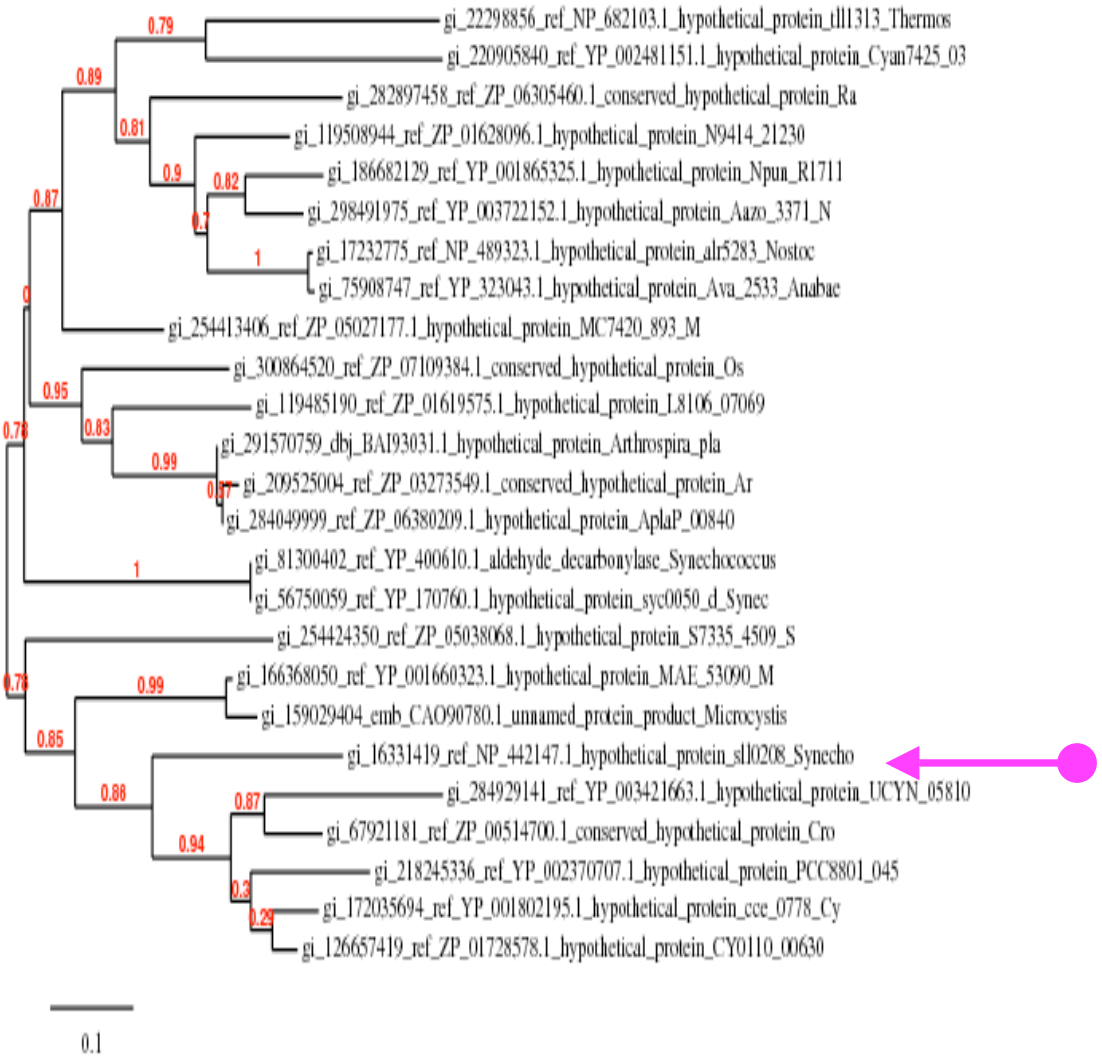


Figure 2.6 A phylogeny tree constructed using a PAM demonstrating the evolution background of decarbonylase enzymes in nature. The x-axis indicating the decreasing homology between sequences and therefore the decreasing evolutionary relationship. The pink arrow shows the location of *Synechocystis* AD (sll0208) protein on the tree.

From the phylogenetic tree above it is clear that NP_682103.1_hypothetical protein is the most distantly related decarbonylase and CA090780.1_unnamed *protein and YP*003421663.1_hypothetical protein are the most closely related.

2.3.3 Comparisons between *Synechocystis* AD (sll0208) and some of other fatty aldehyde decarbonylase proteins present in nature as shown in the phylogeny tree in Figures 2.6.

a. 682103.1 vs 442147

```
682103          MTTATATPVLDYHSDRYKDAYSRINAIVIEGEQEAHDNYIDLAKLLPQHQEELTRLAKME 60
442147          MPELAVRTEFDYSSEIYKDAYSRINAIVIEGEQEAYSNYLQMAELLPEDKEELTRLAKME 60
                *.  :. . :** *: ********************:.**:::*:***:.:**********

682103          ARHKKGFEACGRNLSVTPDMEFAKAFFEKLRANFQRALAEGKTATCLLIQALIIESFAIA 120
442147          NRHKKGFQACGNNLQVNPDMPYAQEFFAGLHGNFQHAFSEGKVVTCLLIQALIIEAFAIA 120
                 ******:***.**.*.*** :*: **  *:.***:*::***..***********:****

682103          AYNIYIPMADPFARKITESVVKDEYSHLNFGEIWLKEHFESVKGELEEANRANLPLVWKM 180
442147          AYNIYIPVADDFARKITEGVVKDEYTHLNYGEEWLKANFATAKEELEQANKENLPLVWKM 180
                ********:** *******.******:***:** *** :* :.* ***:**: ********

682103          LNQVEADAKVLGMEKDALVEDFMIQYSGALENIGFTTREIMKMSVYGLTGA 231
442147          LNQVQGDAKVLGMEKEALVEDFMISYGEALSNIGFSTREIMRMSSYGLAGV 231
                ****:.*********:********.*. **.****:*****:** ***:*.
```

b. AD_*Synochococcus* vs 442147

```
AD_Synochococcus    MTTATATPVLDYHSDRYKDAYSRINAIVIEGEQEAHDNYIDLAKLLPQHQEELTRLAKME 60
sll0208             MPELAVRTEFDYSSEIYKDAYSRINAIVIEGEQEAYSNYLQMAELLPEDKEELTRLAKME 60
                    *.  :. . :** *: ********************:.**:::*:***:.:**********

AD_Synochococcus    ARHKKGFEACGRNLSVTPDMEFAKAFFEKLRANFQRALAEGKTATCLLIQALIIESFAIA 120
sll0208             NRHKKGFQACGNNLQVNPDMPYAQEFFAGLHGNFQHAFSEGKVVTCLLIQALIIEAFAIA 120
                     ******:***.**.*.*** :*: **  *:.***:*::***..***********:****

AD_Synochococcus    AYNIYIPMADPFARKITESVVKDEYSHLNFGEIWLKEHFESVKGELEEANRANLPLVWKM 180
sll0208             AYNIYIPVADDFARKITEGVVKDEYTHLNYGEEWLKANFATAKEELEQANKENLPLVWKM 180
                    ********:** *******.******:***:** *** :* :.* ***:**: ********

AD_Synochococcus    LNQVEADAKVLGMEKDALVEDFMIQYSGALENIGFTTREIMKMSVYGLTGA 231
sll0208             LNQVQGDAKVLGMEKEALVEDFMISYGEALSNIGFSTREIMRMSSYGLAGV 231
                    ****:.*********:********.*. **.****:*****:** ***:*.
```

c. Hypothetical _protein _ MAE_53090 vs 442147

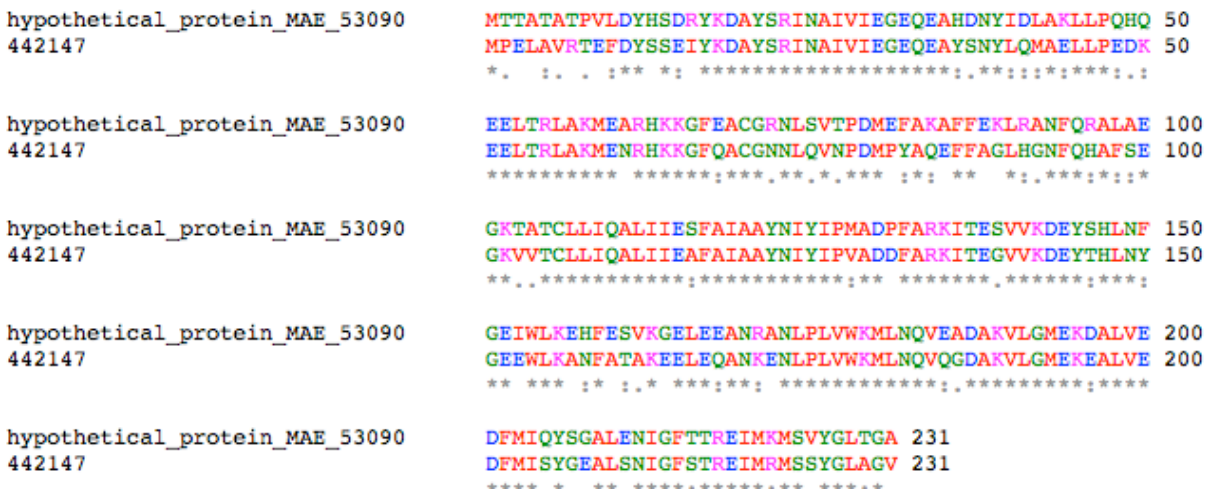


Figure 2.6. Detailed alignment of the protein sequences of three known fatty aldehyde decarbonylase using the ClustalW function in BioEdit (Hall, 1997; Higgins *et al*., 1994 and Thompson *et al*., 1994) AD active site residues are marked with a * shading threshold was 70%.

CHAPTER 3:

PROTEIN EXPRESSION AND PURIFICATION

3.0 Introduction

In order to curry out studies with cAD from *Synechocystis* species, it was essential to use the recombinant protein, which could be expressed, purified to high homogeneity and easy to crystalise. The putative SynADC was isolated from *Synechocystis* sp. PCC 6803 and cloned into pET160/GW/D-TOPO by Dr Christoph Edner (University of Exeter). The research I carried out was to the over-expression and the purification of the protein to carry out crystallisation trials in order to solve the protein structure. Attempts to purify the enzyme without a Tag (SynADC) proved problematic, the enzyme aggregated very easily during purification, dialysis and concentration.

In view of the problem with aggregation of the recombinant SynADC without a Tag, the decision was taken to re-clone the SynADC gene into pCold vector; this was done by Dr Christoph Edner in the laboratory of Prof. Nicholas Smirnoff at Exeter University. This chapter discusses the expression and purification of SynADC gene and resulting effects on the solubility and oligomeric state of the protein.

3.1 Materials and Methods

Sample preparation

All samples were filtered through 0.45 µl filters to remove small particulate matter and to avoid the risk of column blockage, reduce the stringent wash procedure and extend the life of packed chromatography medium. The samples were concentrated down where necessary to approximately 10 mg/ml.

3.1.1 Reagents grade chemicals

Chemicals used in this research are analytical grade or better, and unless otherwise stated all were supplied by Sigma Aldrich. All buffers were vacuum filtered before use through a Whatman No.1 circular 55 mm filter. Aseptic conditions are always followed at all times and a fume hood was used when using chemicals harmful when inhaled.

The SDS-PAGE pre made gels used were obtained from Invitrogen ^{TM}. The water used throughout this work is of a double distilled quality, and of Purite ^{TM} quality.

3.1.2 Growth media

LB Broth (TB)

To 1L of H_2O 12 g Bacto tryptone, 24 g Bacto yeast and 4 ml glycerol were added. The pH was adjusted to pH 7.5 with 10M NaOH and the media autoclaved for 15 minutes, 121ºC. After autoclaving, separately autoclaved 100 ml of Phosphate solution was added (900 TB + 100 ml Phosphates). To make the stock Phosphate solution, 11.6g KH_2PO_4 and 62.7g K_2HPO_4 were dissolved.

LB/Amp

Ampicillin (100mg/ml) was added to 1L of LB broth after autoclaving and cooling to a final concentration of 0.1mg/ml ampicillin unless otherwise indicated.

Agar plates

15g Agar was added to 1L LB broth prior to autoclaving. The agar was cooled to approximately 40ºC and filtered ampicillin added to a final concentration of 0.05mg/ml.

Glycerol stocks

Overnight cultures were used to prepare glycerol stocks for long-term storage. These were performed using PROTECT BACTERIAL PRESERVERS (Fisher Scientific) according to the manufacturer's instructions. These are ceramic beads surrounded by cryo-protectant media and prevent the requirement for freeze thawing. All stocks were stored at -80ºC until required

3.1.3 Cell culture

Prior to this work, the putative SynADC was cloned into pET160/GW/D-TOPO, and over expressed in *E. coli*. 20ml LB/amp was inoculated from a single *E. coli* colony. This was grown overnight at 37ºC in 10 ml 2x-YT media and used to seed 1L LB/amp contained in a 2L flask, which was grown at 37ºC with agitation to approximately 0.6 – 0.8 at 600 nm and induced with 0.5mM IPTG. Cells were harvested by centrifugation (8,000g, 15 min).

3.1.4 Expression SynADC

Expression studies were carried out using *E. coli* BL21. 50ml of LB/Amp media was inoculated with 2% overnight culture and grown at 37ºC with agitation. Expression parameters were investigated and induction occurred at an OD_{600} of approximately 0.6 – 0.8 using 1 ml of 100 mM IPTG at 37ºC. *E.coli* cell cultures were harvested chilled on ice for 20 min then cells were harvested by centrifugation at 8,000g for 15 min at 4ºC. Following centrifugation, the cells pellet was washed once with 50 mM Tris-HCl pH 7.5, then transferred into Falcon tubes, froze in liquid nitrogen and stored at -8ºC. Total and soluble protein fractions were assessed by SDS-PAGE after 4hrs growth (post –induction temp 37ºC) or overnight growth.

3.1.5 Handling and storage of protein solutions

During the purification, to avoid enzyme degradation and loss of activity, all enzyme samples were handled at 4ºC and stored at this temperature on a short-term basis. During use, the enzyme solutions were kept on ice. Smaller protein samples were concentrated using Vivaspin concentrators with a membrane pore size cut-off of 10 kDa. The enzyme samples were placed into the concentrator and centrifuged at 4000g until the appropriate concentration had been reached.

3.1.6 SDS- polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used to analyse protein content. Protein gels were run on a Bio-Rad Mini-protein gel apparatus. Freshly made gels were prepared with 30%

acrylamide /0.8% bisacrylamide to produce a 12% separating gel and a 6% stacking gel (Laemmli, 1970). Protein samples (10 µl of the protein extract, 4 µl of sample buffer, 1.6 µl of reducing agent and 0.4 µl of water) that were to be loaded onto the protein gels were prepared by heating samples on a Techni Dri-Block DB 2A, for 10 minutes.

3.1.7 Pre -made SDS gels

The pre-made gels were prepared according to the manufactures instructions (NuPAGR® Bis-Tris Mini Gel protocol).

3.1.8 SDS-PAGE gel running procedure

The gel was inserted into electrophoresis tank in accordance with the instruction of the manufacturer. The electrode reservoirs were filled with 1x MES running buffer (Initrogen). 10 µl of each sample was loaded onto the gel using a 50 µl Hamilton syringe, molecular weight standards were loaded into the final well. The gel was run at 200 V for approximently 1 hour.

3.1.9 SDS gel staining and De-staining

This was done according to the manufacturer's instructions (SimplyBlue ™ Safestain – Invitrogen).

3.1.10 Purification of SynADC

3.1.10.1 Introduction to Purification

In order to study an enzyme it must be isolated from other proteins in the organism in order to remove any inhibitors or proteins with similar functions. This can be achieved by exploiting enzymes properties, such as its charge and size, in order to separate it from other proteins. Different types of purification include ion exchange, affinity and hydrophobic chromatography and gel filtration.

Ion Exchange Chromatography

Ion exchange chromatography uses a porous resin that is charged with ionised groups on its surface. The resin can be charged with anions or cations depending on the properties of the proteins being purified. Proteins that bind to the column can be eluted by a salt gradient, which alters the number of ions in the buffer and therefore decreases the affinity by competing with the protein for the column media. The protocol is designed so that protein of interest is eluted during the low to high salt gradient.

Gel Filtration Chromatography

Gel filtration or size exclusion chromatography separates proteins according to the molecular weight of the native protein. The medium is formed by beads of a cross linked gel which contain pores of a known size. Smaller proteins will become caught in the pores and their movement will be retarded, whereas larger proteins will travel around the beads and be eluted first. These columns are calibrated with standard proteins prior to use and using the elution date, helps to determine an approximate native Molecular Mass (M_r) for the proteins using the calculations below.

$$K_{av} = \frac{V_e - V_o}{V_t - V_o}$$

Equation 3.1. Calculation of K_{av} where V_e = elution volume of the protein, V_o = void volume the column (44 ml), and V_t = total volume of the column (122 ml).

A rearrangement of the equation of the line from the calibration is used to calculate the log molecule weight from the K_{av} value.

$$\log MW = \frac{1.0185 - K_{av}}{0.3457}$$

Equation 3.2. Calculation of the log of the molecular weight

Affinity Chromatography

Affinity chromatography uses a medium that has a ligand bound to it, which will interact specifically with the protein of interest. This is a common method for recombinant protein that can have specific tags added such as a His-tag that will interact with Ni^{+}. Adding a protease that specifically cleaves the tag or a buffer, which competes with the protein/ ligand interaction, as it has a greater affinity for the ligand, can elute protein. Altering the pH or increasing the ionic strength of the buffer decreases non-convalent interaction between the ligand and protein and will also cause protein to be eluted from the column.

Buffer	Component
A	20 mM Na-phosphate pH 7.5, 500 mM NaCl, 20 mM Imidazole, 0.5 mM PMSF, and 1 mM Tris Hydroxypropyl phosphine (THP)
B	20 mM Na-phosphate pH 7.5, 500 mM NaCl, 100 mM Imidazole, 0.5 mM PMSF, 1 mM THP
C	20 mM Na-phosphate pH 7.5, 500 mM NaCl, 250 mM Imidazole 0.5 mM PMSF, 1 mM THP
D	50mM Tris-HCl pH 7.5, 10% glycerol, 0.1 M NaCl, 1 mM THP

Table 3. Shows buffers used during affinity chromatography

3.1.10.2 Sample preparation

The cell extracts were prepared by adding 40 ml of Buffer A to the cell pellet. The cells were stirred with a magnetic stirrer until no clumps were visible. Lysozyme (1 mg/ml) was added and kept at room temperature for 30 min to lyse the bacterial cell walls, cells were then sonicated 6 x 10 sec on ice at 80% amplitude with 10 second cooling intervals. Centrifugation was then carried out (16000g, 15 min at 4ºC) to remove the cell debris. The cell debris pellet was discarded but retaining the supernatant, which was then, filtered through a 0.2 mM filters before His Trap

chromatography.

3.1.10.3 Purification of recombinant SynADC

15 g (damp weight) of cells were resuspended in 40 ml 50 mM Tris-HCl buffer, pH 8.0, containing 100 mM NaCl, 10 mM imidazole, 5% glycerol and 0.5 mg/ml of lysozyme, protease inhibitor tablet (Roche) for 1 h on ice and lysed by sonication at maximum power using 2 s pulses separated by 8 s to prevent overheating for a total time of 30 min. The supernatant was separated from cell pellet by centrifugation at 15000 g at 4ºC for 30 min.

Purification of SynADC was achieved by affinity chromatography using a Ni-affinity column and taking the advantage of the expressed N-terminal His-tag. The supernatant from cell lysis was loaded onto the column and the column was washed with 25 ml of buffer A: 20 mM potassium phosphate, pH 7.4, 500 mM NaCl, 20 mM imidazole, 5% glycerol at a flow rate of 2 ml/min. SynADC was eluted from the column with 10 ml of buffer B: 20 mM potassium phosphate, pH 7.4, 500 mM NaCl, 500 mM imidazole, 5% glycerol at a flow rate of 2 ml/min. The fractions containing pure protein were pooled and dialyzed against the final assay buffer: 100 mM potassium phosphate, pH 7.2 containing 100 mM NaCl and 10% glycerol. After dialysis, the protein was concentrated using Amicon Ultra-15 centrifugal filters to a concentration >500 mM. The purified protein was judged to be better than 95 % pure as determined by SDS-PAGE (Figure 3.5.2).

3.1.10.4 Affinity Chromatography; His Trap FF (GE Healthcare)

A number of buffer solutions (Table 3) were prepared and used during enzyme purification procedures. The columns (HisTrapFF, 2x1 ml bed volume, GE Healthcare) were washed first with 20 ml of water and then equilibrated with 15 ml buffer A. Crude extracts (Cr) were loaded to the column (2 x ml bed volume at 4^0C, the flow-through (Ft) was collected and then recycled for 30 min at 1 ml / min. The column was washed with 15 ml of buffer A (Wa), followed by elution of 10 ml buffer B (B) and elution of 10 ml buffer C (C), Buffer exchange method followed, 2 x PD-10 columns were eluted with 3.5 ml of Buffer D (D). SDS gels were prepared on the collected fraction samples (Cr, Ft, Wa, B, C and D).

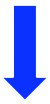
3.1.10.4 Gel 200 Filtration (GF)

All samples were filtered through a 0.45 µl filters to remove small particle matters and to avoid the risk of blockage, reduce the stringent wash procedure and extend the life of packed chromatography medium. The samples were concentrated down where necessary to reasonable concentration.

A Superdex 200 HiLoad 16/60 (GE Healthcare) gel filtration column (column volume 120 ml) was equilibrated with the GF buffer (Figure 3.1) for 2 hr at a flow rate of 1 ml/min. After equilibration, 1 ml of protein sample was applied directly onto the column via AKTA purifier system using a syringe. Samples were eluted isocratically from the gel filtration column, using a single buffer system. After sample application the entire separation took place as a one-column volume of buffer passes through the column. The flow rate was set to 1 ml/min maintained using a pump within the chromatography system; to allow time for molecules to diffuse in and out of the matrix. Fractions corresponding to protein peaks were analyzed using SDS-PAGE.

1. 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% glycerol, 0.5 mM THP

2. 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.5 mM THP (glycerol was removed from the preparation)



3. 25 mM HEPES – NaOH pH 7.8, 5% $(NH_4)_2 SO_4$ w/v, 0.5 mM THP (Tris-HCl was substituted with HEPES - NaCl, 100 mM NaCl was substituted with 5% $(NH_4)_2 SO_4$ w/v).

Figure 3.1 Series of GF and Sample buffers used during SynADC protein purification. All buffers were prepared a night before the experiment; filtered and kept cool in the fridge to reach the same temperature as the column (4ºC), 0.5 mM THP was added before the experiment.

3.1.10.5 Protein concentration determination

The protein concentration of SynADC was both determined by Bradford assay (Bradford 1976) and using the extinction coefficient calculated using Gill and von Hippel 1989 method. The extinction coefficient at 280 nm was 19.9 mM-1cm-1. SynADC concentrations used for protein crystallisation reported in this study were determined by measuring the absorbance at 280 nm based on this extinction coefficient.

Warburg and Christian (1941)

Absorbance at A280 was measured on a WPA UV1101 Biotech Photometer and the result used in the equation

$$A_{280} = \varepsilon cl$$

Equation 3.3. Beer-Lambert Law: A = the absorbance at 280 nm, ε = the extinction coefficient of the protein, c = concentration (mg/ml) and l = cuvette path length (1 cm in this incidence)

Bradford (1976)

800µl of water was mixed with 200µl of Bradford's reagent (BioRad, 500-0006), 20µl of protein and incubated for 10 minutes. 20µl BSA at concentrations 0.1 – 1mg/ml was used to produce a standard curve and the absorbance of each sample read at 595nm. The standard curve of protein concentration *vs.* absorbance was plotted and used to determine the protein concentration of unknown samples.

3.1.10.6 Determination of metal content of cAD

The metal content of the enzyme was determined by inductively coupled plasma-mass spectroscopy (ICP-MS). For routine determination of the iron content of protein preparations, the well established assay based on chelation of Fe (II) by ferrozine was utilized (Stookey, 1970).

3.2 Results And Discussion

3.2.1 Expression of SynADC

The pET160/GW/D-TOPO construct was over-expressed in the *E.coli* strain at 37ºC. The over-expression of SynADC protein was investigated by induction studies to obtain the best condition for the over-expression. The optimal over-expression was achieved by addition of a final concentration of 0.5mM IPTG when the cells had reached an OD_{600nm} of approximently 0.8 with further incubation for 4 hours at 37ºC. The over-expressed of SynADC resulted in soluble protein being produced at the expected size ~ 29 kDa (figure 3.1) as reported by Schirmer *et al*., 2010.

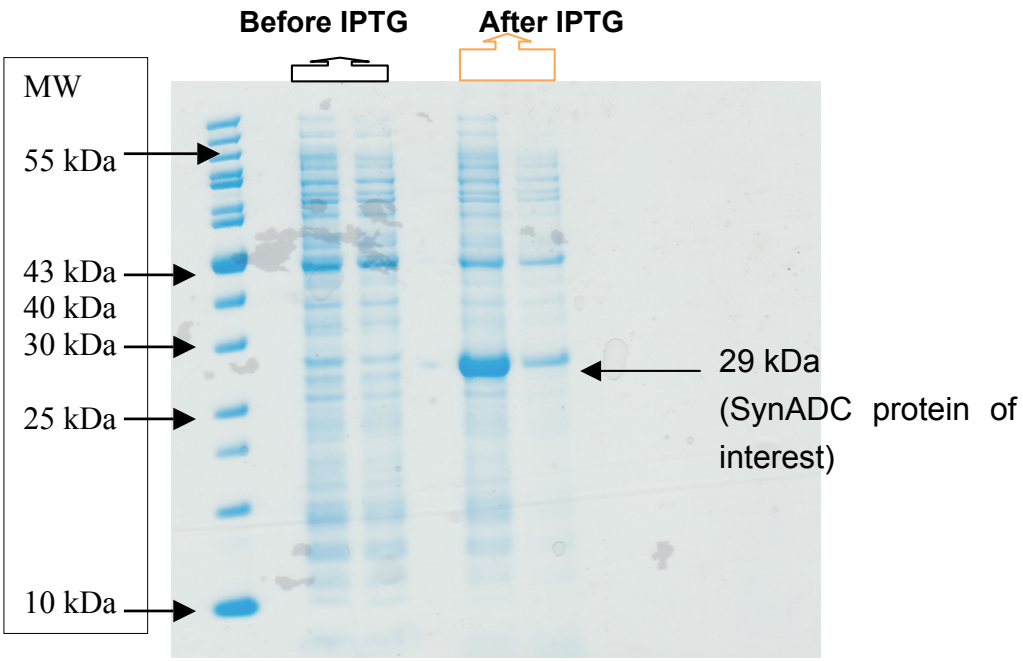


Figure 3.1 SDS-PAGE analysis of the over-expression of SynADC

3.2.2 Gel filtration chromatography

Gel Filtration Chromatography was used as a final step of the purification and as a tool for the estimation of the molecular weight of the protein. The concentrated protein sample was loaded onto an equilibrated HiLoad 16/60 Superdex 200 and run

as described in method (3.1.10.5). The elution profiles for SynADC are shown in (Figures 3.3, 3.4, and 3.5). The peak fractions were analyzed by SDS-PAGE and fractions that contained SynADC were pooled and concentrated. Red colorimetric screen assays were prepared to measure the enzyme activity. This assay uses dye phenol red to monitor a change in pH, which is related to activity. Since no activity could be found using this method, the SynADC concentration was followed on SDS-PAGE results instead. The GF column provided a very good purification step as can be seen from SDS-PAGE (Figure 3.3.2). This step resulted in a double fold increase in purity. The fractions were pooled and concentrated to approximately 10mg/ml for crystallisation trials.

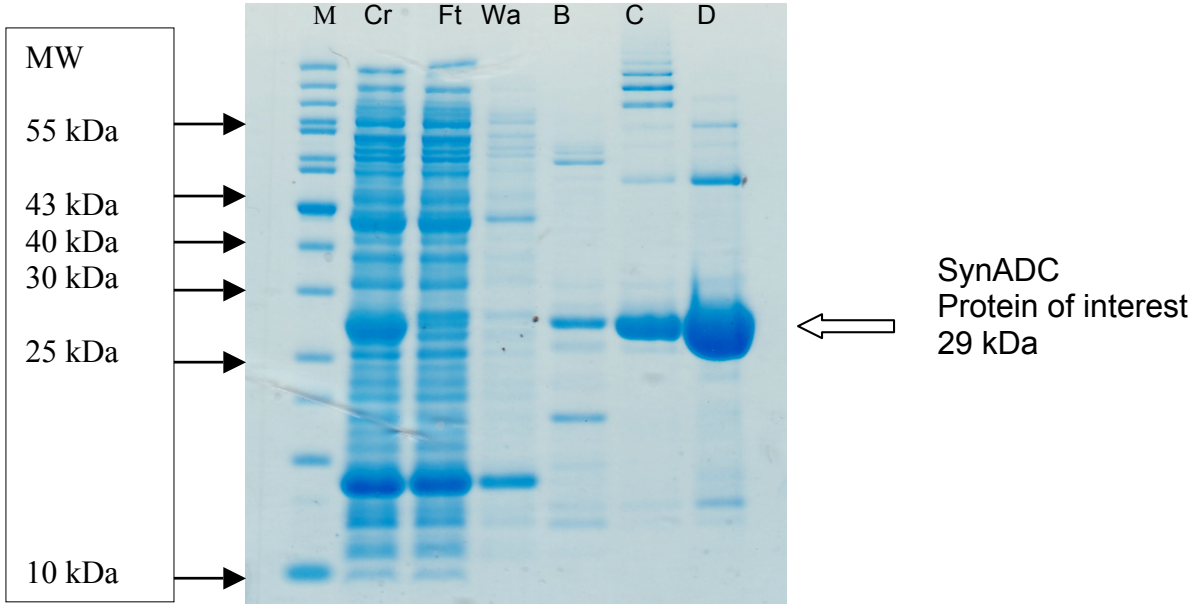


Figure 3.2 SDS-PAGE analysis of SynADC purification protocol after Ni-affinity column

(M) = Marker (Cr) = Crude extract

(Ft) = Flow-through (Wa) = Wash

(B) = Elution buffer B (100 mM Imidazole)

(C) = Elution buffer C (250 mM Imidazole)

(D) = Final enzyme preparation (buffer exchange)

Sonication of cells, centrifugation and then running the supernatant onto His Trap FF

column removes *E. coli* proteins prior to the second column being run. As can be seen from the SDS-PAGE (Figure 3.2) this step was the most effective in removing unwanted proteins from the preparation. SynADC activity was not detected neither in the pellet produced after centrifugation nor in the wash fractionation after His Trap column indicating little SynADC is lost during this step, the protein binds onto the column very well.

3.2.3 GF elution profiles under different buffer conditions

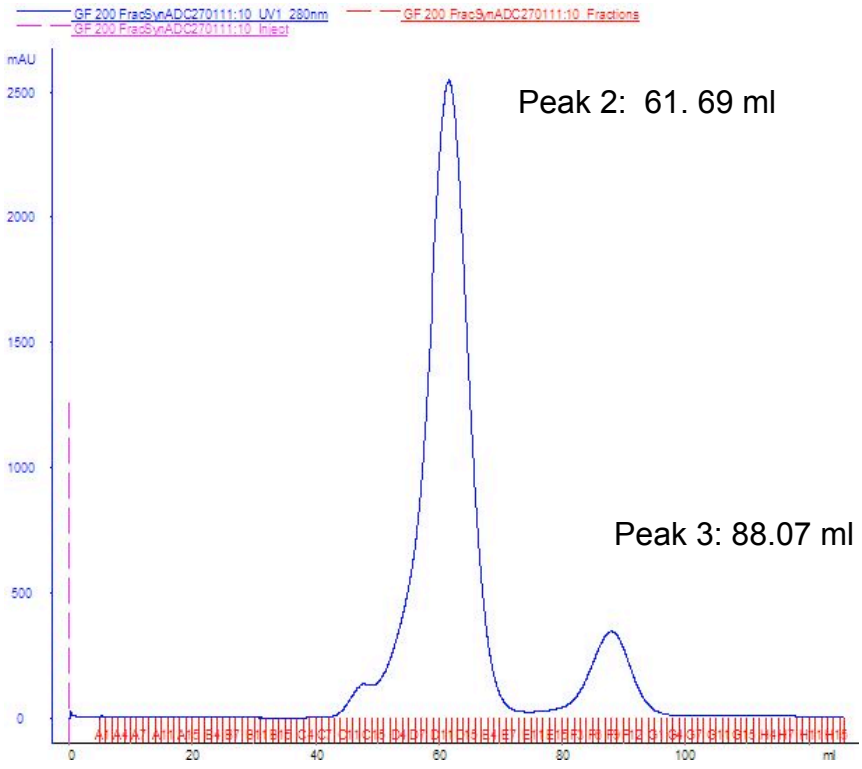


Figure 3.3 shows the elution profile of Superdex 200 GF. Using 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% glycerol, and 0.5 mM THP GF buffer, run at a flow rate 1 ml/ min. V_0 = 46.82, Peak 2 elution (D6 – E 5) samples were collected and concentrated to 10 mg/ml for crystallisation trial plates.

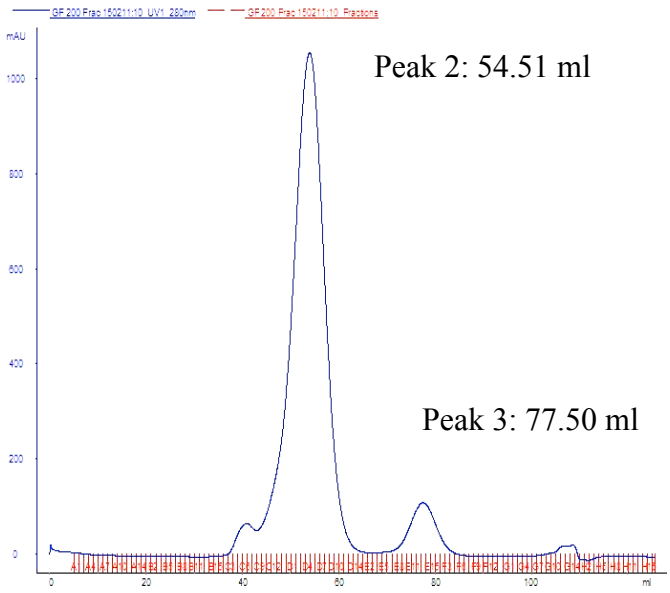


Figure 3.4 Shows the elution profile of Superdex 200 GF fractionation. GF buffer used: 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.5 mM THP at a flow rate I ml / min. D4 – D13 of peak 2 and E8 – F3 elution samples were collected, concentrated to 10 mg/ml for crystallisation trail plates.

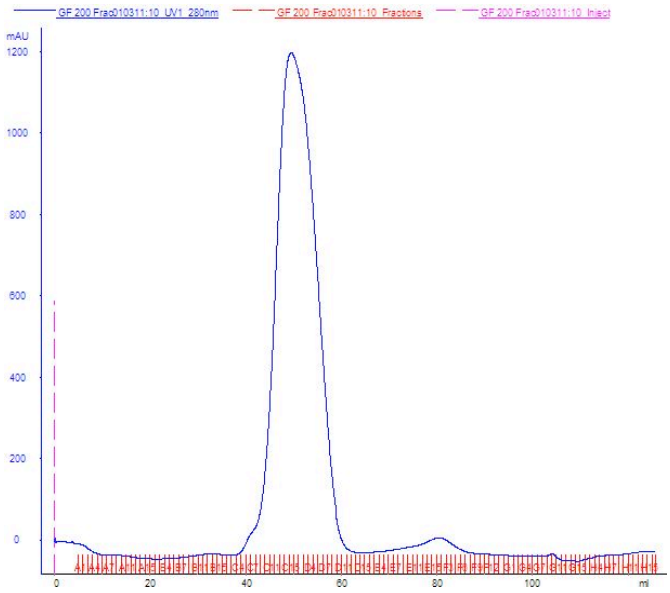


Figure 3.5 Shows the elution profile (chromatogram) of Superdex 200 GF buffer used; 25 mM HEPES – NaCl pH 7.8, 5% $(NH_3)_2 SO_4$ w/v, 0.5 mM THP at a flow rate 1 ml/min. C9 – D11 of peak 2 elution samples were collected, concentrated to 10 mg/ml for crystallisation trail plates.

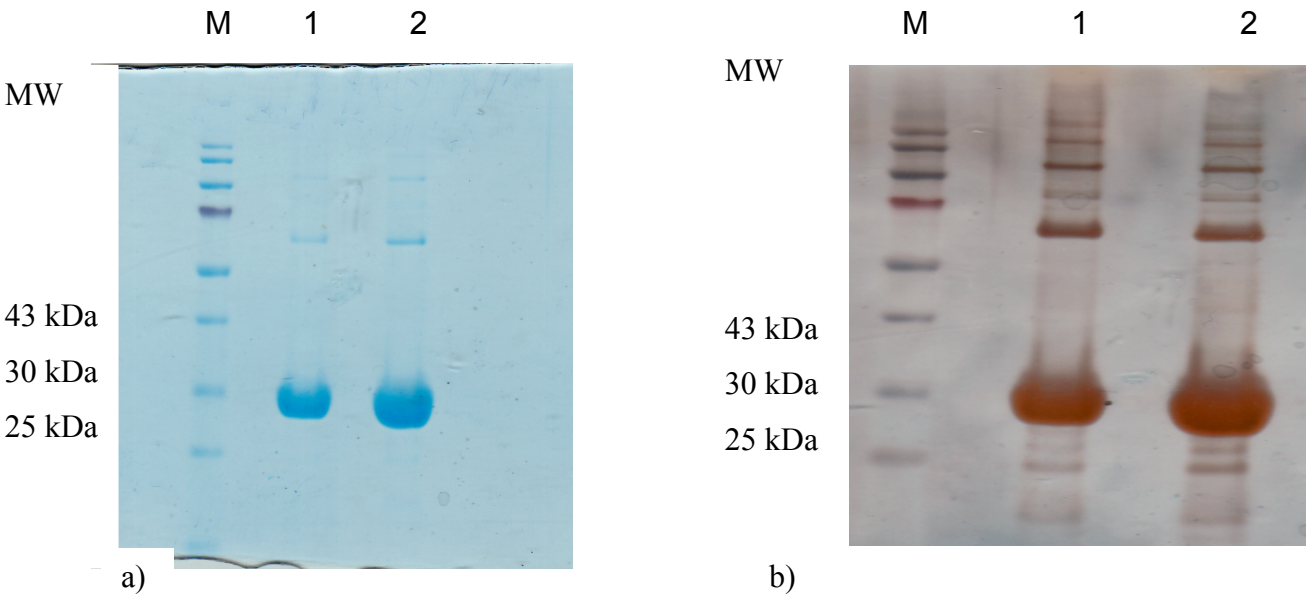


Figure 3.5.1 SDS-PAGE analysis of SynADC purity after GF purification.
b) Silver stained SDS-PAGE.

(M) = Marker (1) = Peak 2 SynADC samples (2) = Peak 3 SynADC samples

3.3 Discussion

During early purification of recombinant SynADC, a number of difficulties were encountered during purification and concentration. During the first attempts, the lysising buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% glycerol, 0.5 mM THP and 0.5 mg/ml of lysozyme enzyme for cell lysis), the approach caused numerous problems. The lysozyme enzyme was interfering with the final purity of SynADC protein after Ni-affinity columns; the SDS-PAGE gel prepared from these samples was smudged. Withdraw of lysozyme enzyme from the purification protocol solved the problem and sonication at 80% amplitude (6 x 10 sec) with 10 seconds cooling on ice was enough to open up cells. There was a need to change salt concentration within the buffers as well as the buffer pH to achieve the best elution from the column (Figure 3.5) this was because the protein was in good oligomeric state with perfect conditions in place therefore, the protein was happy to bind onto the column very well.

The three distinct peaks obtained after GF separation; a small shoulder of a void volume V_o= 46.82, peak 2 - the major peak and small aggregation peak 3.

Gel filtration chromatography was used as a final step of the purification and as a tool for the estimation of the oligomeric state of the protein. Purified SynADC decarbonylase samples were achieved through three steps: sonication at 80% amplitude, HisTrapFF, 2x1 column, and gel filtration chromatography on Superdex 200. During the first gel filtration run the protein eluted at 61.69 ml, which corresponds to a MW of ~ 26.1 kDa, the molecular weight corresponding to an octamer.

There are a number of factors thought to influence the final resolution (the degree of separation between peaks of a gel filtration separation): sample volume, the ratio of sample volume to column volume, column dimensions, particle sizes, pore size of the particles, flow rate, and viscosity of sample and buffer. The viscosity of the GF buffer used in the first protein purification protocols had a major influence to the outcome of the GF separation (Figure 3.3). Initially, 10% glycerol was included to the buffer this was to stabilize the enzyme but later on it was discovered that the enzyme was stable without glycerol therefore, stop using glycerol in the buffer preparation procedures a drastic improvement in the final resolution (Figures 3.3 and Figure 3.4) was achieved. This is because glycerol was maybe stopping the protein from binding freely onto the column by increasing the viscosity of the protein particles.

Change of buffers his involved switch between salts used in GF buffers (from 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% glycerol, to 25 mM HEPES – NaCl pH 7.8, 5% $(NH_3)_2 SO_4$ w/v) as shown in figures 3.3 and Figure 3.5 above improved the SynADC protein elution profile. The best conditions was the, 25 mM HEPES – NaCl pH 7.8, 5% $(NH_3)_2 SO_4$ w/v, 0.5 mM THP, the buffer was found not to be far off the conditions from which the initial crystals occurred (figure 6.2).

CHAPTER 4:

SCREENING OF DIFFERENT VECTOR CONSTRUCTS

1.0 Introduction

The protein crystallisation results (section 6.3.2) confirmed that, the SynADC protein can only be crystallised without a His tag. Dr. Christoph Edner at the University of Exeter previously cloned the gene that encodes SynADC protein into pET160. The vector had a His tag and a big lumio^{TM} tag (Met, Asn, Lys, Val, His, His, His, His, His, His and Met) at the N-terminal of the protein (Appendix 1). There was a need for a new construct with easily cleavage His-tag.

4.1 **SynADC – Tag into pET22b**

The cloning of the SynADC protein into pET22b was carried out by Dr. Sabine Middelhaufe in the laboratory of Prof. Nicholas Smirnoff at Exeter University. I carried out the expression studies of the SynADC-Tag protein in *E. coli*.

4.2.1 Materials and Methods

4.2.1 Expression of SynADC- Tag

An overnight culture of *E.coli* (BL21) harboring pET22b SynADC-Tag was diluted 1:100 into 5X 200 ml LB broth supplemented with 100 µg/ml ampicillin and grown at 37ºC to OD_{600} value of approximately 0.6 – 0.8.

To optimize the amount of IPTG required for over-expression of the SynADC protein in *E. coli* a screen was carried out by addition of 500µl to flask 1, 250 µl to flask 2, 50 µl to flask 3, 10 µl to flask 4, and 0 µl to flask 5 and these were cultured for 3 h at 37ºC in a shaking incubator. The *E. coli* cell cultures were harvested; first chilled on ice for 20 minutes then centrifigated at 9,000g for 15 min at 4ºC. Following centrifugation, the cell pellets were washed once with 50 mM Tris-HCl pH 7.5, then

transferred into Falcon tubes and frozen in liquid nitrogen before being stored in a -80ºC freezer.

The cell pellets were dissolved in 70 µl of lysis buffer A by pipeting the solution up and down until there were no clumps left, then samples were left to sonicate for 30 minutes in the water bath of a Bransonic ultrasonic cleaner machine.

4.2.2 SDS Gel samples

10 µl lysed extracts
4 µl 4 x sample buffer (Invitrogen NuPAGE)
1.6 µl reducing agent
0.4 µl water

The electrophoresis tank was filled with 1x MES running buffer (Invitrogen). Samples were prepared as described (section 3.1.8), 10 µl of lysed extracts were used for the gel. One well was reserved for 10 µl of molecular weight marker (MW). The lid to the electrophoresis tank was replaced and a constant voltage of 200v was applied for approximately 45 minutes

4.2.3 Ammonium sulfate precipitation of SynADC -Tag

This was carried out in a step-by- step precipitation manner, and then the supernatant was brought to 80% saturation with ammonium sulfate and centrifuged as before. The pellet was resuspended in a minimum volume of Buffer A (50 mM HEPES pH 7.8) and dialysed overnight against 1L of the same Buffer A prior to loading onto DEAE Sephadex.

4.2 SynADC + small Tag in Cold-Shock Expression Vector pCold™ II DNA

Low expression levels and low solubility often hamper the production of recombinant protein in *E. coli*. A variety of methodologies have been developed including protein production at low temperature, and fusion protein expression using soluble protein tags. Cold-shock expression vectors, pCold™ DNA, are designed to perform

efficient protein expression utilizing promoter derived from *cspA* gene, which is one of the cold-shock genes. At the downstream of the *cspA* promoter, *lac* operator is inserted so that the expression is strictly controlled. In addition, a 5′ untranslated region (5′ UTR), translation enhancing element (TEE), His-Tag sequence, Factor Xa cleavage site, and multicloning site (MCS) are located downstream of the *cspA* promoter. As this product utilizes the promoter derived from *E. coli*, most *E. coli* strains can be utilized as an expression host. There are four kinds of pCold™ vectors, whose arrangements vary in the existence of the TEE, His-Tag sequence and Factor Xa clevage site. pCold II contains TEE and His-Tag sequence; the vector has features, which makes it use friendly.

The pCold vector can be cleaved at a single site by restriction enzymes *Nde* I, *Sac* I, *Kpn* I, *Xho* I, *BamH* I, *Eco*R I, *Hin*d III, *Sal* I, *Pst* I and *Xba* I.

4.2.1 Cloning SynADC into pCOLD VECTOR and expression in *E.coli*

Dr. Christoph Edner at University of Exeter carried out the cloning of SynADC gene into the pCold vector. The SynADC pCold trial expressions were performed by inoculation of 100 ml LB medium supplemented with ampicillin (100 µg/ ml) with 2 ml of overnight culture and incubation at 37ºC with shaking until the OD_{600} reached ~ 0.4 - 0.5. Then, a cold shock was applied to the culture by chilling to 15ºC for 30 min in a water-bath with ice and expression of SynADC induced with 1 mM IPTG (final concentration). Cultures were left at 15ºC for another 24 h with shaking before harvesting. Samples were taken for SDS-gel electrophoresis before and after induction. Before induction: OD_{600} approximately 0.4 – 0.5, used about 800µl after induction; measured OD_{600} of the culture diluted with LB ~ 1:5 dilutions. Cultures were centrifuged for 10 min, at 9000 rpm; the pellet was dissolved in 50 µl lysis buffer, then sonicated for 30 min and used 10 µl used for gel analysis.

4.2.2 Screening of IPTG concentrations for optimum induction

6 X 600 ml of LB medium was inoculated with 100µl ampicillin each plus 12 ml of over-night culture, incubated at 37ºC until the OD_{600} was approximately 0.4 – 0.5.

The cell cultures were then cooled to 15ºC in a water bath for 30 minutes. Optimum induction was achieved using a series of IPTG concentrations (0, 20 µl, 50 µl, 100 µl, 250 µl and 500 µl). 100 ml of cell cultures from each flask were harvested by centrifugation at 4000 rpm for 20 minutes, and then 10 µl samples used for SDS-PAGE analysis.

4.2.3 Purification of SynADC protein from pCOLD Vector

15 g (wet weight) of cells were resuspended in 120 ml of buffer A, and lysed by sonication at maximum power using 2s pulses separated by 8s to prevent overheating for a total time of 30 min. The supernatant was separated from the cell pellet by centrifugation at 15000g at 4ºC for 30 min, and loaded onto the affinity chromatography as in section (3.1.10) above**,** using a Ni-affinity column and taking the advantage of expressed N-terminal His-tag. The fractions from the column were tested for protein purity by SDS-PAGE.

4.2.4 Quantification of SynADC concentration and activity

Bradford assays method was used for the quantification of SynADC concentration as described in (section 3.1.10.5).

SynADC activity was assessed both qualitatively and quantitatively. From Schirmer *et al*., 2010 to test the aldehyde decarbonylase (ADC) for activity, the following assays were set-up: 200µl reactions containing 100 mM sodium phosphate buffer at pH 7.2 with the following components at their respective final concentrations: 30 µM of purified ADC, 200 µM substrate (octaecanal, C_{12}, C_{13}, and C_{18}), 50 µg/ml spinach ferredoxin (Sigma), 0.05 units /ml spinach ferredoxin reductase (Sigma), and 1 mM NADPH (Sigma). Negative controls included the above reaction without ADC, without substrate, or without spinach ferredoxin, ferredoxin reductase and NADPH. Each reaction was incubated at 37ºC for 2 hours before being extracted with 100µl ethyl acetate. All samples were analyzed for enzyme activity by GC/MS.

Chemicals

- 10 µl of corresponding aldehyde i.e. C_{10}, C_{12}, C_{13}, and C_{18}, (4 mM hexane, dry under N_2)
- 100 µl SynADC (1.6 mg/ ml)
- 3.5 µl Ferredoxin (Sigma: F5875, from spinach) 10 µg
- 5 µl Fed/NADP reductase (Sigma: F0628, from spinach) 10mU
- 10 µl 20 mM NADPH
- 80 µl 100 mM Na-phosphate, pH 7.5
- 2 µl 10% Triton X-100

Method

Solution mixtures were left to incubate for 2hrs at 37ºC, in an intermittent shaker. The entire reaction mixtures were transferred into a 1.5 ml Eppendorf tubes and 100 µl ethyl acetate was added using a glass syringe. The solution mixture was vortexed for 3 minutes, then spin for 3 minutes at 13000 rpm and then the top organic phase was carefully transferred to auto-sampler vials with inserts then GC/MS was used to measure the activity.

4.3 Results and Discussion

Cloning of SynADC into pCold vector and over-expression in *E. coli* was successful. Optimum induction was achieved by the addition of 1 ml of 100 mM IPTG to 1L of cell cultures (Table 4).

Protein purification of SynADC –Tag (without a His Tag attached) was difficult to purify and hard to collect enough protein for crystallisation even after ammonium sulfate fractionation (20%, 30%, and 40%, to 80% precipitation) was used. Most of the protein was being lost when samples had to go through a large number of columns.

Final IPTG concentrations	OD_{600} after 24 hr
0	0
20	0.58
50	0.54
100	0.51
250	0.53
500	0.82

Table 4.1 Screening of IPTG concentrations for optimum induction

Graphs to show the conversation of aldehydes to (n)- alkanes by cyanobacterial decarbonylase (SynADC) to confirm SynADC enzyme activity potential.

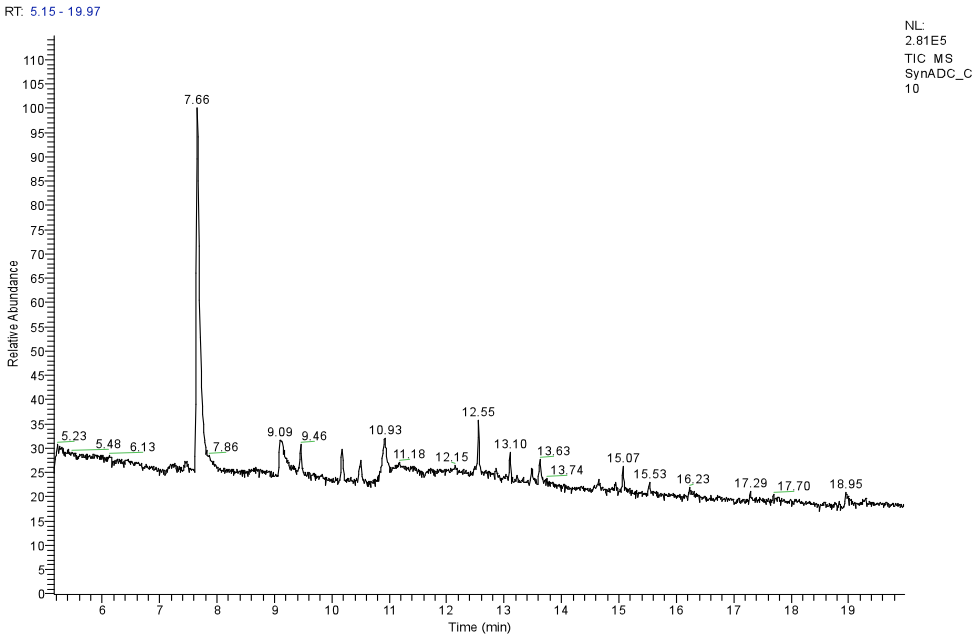


Figure 4.1 GC-MS data for the alkane production by cyanobacterial decarbonylase (SynADC) to demonstrate the metabolism of aldehyde to alkane by recombinant SynADC enzyme. 10 μl C_{10} aldehyde were used in the experiment substrate however; there was no activity was found.

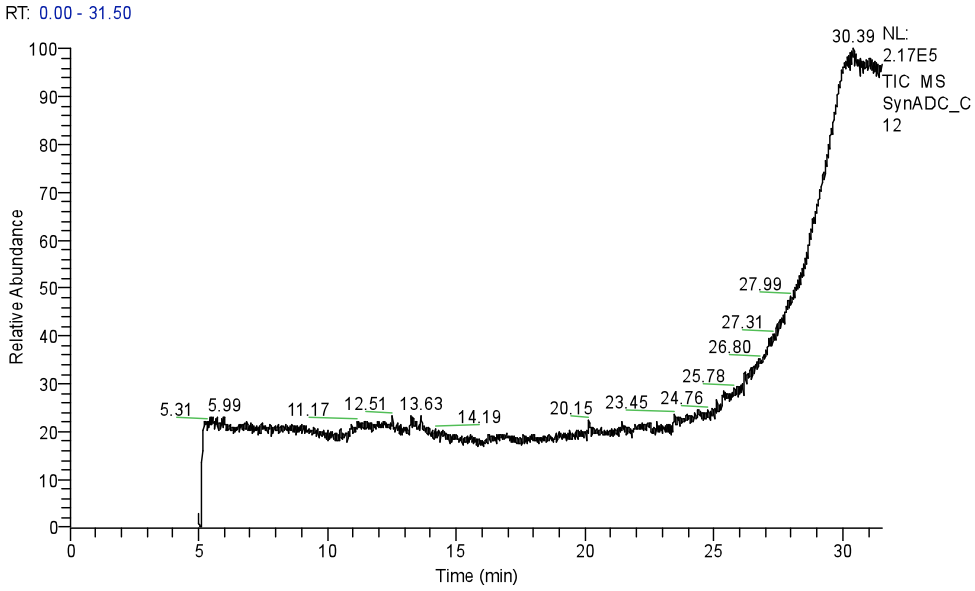


Figure 4.2 GC-MS data for the alkane production by cyanobacterial decarbonylase (SynADC) to demonstrate the metabolism of aldehyde to alkane by recombinant SynADC enzyme. 10 μl C_{12} aldehyde was used as substrate however; there was no activity was found.

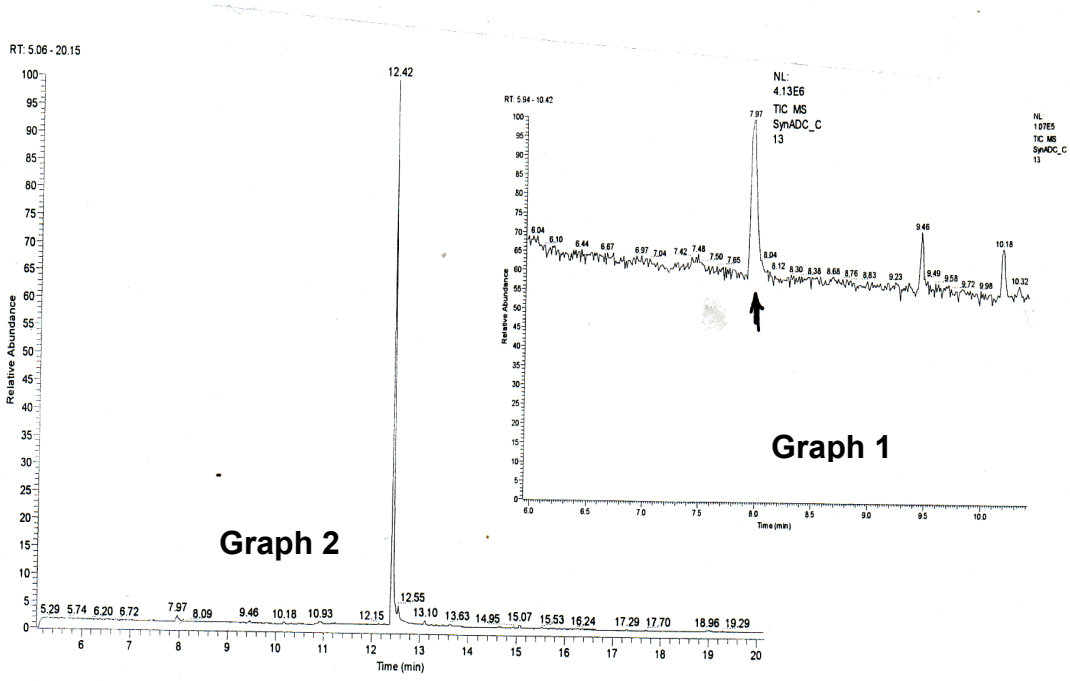


Figure 4.3 GC-MS data show the activity potentials of cyanobacterial decarbonylase (SynADC), converting 10 μl C_{13} aldehyde substrate to C_{12} alkane (graph 2). Graph 1 shows negative control samples results.

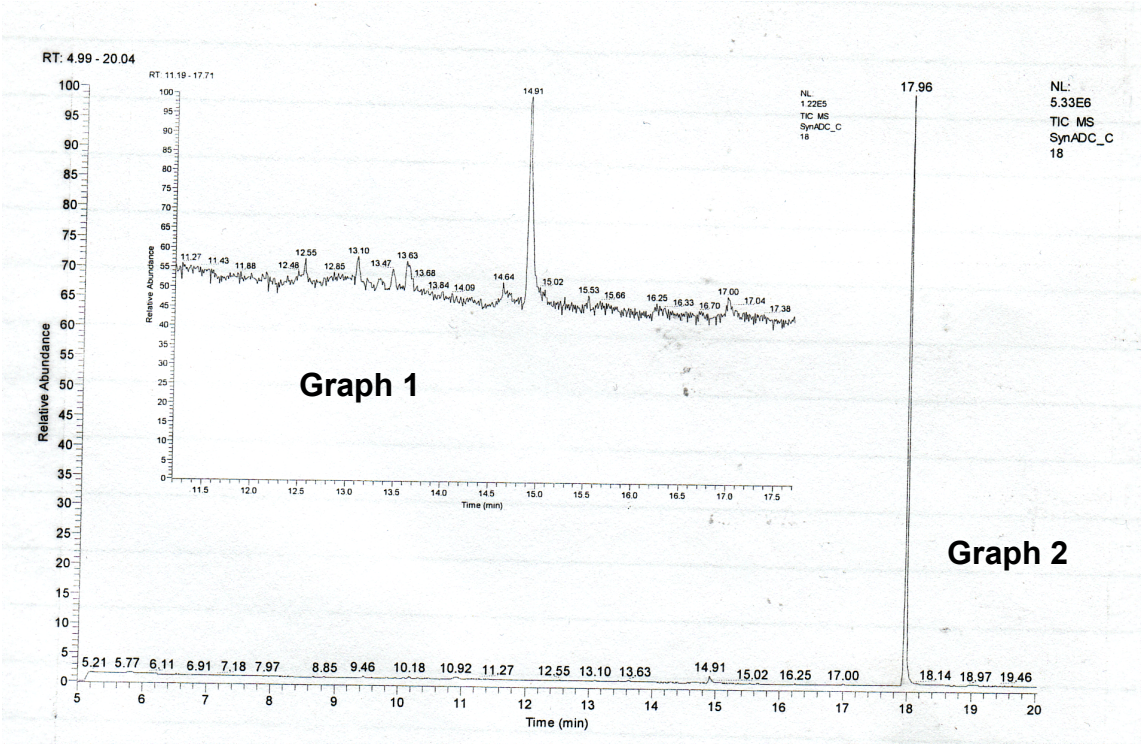


Figure 4.4 GC-MS data show the activity potentials of cyanobacterial decarbonylase (SynADC), converted 10 μl C_{18} aldehyde substrate to C_{17} alkane (graph 2). Graph 1 shows negative control samples results.

Discussion

To evaluate in vivo alkane formation using SynADC enzyme, the purified aldehyde decarbonylase was used in vivo activity with various aldehydes as substrates in the presence or absence of different cofactor as described in section 4.2.4 above. The following graphs were obtained (Figures 4.1 to 4.4). A number of nonheme diiron enzymes require ferredoxin, ferredoxin reductase, and reducing equivalents for activity (Das *et al*., 2011). The commercially available ferredoxin from spinach and ferredoxin reductase and NADPH were used.

In vivo decarbonylation of aldehyde to alkane was only observed in the presence of ferredoxin, ferredoxin reductase and NADPH, and omitting any one of these cofactors completely stopped activity.

SynADC enzyme prepared using pCold vector converted C_{13} aldehyde to C_{12} alkane and C_{18} aldehyde to C_{17} (Figures 4.3 and 4.4) respectively alkane however, using the same enzyme preparations samples as before towards C_{10} and C_{12} aldehyde samples (Figures 4.1 and 4.2) no activity was detected. This maybe SynADC enzyme only acts only upon long chain aldehydes but not on short chains or on shorter hydrocarbons more SynADC enzyme is required for activity! However further more vigorous experiments still needs to be curried out to prove the hypothesis.

Synechocystis l decarbonylase (SynADC) enzyme prepared from pCold vector over-expressed in *E.coli*, induced with 1 ml IPTG (100 mM), formed C_{12} alkane from C_{13} aldehyde (substrate). C_{17} alkane was formed C_{18} aldehyde substrate however; there was no activity notice when C_{13} and C_{18} substrates were substituted for shorter fatty acids (C_{10} and C_{12}).

Given the structural similarity of the cyanobacterial ADs to the other oxygenases and the example of the house fly AD which are known to work best on longer fatty acids chains than on shorter chains (Reed *et al*., 1995), it is logical why no activity obtain when shorter fatty acids were used in the activity assays (Figures 4.1 and 4.2) but enzyme activity was found when longer chain hydrocarbons were used instead (Figures 4.3 and 4.4). However, it will be logical also to consider some other possible oxidative outcomes of these assays as positive results such as formate. Formate is a byproduct of alkane biosynthesis in cyanobacteria through the oxidative decarbonylation by a ferredoxin-dependent oxygenase (Schirmer *et al*., 2010), getting a system, which can measure the formation of formate, can give another alternative approach which might deliver better and reliable results as Das *et al*., 2011 describes it. And the amount of formate formed in the enzyme assay can be quantified by HPLC method and standard curves.

CHAPTER 5:

SPECTROSCOPY STUDIES

5.0 Introduction

Metalloproteins - Iron-sulfur proteins:

Iron-sulfur (Fe-S) proteins are proteins characterized by the presence of iron-sulfur cluster containing sulfide-linked of di-, tri-, and tetrairon centers in variable oxidation states Figure 6.1. These clusters are found in a variety of metalloproteins, such as the ferredoxins and as well as NADH dehydrogenase, hydrogenases, Coenzyme Q – cytochrome C reductase, Succinate coenzyme Q reductase and nitrogenase (Lippard, 1994). Fe-s clusters are commonly known for their major role in the oxidation –reduction reactions for the mitochondrial electron transport.

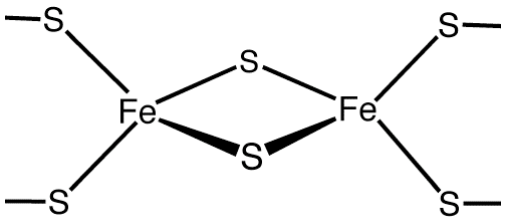


Figure 6.1 A diagram to show 2Fe-2S proteins

Since ADs are suggested to be members of the ferritin-like or ribonucleotide reductase-like family of nonheme diiron enzyme (Stubbe and Gelasco, 1998).

Non-heme iron enzymes promote a number of important biological reactions, including serotonin, leukotreine and DNA synthesis (Kovacs, 2004). Non-heme iron enzymes are of two forms; mononuclear non-heme iron enzymes contain iron ligated by oxygen and/or nitrogen ligands (Figure 6.2) or non-heme diiron (Figure 6.3), in which ADs are suggested to be a member (Stubbe and Gelasco, 1998).

Many of these enzymes promote dioxygen activation, resulting in the formation of highly reactive iron- peroxo (Fe^{III}-OOH, Fe^{III}-O_2-) (Karlsson *et al*., 2003) or iron-oxo ($Fe^{IV=O}$ or $Fe^{V}=^{O}$) (Rohde *et al*., 2003) oxidation catalysts. The flexible coordination

environment of enzymes containing 2-His-1-carboxylate (N_2O)-ligated iron leaves room for substrate as well as dioxygen activation to occur at the metal site.

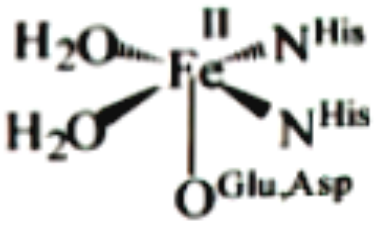


Figure 6.2. Active sites of 2-His-1-carboxylate (NO)-ligated non-heme iron enzymes

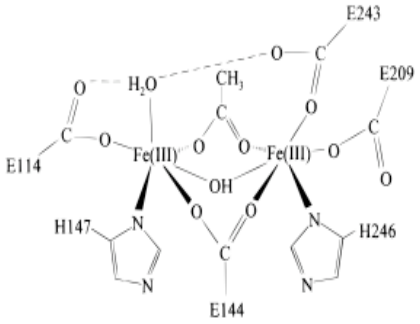


Figure 6.3. Structural representation of the binuclear iron center of diferric Methane Monooxygenase

5.1 Materials and Methods

5.1.1 Spectrophotometer scanning

A scan of the protein sample at 5 mg / ml was carried out using spectrophotometer equilibrated at wavelength parameters from 300 – 700 nm at medium speed. The glass cuvette was filled with buffer D, the machine was auto zeroed. When this was completed, the cuvette was filled with 300 μl of SynADC sample to measure the absorbance in the visible region.

5.1.2 Pull down assays to determine other proteins, which interact with *Synechocystis* AD

This work was performed to test the hypothesis that SynADC interacts with other protein. Pull down experiments was carried out as described in Schemes 1 - 3 below.

Spin down 3x30 ml culture in three 50 ml Falcon tubes (20 min, 4000g at RT)

Carefully pipette supernatant off without disturbing the pellet

Wash pellets with 5 ml binding buffer (each pellet in a different pH buffer)
Spin as above

Suspend pellets in 3 ml binding buffer (each pellet in different pH buffer) + 6 µl 0.5 M THP

Add 0.5 ml of 7x protease inhibitor stock (Roche) to each sample

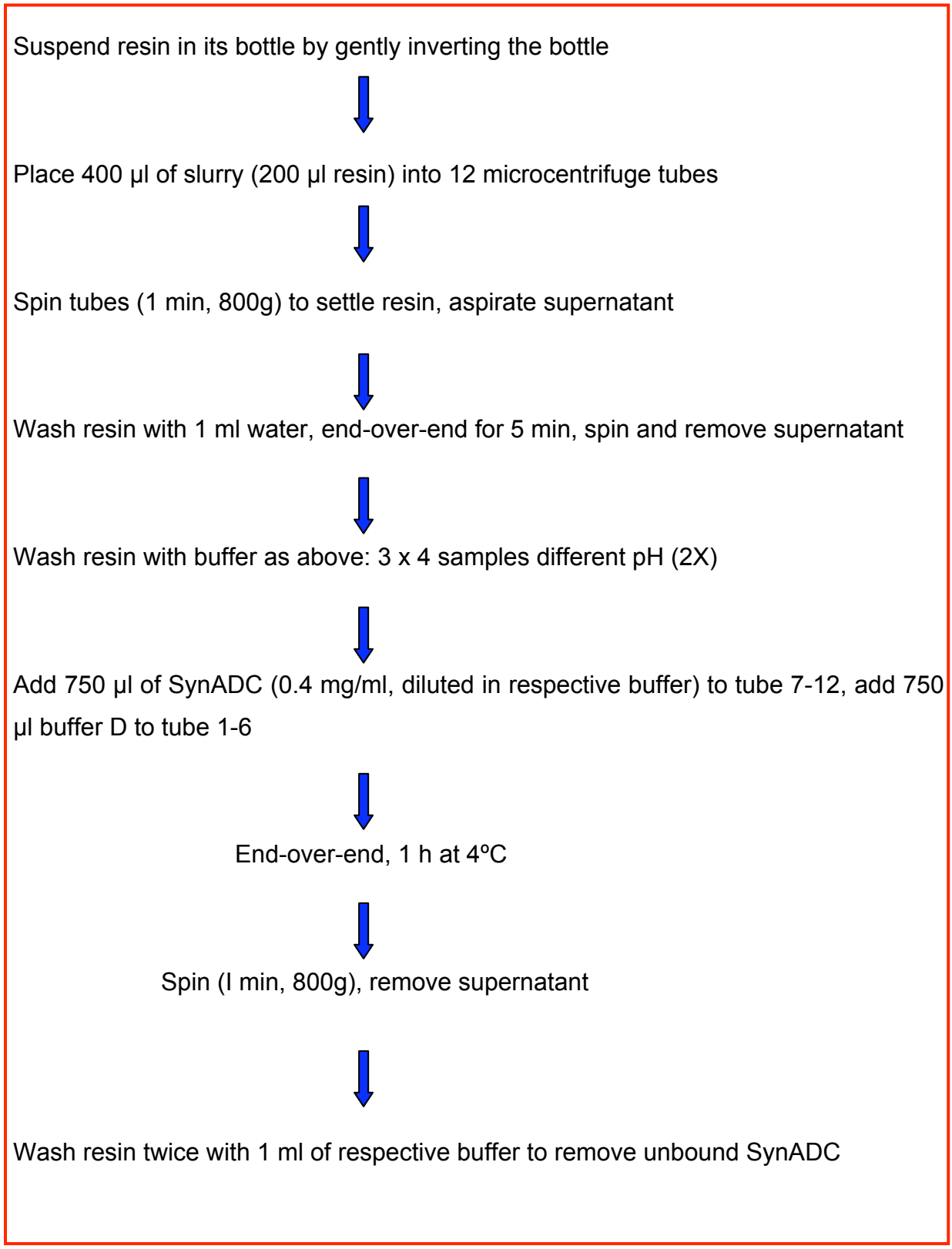
Split each sample into two 1.5 ml batches and transfer to 2 ml-bead beater tube (Lysing matrix C)

Break cells open using 45 sec blasts (max speed) with 5 min cooling of samples on ice

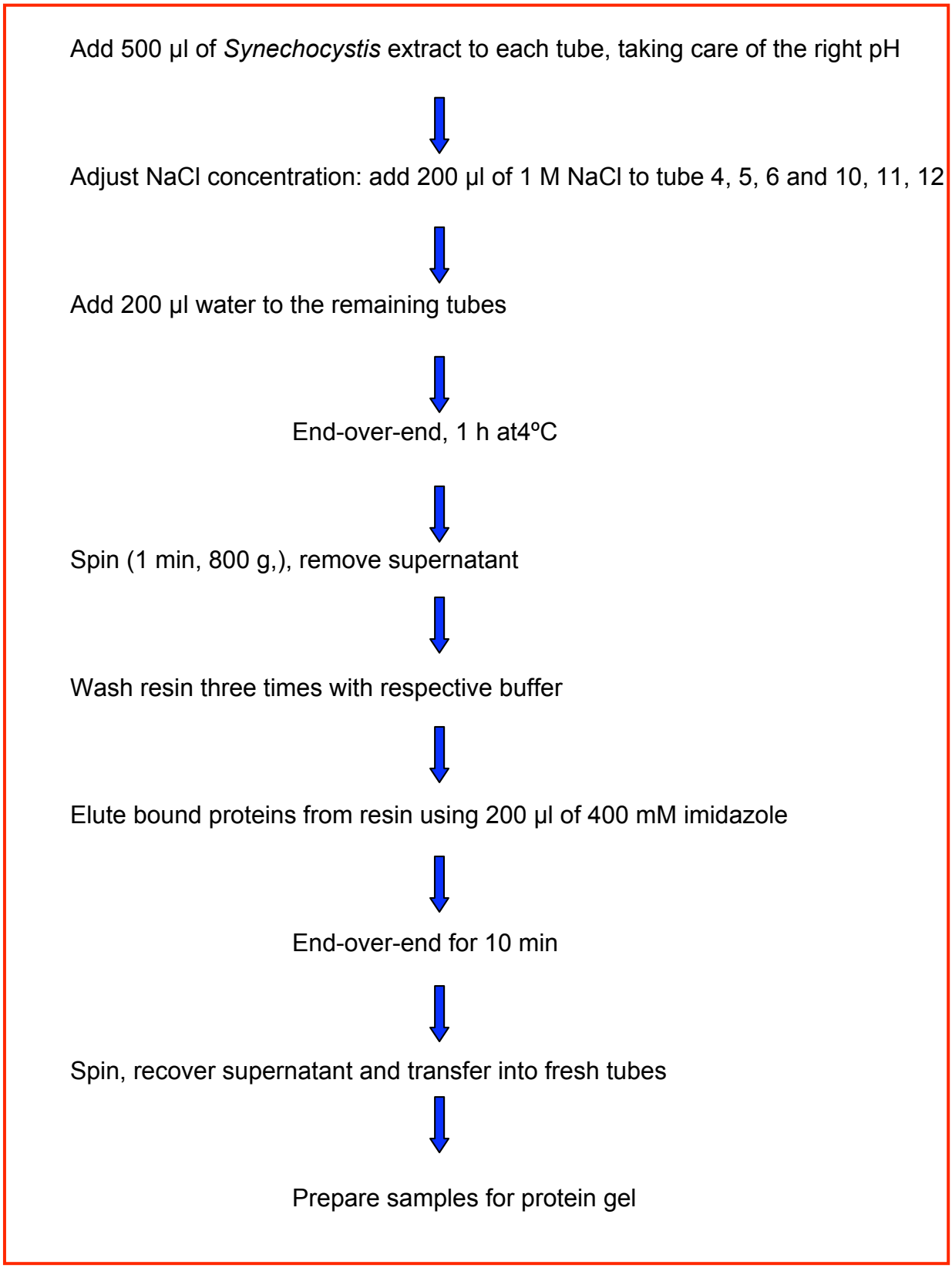
Remove cell debris and matrix by spinning the tubes at 5000g for 10 min at 4ºC

Combine supernatants of two samples with the same buffer pH into 15 ml Falcon tubes

Scheme 1. Preparation of protein extract of *Synechocystis* wild type



Scheme 2. Binding SynADC to PROBOND resin (Invitrogen)



Scheme 3. Challenge resins with *Synechocystis* extract

Sample	1	2	3	4	5	6	7	8	9	10	11	12
pH	7.0	7.5	8.0	7.0	7.5	8.0	7.0	7.5	8.0	7.0	7.5	8.0
NaCl	-	-	-	250	250	250	-	-	-	250	250	250
SynADC	-	-	-	-	-	-	+	+	+	+	+	+

Buffer: 50 mM HEPES-NaOH, pH 7.0/7.5/8.0; 1 mM THP (prepare 50 ml each)

Table 5.1 Summary of preparation extracts of *Synechocystis* AD used in the pull down assays above.

After failing to obtain any form of interaction from all the samples with and without SynADC added (Figure 6.3A & B) using mass spectrometry (on both in-gel and in solution digests) samples, the method was restructured. Instead of using 30 ml culture to start with, the volume was scaled-up to 50 ml of *Synechocystis* wild type and then followed the same preparation steps as in (Scheme 1) above. Then the *Synechocystis* extracts were challenged with 250 µl of purified recombinant SynADC (10 mg/ml) to make a positive control (+), while on the negative (-) control only added 250 µl of protein buffer for 1h at 4ºC in a rotary mixer. The SynADC was pulled down with Interactors using “PROBOND” resin as in (Scheme 3) above but with the following changes; the resins were re-suspended in its bottle by gently inverting the bottle. 2X4 ml of slurry (2X2 ml resin) was placed into two 15 ml Falcon tubes. The tubes were spun (I min, 800g) to settle resin, then washed resin with 10 ml water, rotary mixed for 5 min, spun and the supernatant discarded, followed by the addition of the entire sample from (+) and (-) to their respective tubes containing the “PROBOND” resins and then rotary mixed for 30 min at 4ºC. The reaction tubes were spun for 1 min at 800g removed supernatant and kept for protein gel analysis. The resins were washed three times with 10 ml of HEPES buffer with THP added to the buffer to remove unbound SynADC and unbound *Synechocystis* proteins. Bound proteins were eluted from the resin using 0.5 ml of 1.25 M imidazole.

Sample	(+)	–
SynADC	+	_
Cyanobacterial extracts	+	+

Summary

Buffer: 50 mM HEPES-NaOH, pH 7.8

5.2 Results and Discussion

5.2.1 Spectrophotometer scanning

The enzyme purified from *E.coli* was pale brown in color; the absorption spectrum (Figure 5.2) has a broad peak at around ~350 nm (ε350 nm ~ 5 $mM^{-1}cm^{-1}$), which is typical of non-heme di-iron enzymes, due to the presence of the oxo to Fe (III) charge transfer band (Das *et al*., 2011). ICP-MS a more sensitive technique was used to determine other metals present in the SynADC protein. The following metals (Zn, Fe, Ni, and Mn) were found to be present in the SynADC proteins, in the concentration (2.37 mg/l, 1.16 mg/l, 0.137 mg/l, and 0.032 mg/l) respectively.

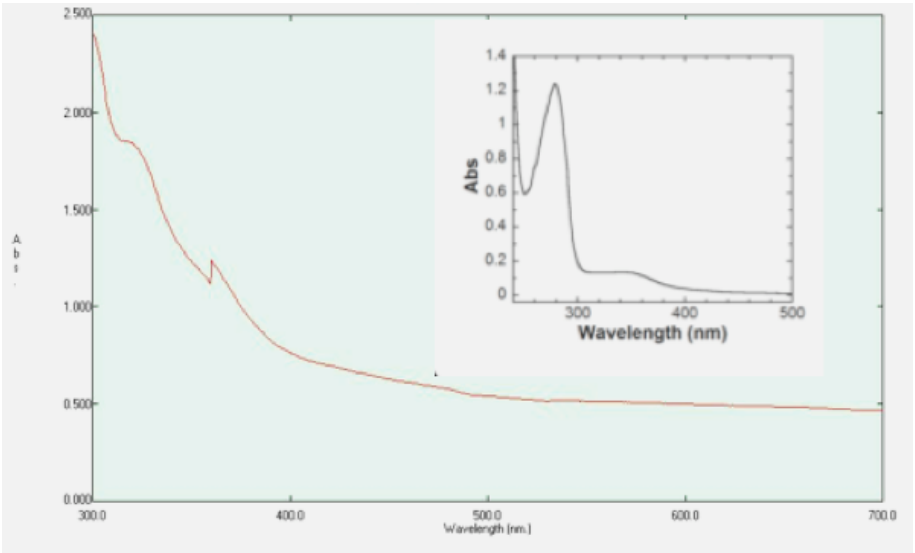


Figure 5.2 The absorption spectrum of SynADC (5 mg/ml) showing charge transfer bands at ~350 nm and at longer wavelengths; inset is the full UV-Visible absorption spectrum of SynADC

5.2.2 Detection of proteins interacting with SynADC protein

The initial approach to detect protein interaction with SynADC failed; there was no visible difference from the gel results could be seen between control samples and experimental samples (Figures 5.3A and 5.3B). However, high salt content encouraged more binding of the protein onto the resin according to lanes 10, 11, and 12 (Figure 6.3A).

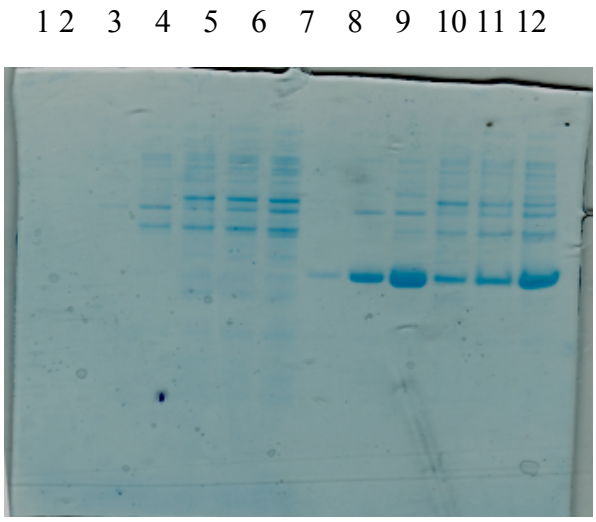


Figure 5.3A SDS-PAGE analysis of SynADC protein interactors lanes 1- 6 represent control samples and lanes 7 -12 are experimental samples (Table 5.1). SDS-PAGE samples were prepared by mixing the following 20 µl SynADC, 8 µl sample buffer, 3.20 µl reducing agent and 0.8 µl of HO_2, then heated up in heating block for 10 min.

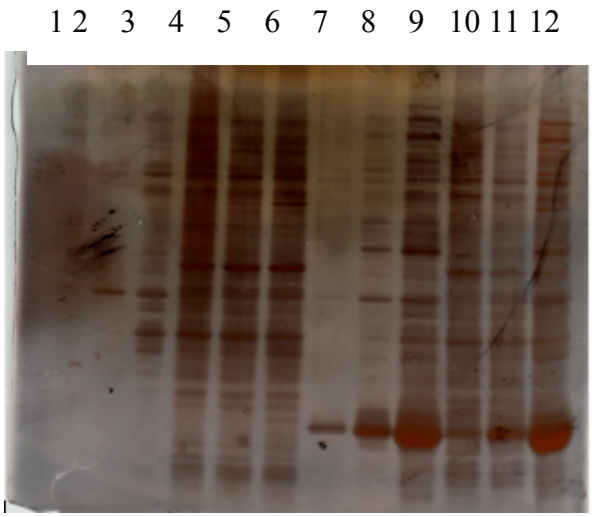


Figure 5.3B Silver stained SDS-PAGE

After modifying method used to prepare samples for experiment (Figure 5.3A), positive results were achieved for the presence of proteins interacting with SynADC confirmed by silver stained gel (Figure 5.4).

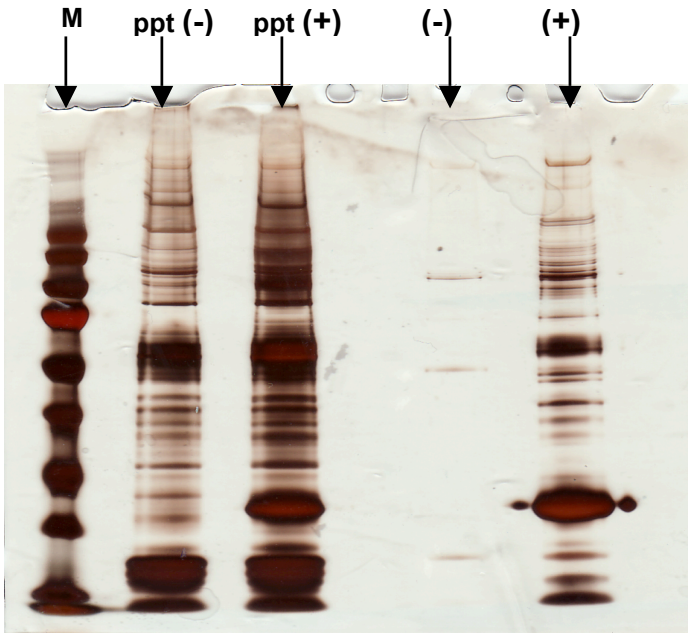


Figure 5.4. Analysis of protein interactions with SynADC decarbonylase using silver stained SDS-PAGE gel.

(+)	=	SynADC + cyanobacteria extracts added
(-)	=	Cyanobacteria extracts
ppt (+)	=	Precipitants from (SynADC + Cyano extracts) samples
ppt (-)	=	Precipitants from (Cyano extracts) samples
M	=	Molecular maker

From figure 5.4 above, we noticed that there was significant numbers of protein bands present in the positive (experimental) samples, which were not present in the negative control samples. Attempts to analyse these differences using mass spectrometry MS for both in-solution and in-get digests samples failed to work.

Discussion

ICP-MS scans for the determination of metals and their content present in the SynADC protein were in good agreement with the Schirmer *et al*. results from the ICP-MS data however, there was tiny difference in the abundance of the metals (Zn, Fe, Ni and Mn) contents. This was due to the difference in the cell growth medium used, different from Schirmer *et al*., 2010.

The protein interactors data conclusively show that SynADC protein interact with some other proteins, which are yet to be established. The interaction of SynADC with other proteins (Figure 5.4); comparison of two samples (-) and (+) lanes was a clear cut from sliver stained SDS-PAGE analysis although attempts to identify these interacting proteins with Mass spectroscopy (MS) in-solution and in-get digests samples failed to support the findings. Silver staining is more technique; in-solution and in-get digests samples for MS use requires large amounts of protein to be loaded on SDS-PAGE. TCA precipitation is a technique in which low Sample preparation for proteomic analysis involves precipitation of protein using 2,2,2-trichloroacetic acid (TCA). Charles Tanford proposed that (TCA)-induces protein precipitation by forces protein to precipitate by sequestering the protein-bound water.

This method can be applied on SDS-PAGE gel samples to boost the protein concentrations for the protein samples. And lastly, scaling-up the amount of protein loaded on the SDS gel and to a bigger nickel affinity chromatograph will be ideal.

CHAPTER 6:

PROTEIN CRYSTALLIZATION

6.1 Introduction

One aim of this research project is to solve the protein structure of SynADC (Cyanobacterial alkanal decarbonylase). The methodology has involved the expression of SynADC in *E.coli*, purification of the protein and screening of various crystallization conditions kits to come up with the perfect crystallization condition for the protein.

The knowledge of the structure of a protein can provide important details about the protein's function, mechanism of action and its interaction with other molecules. The techniques of NMR or X-ray diffraction can be used to solve the three dimensional structure of protein. In order to solve protein structures by X-ray diffraction, high quality well ordered crystals are used.

The crystallisation of protein is dependent on many inter-related factors. These include protein purity (homogeneity), solubility, aggregation and concentration of precipitant, additive, temperature, and pH while as salt concentration is some of the other external factors.

There are two phases of crystallisation, nucleation followed by growth. Each of these stages is dependent on both protein concentration and the crystallisation agent concentration. To promote either stage, supersaturation needs to occur. This is a condition where there is more protein than can be dissolved in the volume of fluid. Supersaturation leads to either a crystalline or amorphous phase when equilibrium returns. Supersaturation is induced in crystallisation by introducing factors that reduces protein solubility, these include; allowing water to evaporate, changing temperatures, or adding an ionic solute (a precipitant). Protein precipitants include salts and high molecular weight PEGs. Salts bring protein out of solution by competing for the water of hydration.

In the first nucleation phase, protein molecules need to overcome an energy barrier to form a periodically ordered aggregate of a critical size that can survive in a thermodynamic sense. Nucleation can happen throughout the solution or on a solid surface such as a dust particle or the bottom of the crystallisation well. In some case nucleation crystal growth can be induced by addition of artificial nucleants such as horse hair. When nucleation occurs, protein is removed from solution; therefore the level of supersaturation will be reduced to the metastable phase. This allows growth to occur and reduces any further nucleation events. If the crystallisation has reached equilibrium, the crystal growth will push the supersaturation further into the metastable phase. If however there is no equilibrium, the solution may remain in the nucleation phase and more nucleation events occur resulting in many small crystals.

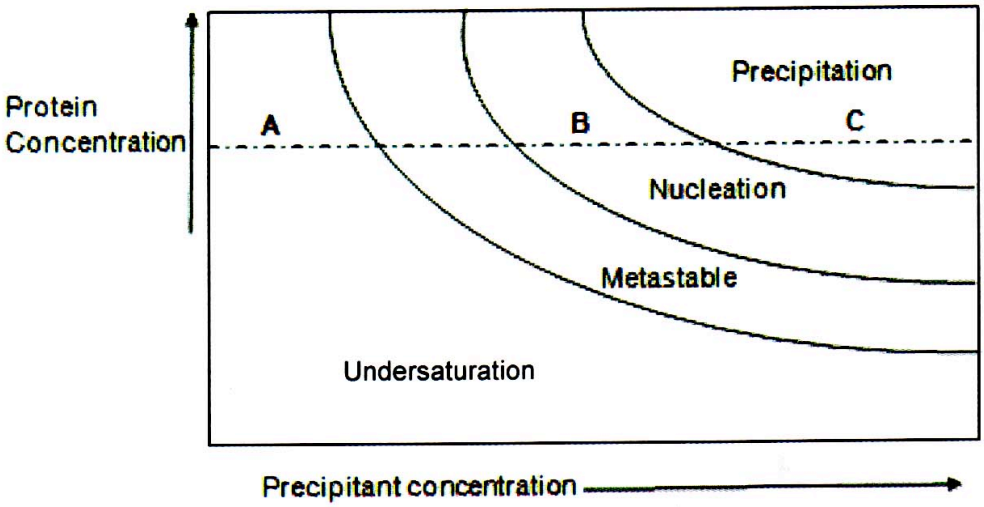


Figure 6.1 The phase diagram showing the solubility of the protein as the precipitant concentration changes. At point A the protein will stay under-saturation. At point B the protein will crystallise and the concentration of the protein in solution drops to saturation as the crystal grows. At point c the protein will precipitate out but crystals may still grow. Figure adapted from (James, 2010).

In order to determine in which condition a protein will crystallise it is common to carry out a wide range of plate screening with a variety of conditions, commonly using commercial crystallisation kits. The structure of Cyanobacterial Fatty Aldehyde

Decarbonylase enzyme has been solved from *P. marinus*, but the one from *Synechocystis* is yet to be solved. Schirmer *et al*., 2010, suggests that the cyanobacterial decarbonylases are members of the ferritin–like or ribonucleotide reductase- like family of nonheme diiron enzymes this was confirmed by the crystal structure of the enzyme from *P. marinus*. *Synechocystis* is known to have 64.9% sequence identity to the cofactor containing decarbonylase from *P. marinus,* **(**Figure 1.3). Since homologous of proteins crystallise under different conditions it was decided to carry out a new screen with the *Synechocystis* enzyme.

There are several different methods of crystallisation such as microbatch and vapour-phase diffusion. Microbatch experiments are where the protein and precipitant are mixed directly under oil. In the true microbatch method there is little to no diffusion as H_2O cannot travel through the paraffin oil placed on top of the droplet (Chayen *et al*., 1992; Brumshtein *et al*., 2008). Protein crystallisation using this method requires that the precipitants concentration is correct for crystal formation. Microbatch crystallisation using Al's oil (50/50 paraffin/silicon oil mix) shows some diffusion increasing the concentration of precipitant and protein and similar in this way to vapour diffusion.

Vapour diffusion consists of a reservoir, which contains the precipitating condition that is separated from the protein droplet mixed with a small volume of precipitant. An equilibrium state is created when water diffuse from the drop to join precipitants in the reservoir. This increase in precipitant and protein concentration within the drop can bring the protein into the super-saturation state. As protein crystals begin to grow the solution is brought into the metastable crystal growth zone.

6.1.1 X-ray crystallography

X-rays can be used to determine the 3D structure of protein. Usually, X-rays with a wavelength between 0.5 and 1.6 Å are used as these wavelengths are comparable to the inter atomic distances in protein crystals. They are also capable to penetrate and diffract, scatter strongly enough by the crystals. X-ray waves are shot, right onto the crystal to cause a diffraction pattern, which is produced according to the different

properties of the crystal. From the diffraction pattern, the three dimensional structure of the proteins is then determined.

Protein crystals are made up of regular repeating array of unit cells, which amplify X-ray diffraction patterns. These unit cells are defined by three length (a, b, and c) and these are the three angles (alpha, beta, and gamma). The more unit cells present within the crystal, the stronger the resulting diffraction. The crystal is thought of being divided into a number of planes, which runs through the unit cell, and those planes have various orientations and spacing run between them. Collections of numerous individual reflections (spots) make-up the diffraction pattern, with the spots corresponding to the diffraction from the crystal planes.

Bragg, 1913 (Bragg's law) illustrated the relationship between the diffraction pattern and the spacing of the crystal planes, which help our understanding into this mystery. Bragg's law describes the relationship between the reflection angle (θ), the distance between the planes (d) and the wavelength (λ) shown in the equation below.

$$n\lambda = 2d\sin\theta$$

Where: n is an integer, λ is the wavelength of the radiation, d is the spacing between the lattice planes and θ is the angle of incidence of the X-ray beam.

From Bragg's law, all the planes that run through the crystal are treated as mirrors that can reflect the X-rays therefore these reflections give rise to the reflection spots. X-rays might constructively add up if rays start or arrives in phase at the crystal. Constructive diffraction will occur according to Bragg's law within a crystal with two planes separated by distance (d) if the difference in path length ($2d\sin\theta$) between the waves is equal to the integer number of wavelength (Figure 6.2).

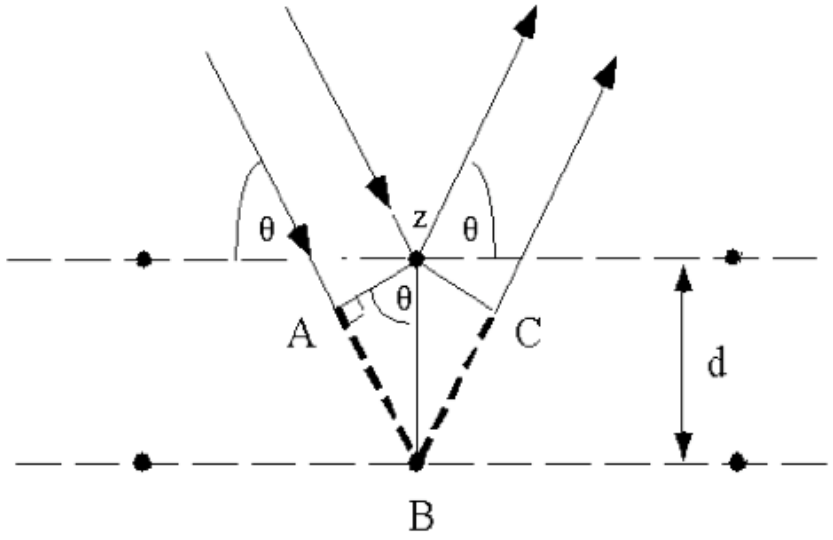


Figure 6.2 Conditions that satisfy Bragg's law. Figure adapted from Paul (2010).

From the figure 6.2 above, we can conclude that the shorter the spacing between the crystals planes then the greater the angle of diffraction. The various lattice planes that divide the crystal across its three dimension (a, b and c) are described by a set of Miller incidences (h, k and l). A value of the reflection angle (θ) can be calculated using simple trigonometry and from this the spacing (d_{hkl}) can be calculated. Thereafter the dimensions of the unit cell are able to be calculated.

During data collection, it is important to rotate the crystals to obtain a diffraction pattern because for a stationary crystal only a limited number of the crystal planes will satisfy Bragg's law. Rotating the crystal allows data to be collected from all the crystal planes. This is better explained by the Ewald's sphere, which describes Bragg's law in three dimensions. In the Ewald's sphere, the crystal is placed at the centre and with a radius of $1/\lambda$. Each reciprocal lattice point has a position relative to its family of crystal planes, which are located at a vector from the crystal with a length that is inversely proportional to the d spacing. A diffraction spot is produced if the lattice point lies on the surface of the sphere (Figure 6.3). Rotating the crystal to a reciprocal lattice points, this changes the points that lie on the sphere's surface and this produces all the required reflections.

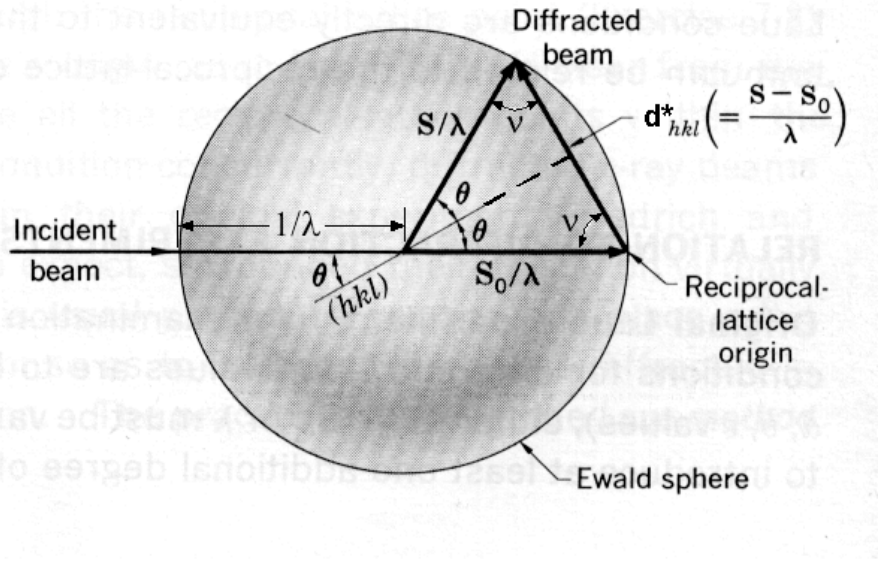


Figure 6.3 Ewald's Sphere
In the Bragg construction both the incident beam and the scattered beam are inclined to the (hkl*)* plane at an angle q making the angles of incidence and "reflection" equal. Figure adapted from Dauter (1999)

6.2 Materials and Methods

6.2.1 Expression of SynADC using a pCold vector

The gene that encodes for the recombinant SynADC protein was transformed into BL21 (Novagen) cell line as described in 3.1.3. Induction studies were conducted as described in 4.1.1. The cells were then harvested by centrifugation (9000 X g, 20 mins, 4ºC) using a Beckman JLA 10.500 rotor. The cell pellets were stored at -80ºC until further use.

6.3 Protein purification of SynADC (pCold SynADC)

6.3.1 Cell lysis

The cell paste was re-suspended in 50 ml of 20 mM Na-phosphate pH 7.5, 500 mM NaCl, 20 mM imidazole and 0.5 mM PMSF). The cells were disrupted by sonication as described in section 4.1.1.

6.3.2 Nickel affinity and gel filtration chromatography

The nickel affinity chromatograph was performed as in method 3.1.10.4. The fractions eluted from the column were analyzed using SDS-PAGE. Size exclusion chromatography was carried out as in method 3.1.10.5 with fractions collected over one column volume. Fractions containing protein were analyzed using SDS-PAGE.

6.4 SynADC cleavage of N-terminal His-tag using AcTEV protease (supplier)

Earliest SynADC protein samples (cloned into pET160 vector) required the His-Tag to be cleaved off before the protein could easily crystallise. The N-terminal His-tag was removed from the protein by carrying out a protease digest reaction on the crude samples eluted from the nickel affinity chromatograph (elution D). The reaction was set at 4ºC in the cold room, using 350 µl of 20.26 mg/ml SynADC, 750 µl 20x TEV buffer, 150 µl 0.1 M DTT, 100 µl ACTEV (10 units), and 119.3 µl water. After overnight incubation, the samples were loaded onto the HisTrapFF column for purification as described in section 3.1.10.4 above. The flow through and wash fractions were collected and loaded onto Size exclusion chromatography as described in method 3.1.10.5. SynADC protein fractions were collected and concentrated to 10 mg/ml for crystallisation trial experiments.

6.4.1 Sample preparation for His-tag cleavage

GF elution fractions samples (peak 2) were pooled and concentrated to 10 mg/ml concentrations for crystallisation. This was achieved by concentrating the SynADC decarbonylase protein in a 15 ml concentrator with a 10 kDa cut-off regenerated cellulose membrane (Millipore). The protein was concentrated in either buffers; 1, 2, or 3 (see table 6.1) before crystallisation trials.

Buffer			pH	Glycerol and THP concentration
A	50 mM Tris-HCl	100 mM NaCl	7.5	10% glycerol, 0.5 mM THP
B	50 mM Tris-HCl	100 mM NaCl	7.5	0.5 mM THP
C	25 mM HEPES-NaOH	5% $(NH_3)_2 SO_4$ w/v	7.8	0.5 mM THP

Table 6.1 Showing the different GF buffers used.

6.4.2 Protein concentration determination

Protein estimation during the concentration was determined by preparing triplicates Bradford Assays i.e. a mixture of 200 µl of Bradford, 790 µl of water, and 10 µl of the protein sample and left to stand for 10 min at room temperature, and then measured the OD_{595}.

6.5 Crystallisation of SynADC

6.5.1 Initial crystal trials (using microbatch method)

Microbatch crystallisation trials were set up in a Hampton 96 well plate using an Oryx 6 crystallisation robot (Douglas Instrument, UK). SynADC peak 2 elution (the major peak samples) were concentrated to 10 mg/ml in 50 mM Tris-HCl, pH7.5, 100 mM NaCl, 10% glycerol and 0.5 mM THP. Trials were set up, screening a wide range of crystallisation kit conditions: JCSG, Sigma, pH, Clear, MDL 1 and 2 (Molecular Dimension Laboratories) (Appendixes 9.8 – 9.11). The robot mixed 0.5 µl screen, 0.5 µl of protein and covered it with a drop of oil at the top. Crystallisation trays were left to incubate in a still 19ºC incubator, checked under a microscope for any signs of crystal formation on a weekly basis. Crystallisation conditions that showed successful crystal hits were repeated with optimized conditions.

6.5.2 Using vapour diffusion method

Each well comprised 2 μl protein solution mixed with 2 μl reservoir solution in a sitting drop, over 1 ml reservoir solution and incubated in a still incubator at 19 ºC.

6. 6 Preparation of apo-SynADC (stripping off metals)

Endogenously bound transition metals were removed from SynADC by incubating the protein at 4ºC overnight in 100 mM HEPES (pH7.2) containing 0.1 M KCl, and 10% glycerol, to which ferrozine (10 mM) and sodium dithionite (5 mM) were added followed by desalting to remove ferrozine and dithionite on a column of Sephadex G 25 fine resin. The protein from this step was then dialyzed against the same HEPES buffer containing the metal chelators EDTA (10 mM) and NTA (10 mM) at 4ºC overnight. Finally, the protein was dialyzed at 4ºC against several changes of the HEPES buffer without metal chelators before being purified on the GF 200 column

6.7 Crystallisation Optimization

Crystallization is often divided into two steps: in the *screening* step, solutions that have previously given crystals are used to find new conditions where a protein crystallizes. Then, in the *optimization* step, the resulting crystals are improved by making small changes to the crystallization conditions identified in the screening step.

Different robots and crystallization approaches are used for these two steps. The screening step is often automated, while the optimization step is slow, and it is frequently carried out by hand.

Microseeding has been used routinely for optimization by a minority of crystallizers for many years (Bergfors, 2003). However this is painstaking work because you need to identify conditions where seeding is likely to work. A new approach allows microseeding to pick up entirely new conditions - conditions where crystals would not

form in the absence of crystal seeds. Moreover the quality of the crystals found is generally better too.

Ireton and Stoddard (2004) introduced a new approach to microseeding, which they dubbed “microseed matrix screening”. Crystal seeds were systematically added to diverse conditions as part of the optimization procedure. These conditions included ingredients that were not present in the original hits.

For example, Ireton and Stoddard identified crystallization conditions for the protein yCD that contained sodium acetate in a standard screening experiment. The resulting crystals, however, could not be used to solve the structure of the protein because the mosaicity was too high. However, by seeding into conditions where the sodium acetate in the initial hit has been replaced with calcium acetate, well-diffracting crystals were grown that allowed the structure to be solved.

However, D’Arcy et al. later changed this approach by the introduction of two practical changes, which together allowed automation: (1) seeding experiments were carried out using ordinary commercial crystallization screens. (2) Microseeding was carried out using solutions containing seeds prepared by crushing crystals using the “seed-bead” kit from Hampton Research.

6.8 Microseed Matrix Screening

Micro-crystals were extracted from the crystal droplet using a pipette. The crystals were broken down into very tiny particles by pipeting up and down. The seeds were diluted using a mother liquor (same as the condition the crystals grew from), 1 in 100 and 1 in 1000. Crystal dishes were prepared with conditions similar to the conditions where previous crystals tend to appear but with slight changes with precipitants. 1 μl of seeds were mixed with 10 μl fresh protein reservoir. Usually microcrystals are formed if the conditions are suitable for nucleation but not the right conditions for crystal growth, hence lots of microcrystals. The aim is to reduce the nucleation events and improve the growth. The seeds act as nucleation, which means the right condition for growth of crystal without having to find nucleation at the same time.

6.9 Soaking of protein crystals in metals ions (Fe^{2+} and Zn^{2+})

E.coli is a convenient host as it is well understood and easily manipulated within the laboratory environment. Many host and vector strains have been developed and in optimal conditions up to 50% of the protein produced can be the protein of interest. Although the simplicity of *E.coli* often makes it the host of choice, as more proteins are being targeted for structural studies, it is becoming apparent that *E.coli* is sometime is unable to fold some proteins in their correct oligomeric forms, sometimes metals are poorly taken up by *E.coli* during cell culture, even metals have a tendency of leaking out of the protein during purification and buffer exchange. There are many ways in which eluted or lost metals from the protein can be replaced; by metal soaking, by adding the lost metals in crystallization conditions, or by adding excess metals in the purification buffers and growth media.

Protein crystals that tested to diffract well but with low metal occupancy were soaked into 700 μl of 10 mM $FeSO_4$ and 10 mM $ZnCl_2$ ion solution for 72 hours and then frozen in liquid nitrogen for data collection.

6.10 Soaking of protein crystals with ligands

Soaking of protein crystal with ligands was carried out by dissolving the ligand in a solvent (DMSO). The mixture was added to the “mother liquor”, made out of conditions similar to the crystallisation conditions of the crystals also containing a cryo-protectant. The protein crystals were then soaked in the mixture of 700 μl of 10 mM Fe^{2+} and 5 mM ligand first for 24 hours then for 72 hours depending on the physical condition of the protein crystal under the cryoprotectant solution. Protein crystals were then frozen in liquid nitrogen ready for data collection.

6.11 Co-crystallization with ligands

Before protein samples were co-crystallised with ligands, the ligands were dissolved in ‘mother liquor’ solutions either of salt or PEG then, to 50 µl SynADC (10 mg/ml) was mixed with 2.5 µl of 5 mM of various ligand (C_4 aldehyde, Hexanoic acid, C_8

aldehyde and valeric acid) and incubated on ice for 2 hours. Crystal dishes in 96 well plates were prepared for the various crystallisation condition kits and the trays were incubated at 19ºC and checked under a microscope for sign of crystallisation at weekly intervals.

6.12 Preparing crystals for data collection

Crystals were frozen using a cryo-protectant (with either salt or PEG buffer). The crystals were removed from the droplet and place in the cryo-protectant before being frozen directly in liquid nitrogen.

6.13 X-Ray data collection

Dr. Michail Isupov carried out all crystallographic data collection and structure determination. I attended a data collection session at the Diamond Synchrotron light source (Oxford, UK).

6.14 Results and Discussion

6.14.1 Expression of the SynADC protein

The SII0208 gene in the pCold vector was transformed into the *E.coli* expression strain BL21. The over-expression of SynADC protein was induced using IPTG as described in methods 4.1.1. The presence of SynADC was confirmed with a band on SDS-PAGE (29 kDa) (Figure 3.2).

6.14.2 Protein concentration determination

I was able to concentrate peak 2 fraction samples from the GF column to 10 mg/ml but peak 3 fraction samples failed; the protein was found to precipitate out of the solution as the concentration increases.

6.14.3 Co-crystallization with ligands

The enzyme was co-crystallized with ligands (C_4 aldehyde, and C_8 aldehyde). Attempts to Co-crystallise SynADC enzyme without metals, no complex were formed. Soaking protein crystals in solutions containing ligands and meters, complexes were formed. X-ray diffraction data from suitable crystals were collected (Table 6.3).

6.14.4 Protein Purification

One litre of culture produced 3.8 g of cell paste and after cell lysis (as in methods 6.3.1 and 4.1.1) and centrifugation the supernatant was purified using nickel affinity and gel filtration chromatography as described in methods 3.1.10.4, 3.1.10.5 and 6.3.2.

6.14.4.1 Nickel affinity and gel filtration chromatography of SynADC (pCold vector)

The clarified cell extracts were applied to a nickel column as described in methods 3.1.10.4 producing a band of 29 kDa on the SDS-PAGE (Figure 3.2). The C elution was pooled, concentrated and then exchanged into buffer D before being loaded onto the GF column. Gel filtration chromatograph was used as a final step of the purification process and as a tool for the estimation of the molecular weight of the protein as mentioned in chapter 3. The peak fractions peak (Figures 3.3, 3.4 and 3.5) were analyzed by SDS-PAGE (Figures 3.5.2a and 3.5.2b) and those of contained the protein of interest (SynADC) were pooled and concentrated.

6.14.4.2 SynADC cleavage of N-terminal His-tag using AcTEV protease

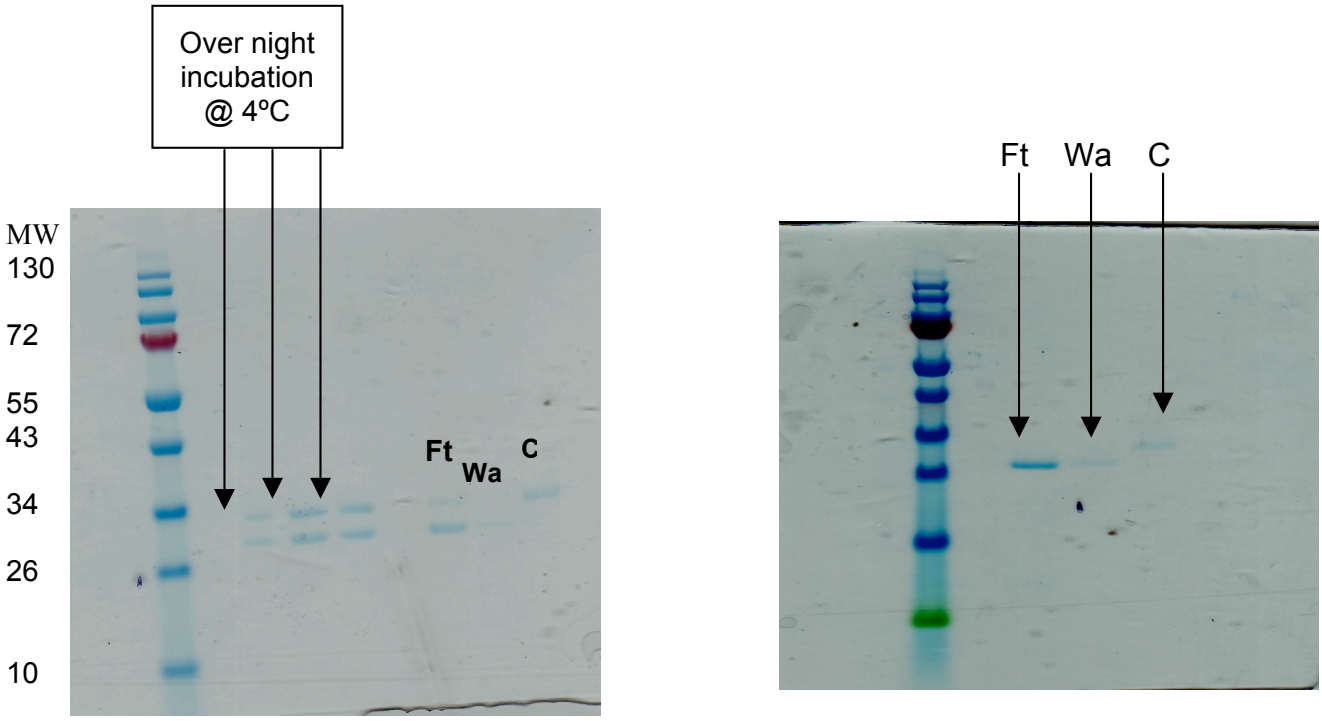


Figure 6.1. SynADC cleavage of N-terminal His-tag using AcTEV protease; Ft= flow through, Wa = wash and C = elution C. In figure (a), protein samples were incubated at 4ºC over –night with 10 units of AcTEV protease then pulled Ft and Wa samples, re-run them twice on the column to remove all the impurities before concentrating the samples to 10 mg/ml for crystallisation trials.

6.3.2 Crystallization Results

The purified enzyme was concentrated to ~ 10 mg/ml and microbatch crystallization experiments were carried out using commercial crystal screens (as in methods 6.5.1) Initial crystallisation trials yielded needle–like crystals after ~ 4 weeks, protein crystals produced in 0.04 M potassium dihydrogen, 16% w/v PEG 8K, 20% v/v glycerol (Figure 6.2). Micro-seeding approach was the attempt, which improved needle crystals to decent crystals suitable for data collection (Figures 6.3 – 6.7).

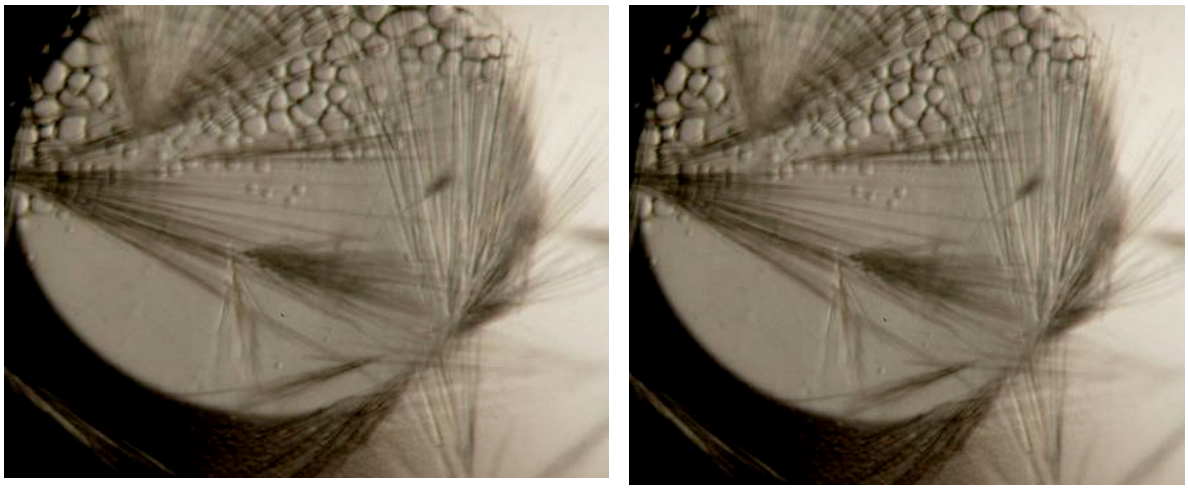


Figure 6.2. Typical 'needle' crystals obtained from screening experiment (JCSG plate). Crystal growth is mainly in one dimension only, resulting in crystals that not suitable for X-ray diffraction and date collection.

Optimization experiments were carried out around those conditions to yield decent crystals Figures 6.3 – 6.7), the crystals were frozen using a cryo-protectant (as in methods 6.12) and these were mainly obtained using JCSG crystallisation screen prepared from GF elution samples where 10% glycerol was removed, and sodium chloride replaced with 5% ammonium sulphate in the GF buffer preparations. Some crystals were co-crystallized with substrates (as in methods 6.11) (Figures 6.9 – 6.12), and other protein crystals were soaked with metal ions (as in methods 6.10) (Figures 6.8 a and 6.8b), and finally the frozen crystals were taken to the Diamond Light Source (Oxford, UK).

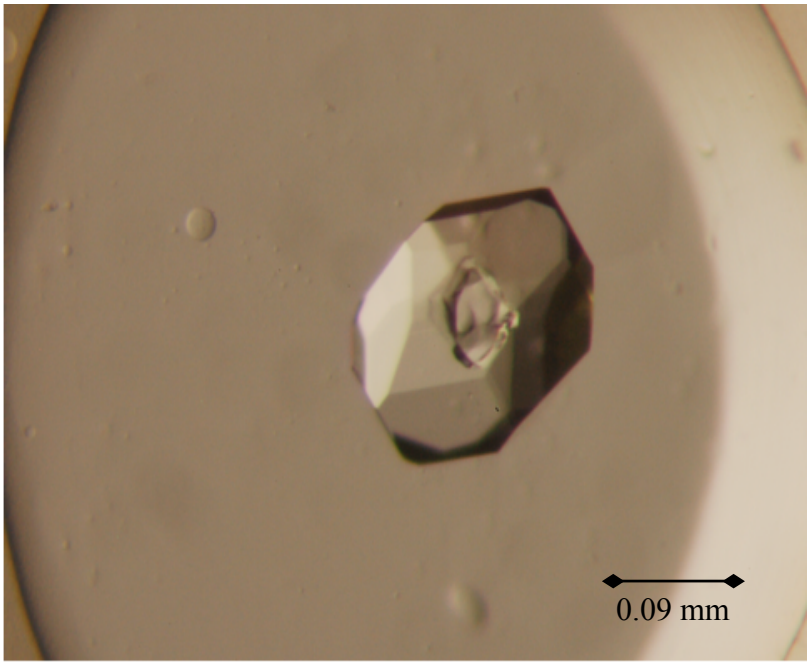


Figure 6.3. Crystal obtained by micro batch method from JCSG plate after one-week incubation at 19ºC, 20% (w/v) PEG3350, 0.2 M Tris potassium citrate with no pH and buffer.

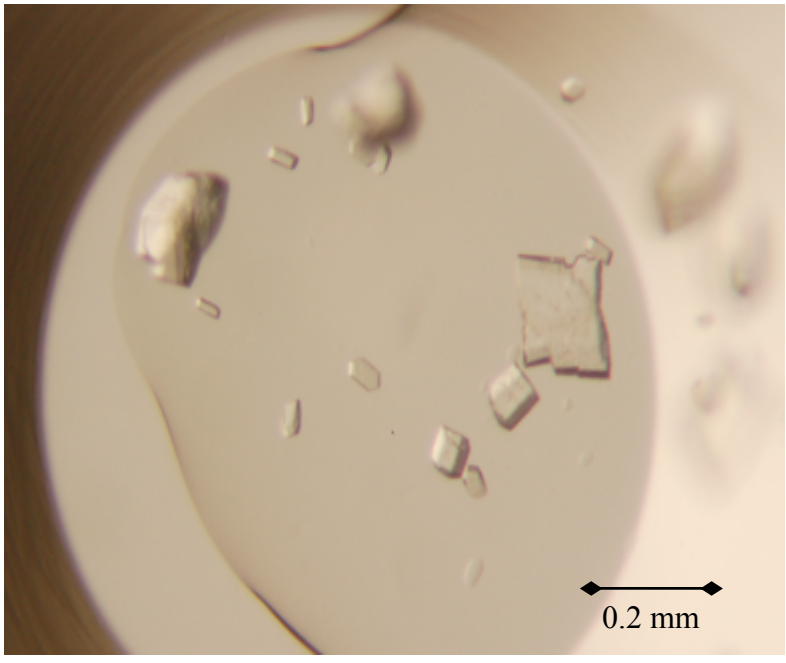


Figure 6.4. Crystal obtained by micro batch method from JCSG plate after one-week incubation at 19ºC, 14% (w/v) propanol / 30% (v/v) glycerol, 0.14 M Tris calcium chloride pH 4.6 and 0.07 Sodium acetate.

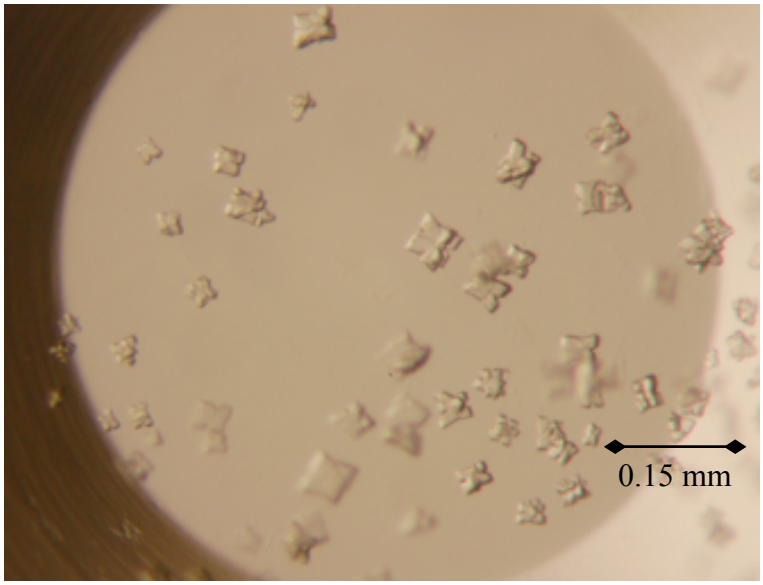


Figure 6.5. Crystal obtained by micro batch method from JCSG plate after one-week incubation at 19ºC, 14.4% (w/v) PEG 8K, 20% (v/v) glycerol, 0.16 M calcium acetate pH 6.5 and 0.08 M sodium cacoylate.

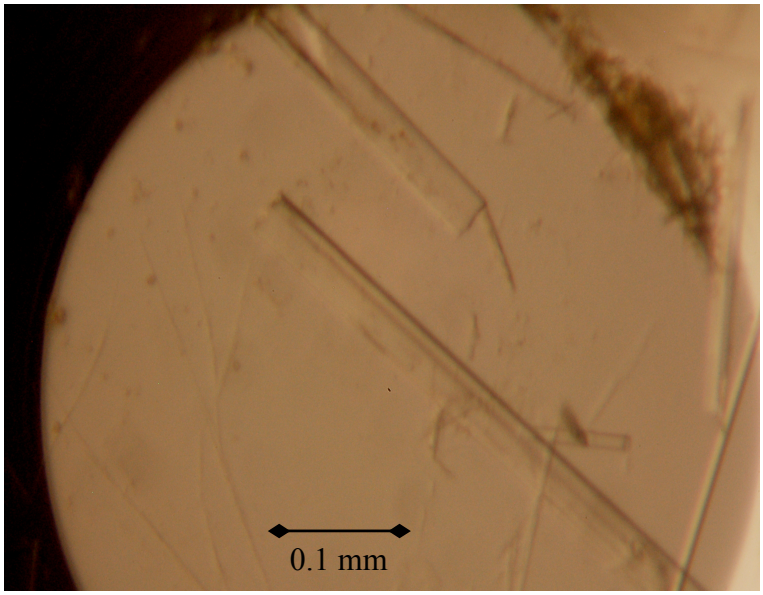


Figure 6.6. Crystal obtained by micro batch method from JCSG plate after one-week incubation at 19ºC, 40% (v/v) MPD, with no salt pH 10.5, and 0.1 M CAPS.

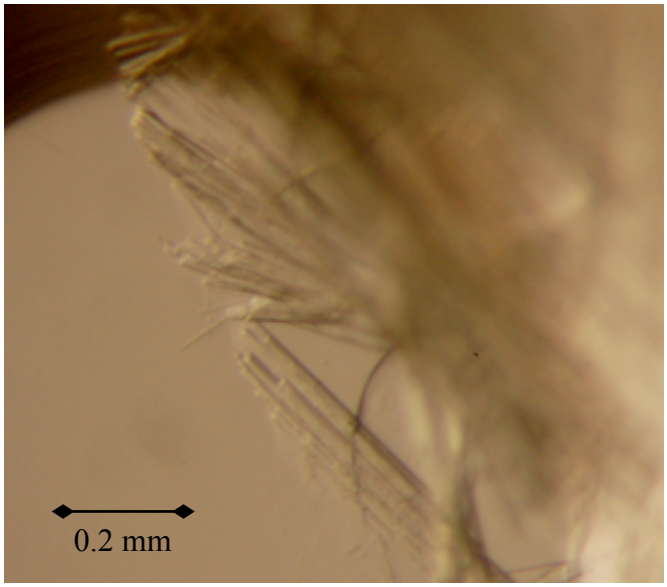
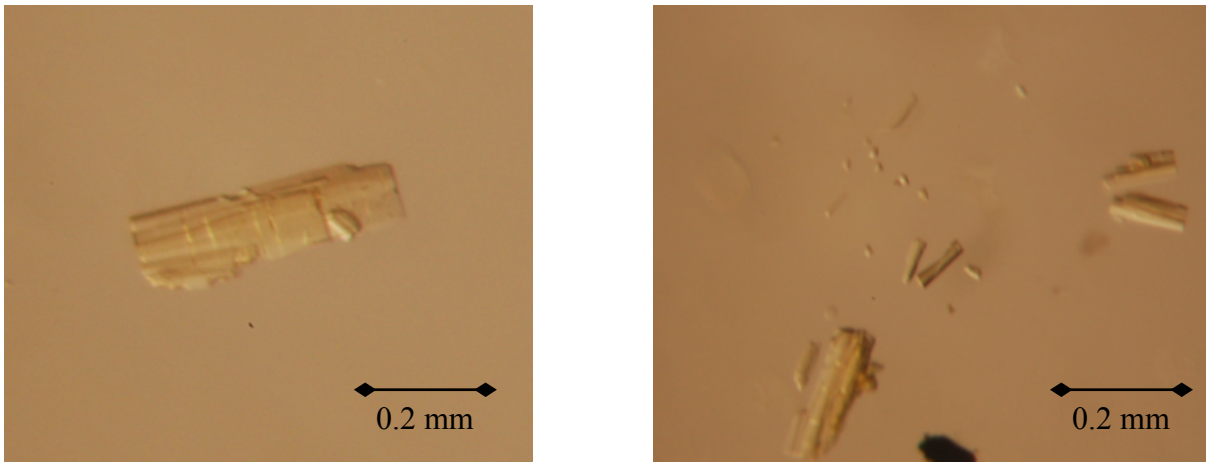


Figure 6.7. Crystal obtained by micro batch method from JCSG plate after one-week incubation at 19ºC, 0.8 sodium dihydrogen phosphate, 0.8 M potassium dihydrogen phosphate, with no salt pH 7.5 and 0.1 M sodium HEPES.

6.3.3 Soaking of protein crystals with ligands



a) **b)**

Figure 6.8. Protein crystals of SynADC soaked with ligands.

a) Protein crystals soaked after 24 hours. b) Protein crystals soaked with ligands for 72 hours.

6.3.4 Co-crystallization with ligands

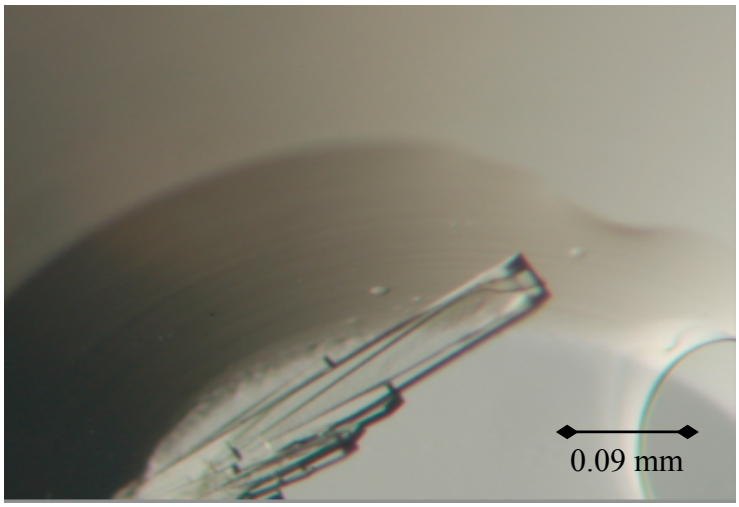


Figure 6.9. Co-crystallization of 50μl SynADC protein with 2.5μl of 200 mM of C4 aldehyde. Crystal obtained by micro batch method from JCSG plate after one-week incubation at 19ºC, under 0.2 M ammonium acetate, 0.1 M Bis Tris, pH 5.5 precipitants 25%, w/v PEG 3350 conditions, data collected at 1.8 Å.

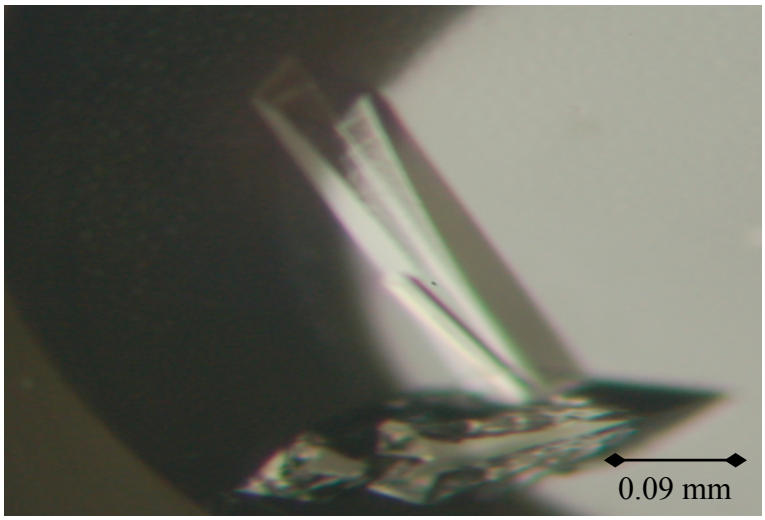


Figure 6.10. Co-crystallization of 50μl SynADC protein with 2.5μl of 200 mM of C4 aldehyde. Crystal obtained by micro batch method from JCSG plate after one-week incubation at 19ºC under; 0.2 M ammonium acetate, 0.1 M Bis Tris, pH 5.5 precipitants 25%, w/v PEG 3350, and data collected at 1.9 Å.

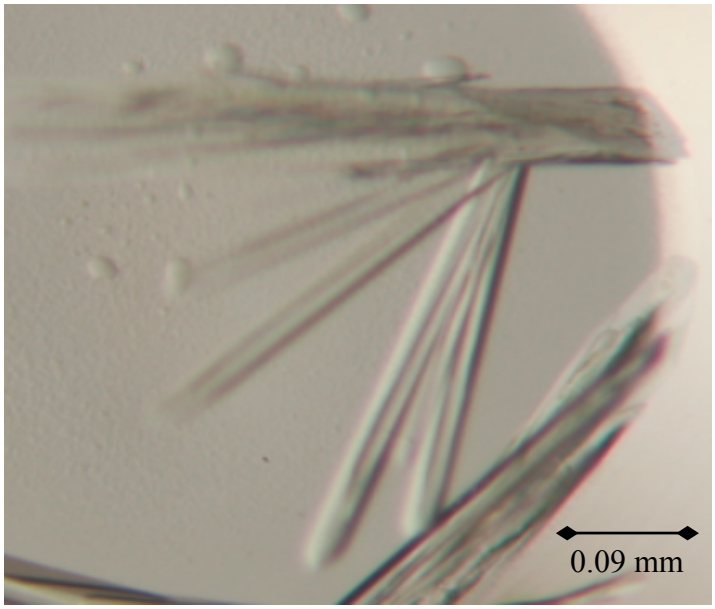


Figure 6.11. Co-crystallization of 50μl SynADC protein with 5μl C8 aldehyde. Crystal obtained by micro batch method from JCSG plate after one-week incubation at 19ºC under; 0.1 M ammonium sulfate, 0.1 M Bis Tris, pH 5.5 precipitants 17%, w/v PEG 10K, and data collected at 1.75 Å

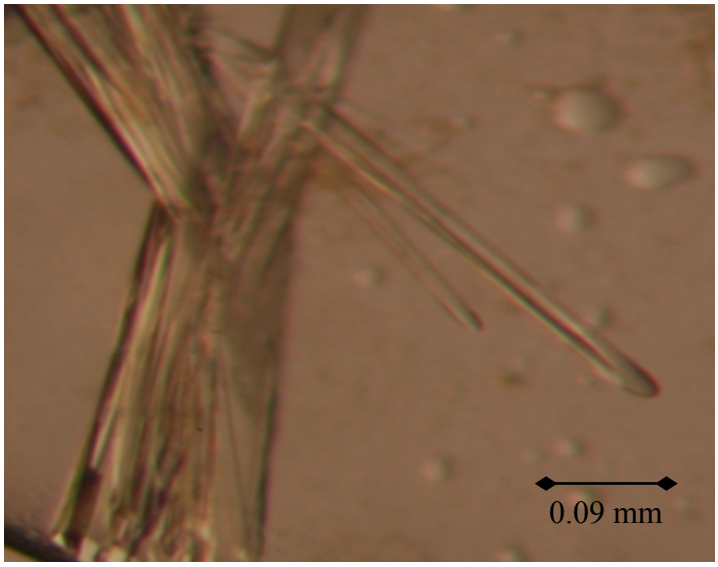


Figure 6.12. Co-crystallization of 50μl SynADC protein with 5μl C8 aldehyde. Crystal obtained by micro batch method from JCSG plate after one-week incubation at 19ºC under; 0.1 M ammonium sulfate, 0.1 M Bis Tris, pH 5.5 precipitants 17%, w/v PEG 10K, and data collected at 2.1 Å

6.3.5 Data collection

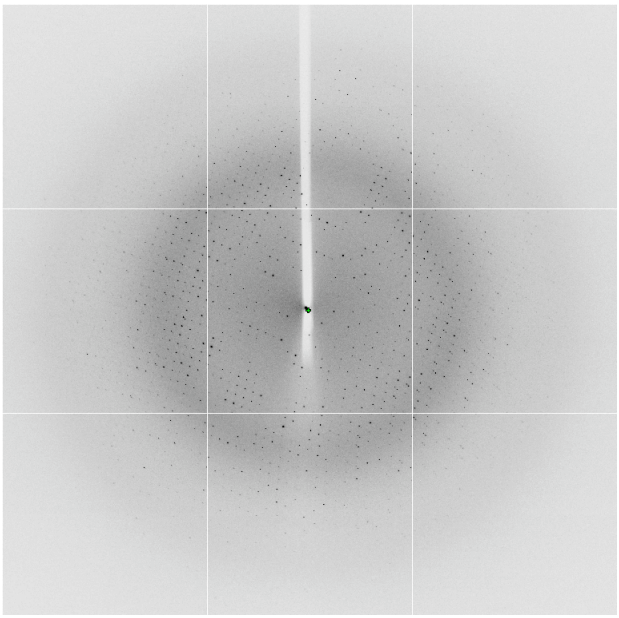


Figure 6.13. X-ray diffraction pattern for SynADC (Fatty Aldehyde Decarbonylase) protein crystals soaked in Fe^{2+} for 72 hours and crystals crystallised from conditions 0.1 M Bis Tris, pH 5.5, precipitants 2.0 M ammonium sulfate. Diffraction was to 1.7Å resolution.

Resolution	Overall	Low	High
Low resolution (Å)	57.76	57.76	1.76
High resolution (Å)	1.71	7.66	1.71
Rmerge	0.068	0.022	0.716
I/sigma	17.1	41.9	2.8
Completeness (%)	7.2	5.9	7.3
Multiplicity	7.2	5.9	7.3
Anomalous completeness	99.6	94.5	99.7
Anomalous multiplicity	3.6	3.6	3.6
Space group	$P2_12_12_1$		
Unit cell dimensions:			
a (Å) 63.9			
b (Å) 65.2			
c (Å) 115.5			
α (°) 90.0			
β (°) 90.0			
γ (°) 90.0			

Table 6.2 Statistics from X-ray diffraction at the Diamond Synchrotron for SynADC protein crystals soaked in Fe^{2+}

The cryo-protectant used helped to prevent to ice rings formation at 1.7Å resolution.

6.3.6 Structural analysis

Dr. Misha Isupov using the P. marinus enzyme as a model solved the crystal structure. The structures obtained are currently being refined.

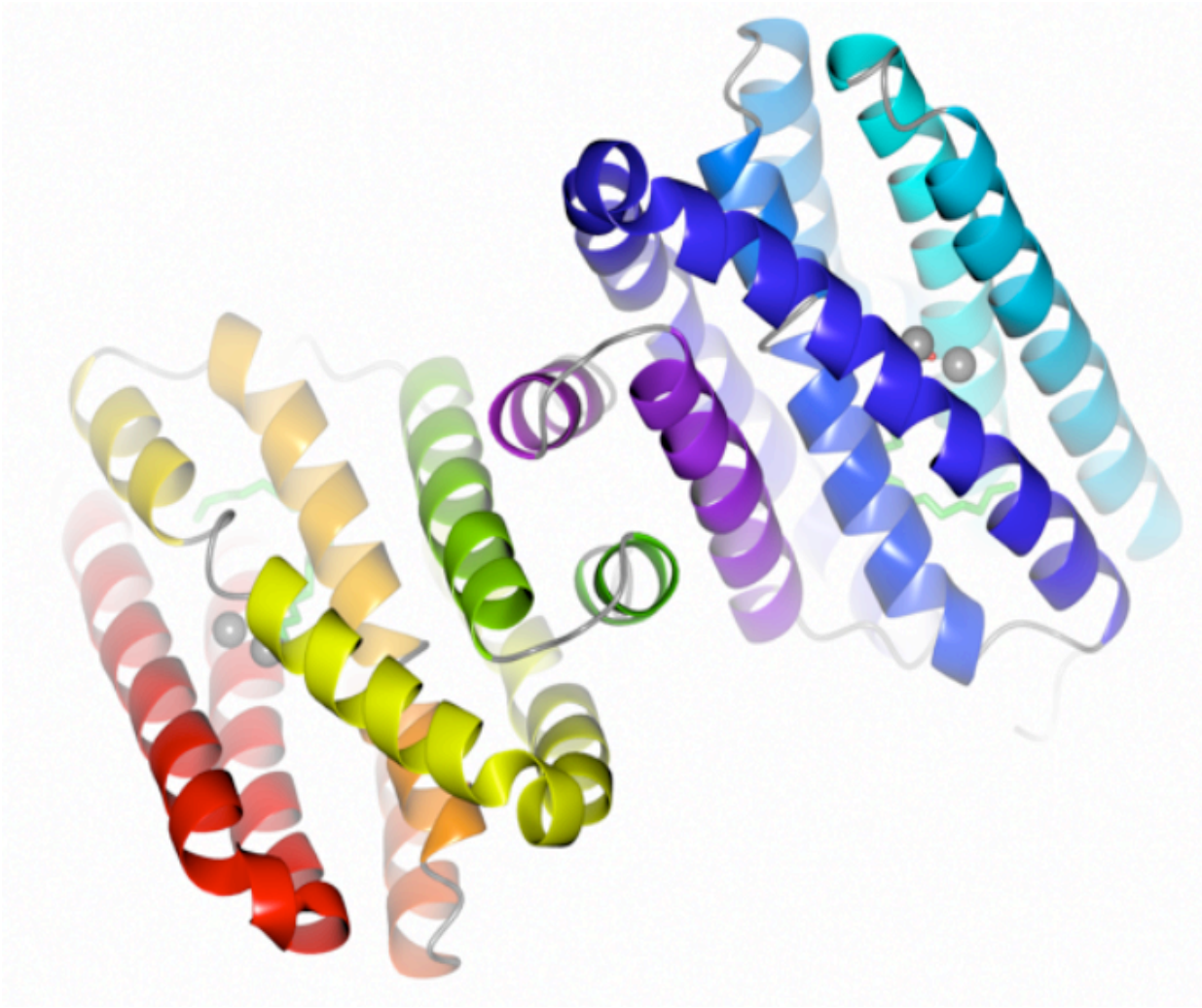


Figure 6.14. Ribbon representation of SynADC dimer viewed along molecular dyad. First monomer is pictured in red and green, monomer two in blue and purple. Picture produced using CCP4MG (McNicholas *et al*., 2011).

The crystal structures of (SynADC from *Synechocystis* and aldehyde decarbonylase from *P. marinus*) were aligned using secondary structure in the program COOT (Emsley *et al*., 2010). Pictures of the superimposition was created using Bobscript (Esnouf, 1997).

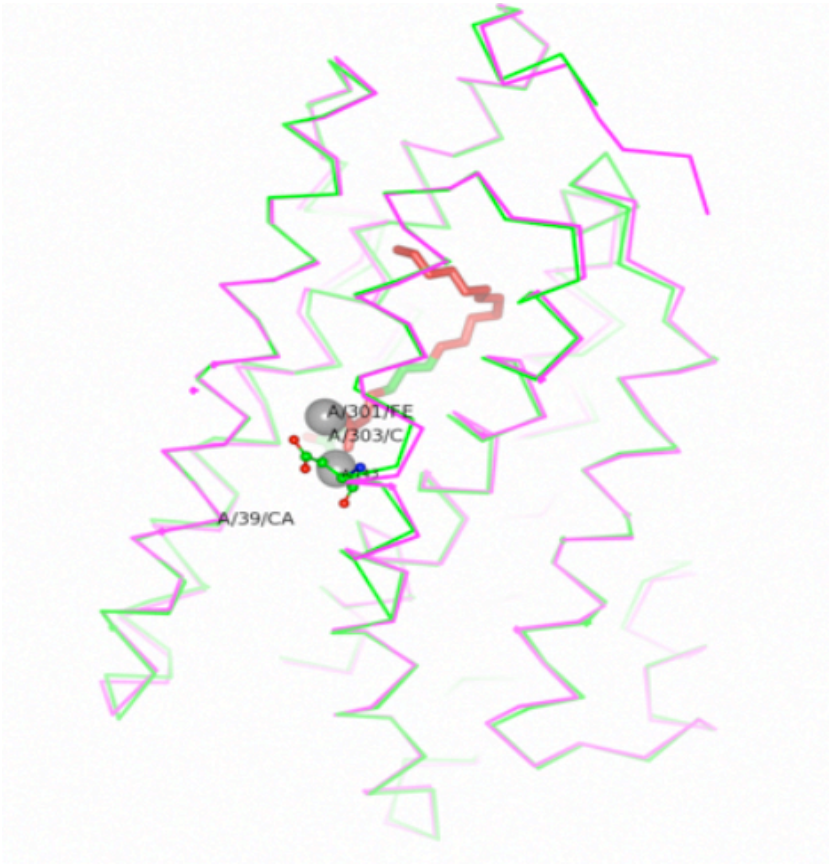


Figure 6.15 Superimposition of SynADC from *Synechocystis sp* is shown as green Cα trace with aldehyde decarbonylase from *P. marinus* (PDB 20C5A) (purple). Fe^{2+} ions are shown as CPK models. Fatty acid is shown as stick models. Catalytic residues are shown as ball-and-stick models. Pictures of the superimposition was created using Bobscript (Esnouf, 1997).

From the superimposition (Figure 6.13) SynADC dimer matches exactly with the AD from *P. marinus* (PDB 20C5A). The regions showing high similarity both in quaternary structure and primary sequences are those involved in the metal binding the four helical bundle motif.

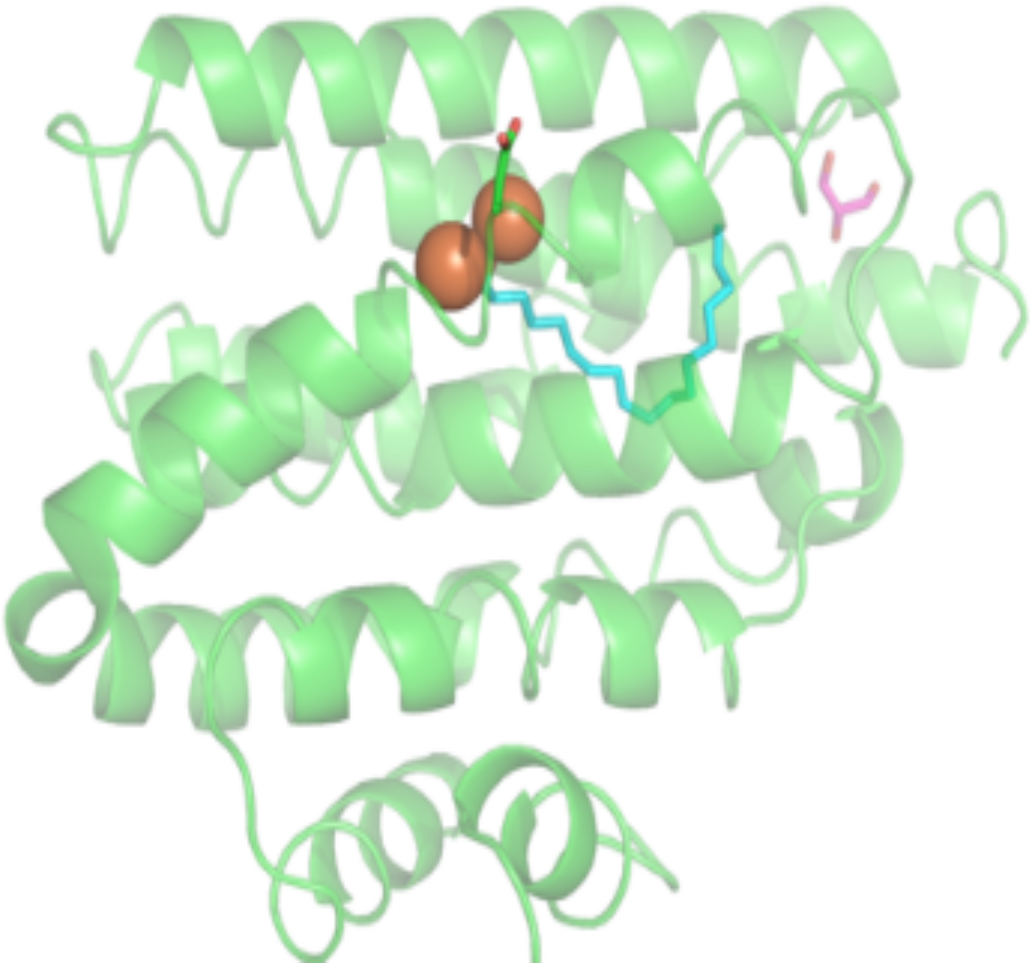


Figure 6.16. Ribbon representation of SynADC monomer. Metal irons (brown) are shown as CPK space-filing models. Fatty acid (blue), Aspartate (green) and Glycerol (purple) are shown as stick models. Picture produced using Pymol and rendered in Raster3D (Merritt and Bacon, 1997).

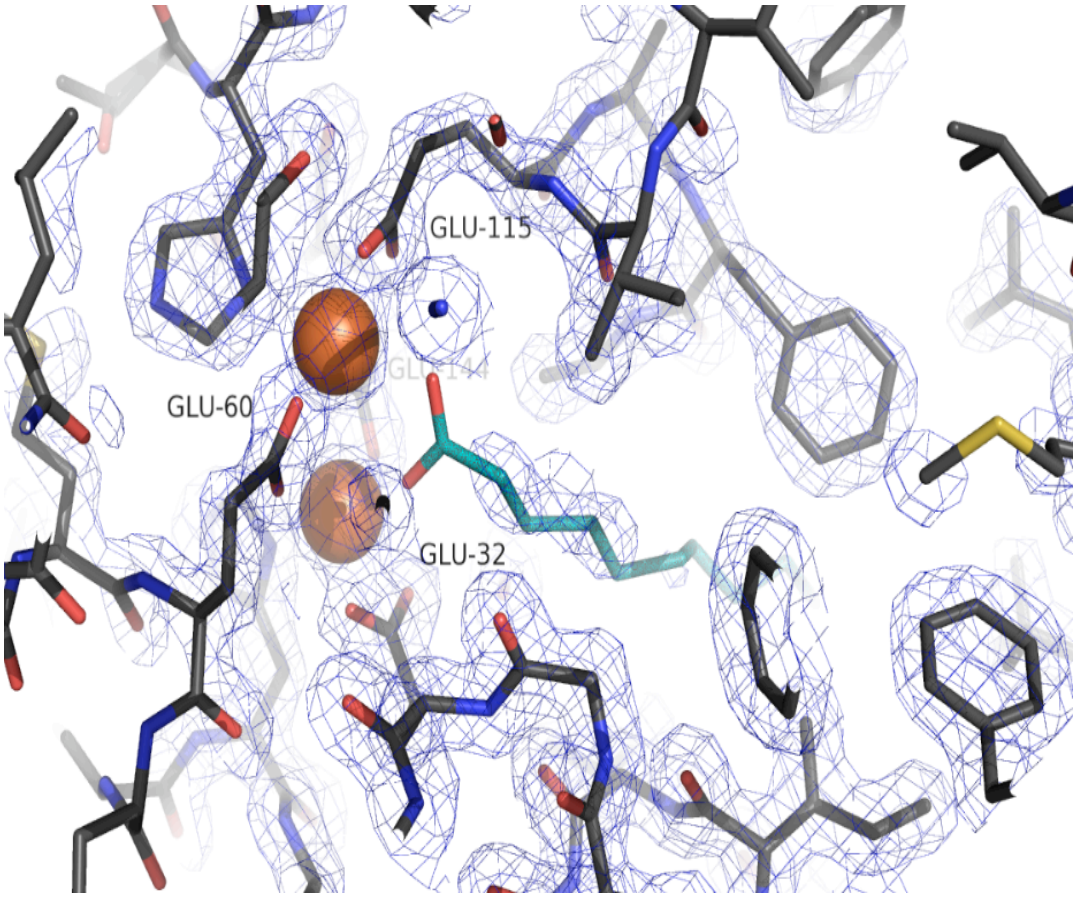


Figure 6.17. Experimental electron density and metal coordination around the enzyme active site; four Glutamic acid residues are involved. Picture produced using Pymol and rendered in Raster 3D(Merritt and Bacon, 1997). Electron density around the fatty acid ligand is not well defined, due to a partial occupancy of the ligand.

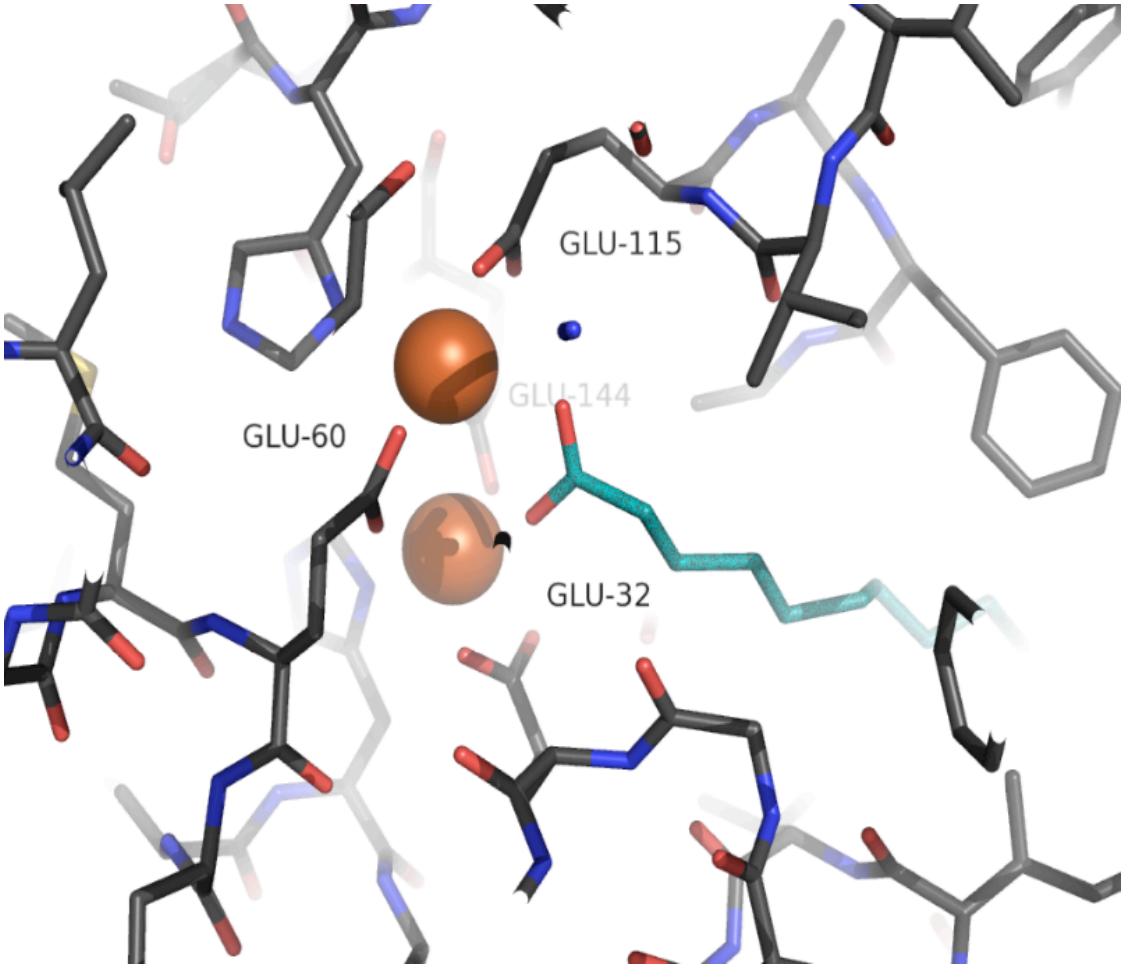


Figure 6.18. The coordination of metal ions in the active site, the two Fe^{2+} (brown) are shown as CPK space-filing models and fatty acid substrate (blue) as stick model. The two metal irons are coordinated with four glutamic residues. Picture produced using Pymol and rendered in Raster3D (Merritt and Bacon, 1997).

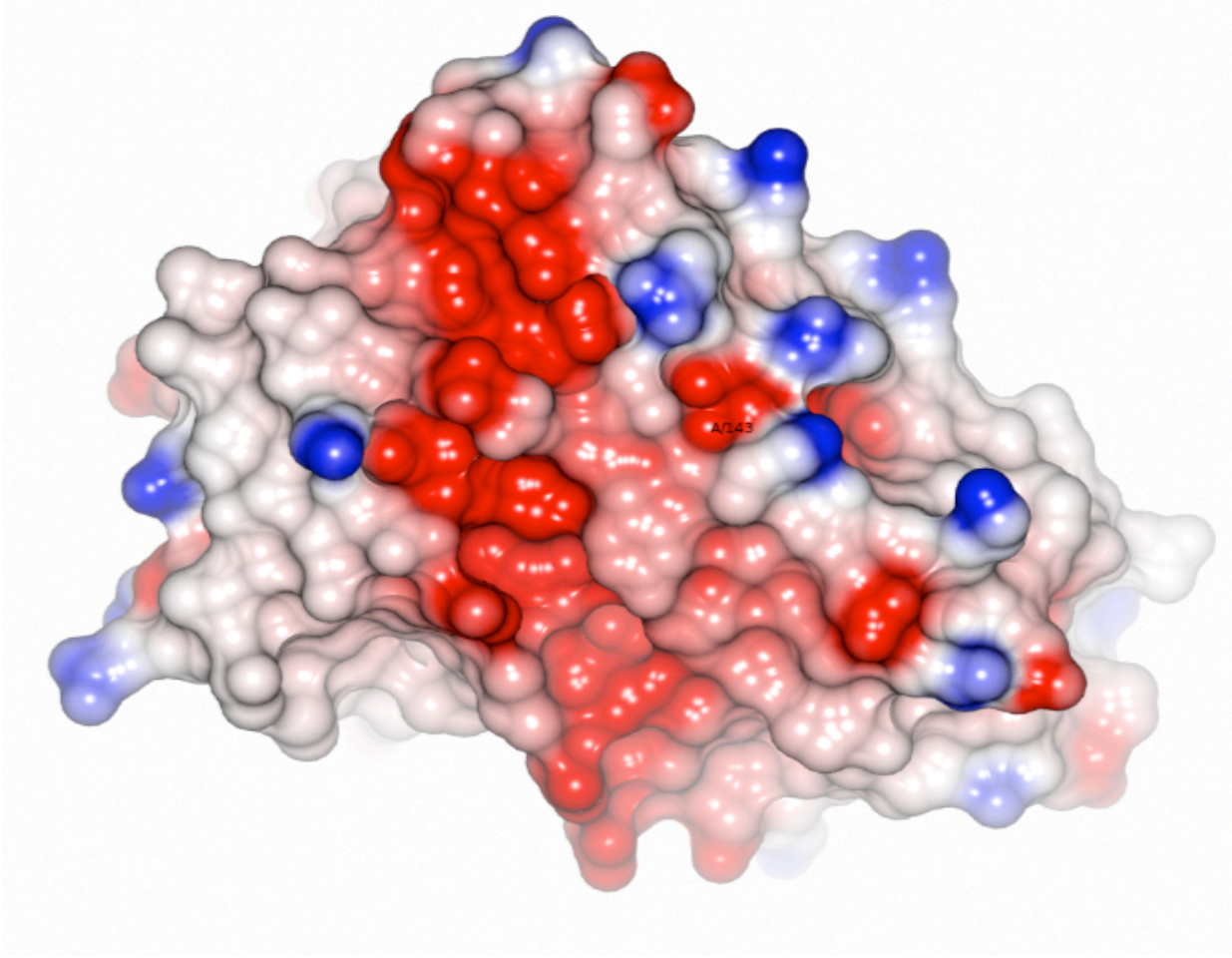


Figure 6.19 Analysis of the electrostatic potential, surfaces showing charge distribution (blue is positive, Red is negative and white is hydrophobicity of SynADC. We speculate distribution of charge to be around aspartate (Asp143). Picture produced in CCP4mg (McNicholas *et al*., 2011).

	Crystallisation Condition	Ligand and Metals added	Resolution	Data collection
1	0.1 M Bis Tris, pH 5.5, precipitants 2.0 M ammonium sulfate	Metal soaking + ligand (Fe^{2+} + C4)	1.5 Å	Collected
2	0.1 M Bis Tris, pH 5.5, precipitants 2.0 M ammonium sulfate	Metal soaking + ligand (Fe^{2+} + C8)	1.8 Å	Collected
3	0.1 M Bis Tris, pH 5.5, precipitants 2.0 M ammonium sulfate	Metal soaking + ligand (Fe^{2+} + Hexanoic acid)	2.3 Å	Collected
4	0.1 M Bis Tris, pH 5.5, precipitants 2.0 M ammonium sulfate	Soaked crystals in Fe^{2+} for 72 hrs	1.7 Å	Collected
5	0.1 M Bis Tris, pH 5.5, precipitants 2.0 M ammonium sulfate	Soaked crystals in Zn^{2+} for 72 hrs	1.6 Å	Collected
6	0.1 M Bis Tris, pH 5.5, precipitants 2.0 M ammonium sulfate	Metal soaking + ligand (Fe^{2+} + valeric acid)	2.05 Å	Collected
7	0.1 M Bis Tris, pH 5.5, precipitants 2.0 M ammonium sulfate	Metal soaking + ligand (Fe^{2+} + valeric acid)	2.7 Å	Not Collected
8	0.1 M Bis Tris, pH 5.5, precipitants 2.0 M ammonium sulfate	Metal soaking + ligand (Fe^{2+} + Hexanoic acid)	1.5 Å	Collected
9	0.2 M ammonium acetate, 0.1 M Bis Tris, pH 5.5 precipitants 25%, w/v PEG 3350	Co-crystallization with C4 aldehyde	1.8 Å	Collected
10	0.2 M ammonium acetate, 0.1 M Bis Tris, pH 5.5 precipitants 25%, w/v PEG 3350	Co-crystallization with C4 aldehyde	1.9 Å	Collected
11	0.1 M ammonium sulfate, 0.1 M Bis Tris, pH 5.5 precipitants 17%, w/v PEG 10K	Co-crystallization with C8 aldehyde	2.1 Å	Collected
12	0.1 M ammonium sulfate, 0.1 M Bis Tris, pH 5.5 precipitants 17%, w/v PEG 10K	Co-crystallization with C8 aldehyde	1.75 Å	Collected

Table 6.3 Summary of crystallographic data collected for SynADC protein using light source beam.

6.15 Discussion

The structure of SynADC is similar to *P*. *marinus* MIT9313 aldehyde decarbonylase reported by Schirmer *et al*. the protein has two iron atoms bound (as the three-dimensional structure shows similarities) (Figure 6.18), but it cannot be excluded that the active enzyme is a manganese/iron protein however, without an improved metal occupancy around the active site it is impossible to confirm. In the solved structures, both proteins have two irons coordinated to histidine and aspartate or glutamate residues, the amino acid sequence of the cyanobacterial ADs suggests they belong to the ferritin-like dimetal caboxylate protein family (Schirmer *et al*., 2010). The comparison of SynADC to the structure of the AD from *Prochlorococcus marinus* MIT9313 (*Pm*) which was solved by the joint Centre of structure Genomics before the function of the protein was known, confirms that these two structures are similar (Figure 6.15).

About the SynADC active site, CCP4mg molecular-graphics software was used to model (locate) the active site funnel but all attempts resulted to nothing (Figure 6.19). Cyanobacterial decarbonylase (SynADC) enzyme was found to have an enclosed active site, the substrate seems to be buried inside the protein with two identical subunits and with Asp143 residue sticking out on the surface, which thought to be the electron carrier chain into the active site. The speculation of how this enclosed active site work is that during activity, one of the subunit opens-up to allow access to the hydrophobic active site then closes immediately. Figure 6.19 shows analysis of the electrostatic potential, surfaces showing charge distribution around the active site with a speculation that distribution of charge takes place at aspartate Asp143. Although not discussed by Schirmer *et al*., 2010, the mutagenesis studies of Asp143 might help solve this mystery.

Analysis of metal ion coordination in active site have shown that the interaction is the same as reported in the previously known X-ray structure from *P. marinus* (PDB 2OC5A) Schirmer *et al*., (2010). However, in conditions where metal ion occupancy was low, the protein seems changing the conformation as was observed for native structure not soaked with Fe^{2+} (Table 6.3).

CHAPTER 7:

SITE DIRECTED MUTAGENESIS OF SynADC

7.0 Introduction

A model has been proposed for how the active site of SynADC (Figure 6.17) works. There is speculation that Asp143 could be the major channel of electron transfer from the outside of the protein into the hydrophobic active site. During activity one of the subunit opens-up, allowing access to the hydrophobic active site then closes immediately. To test the hypothesis it was necessary to mutate Asp143 into three other amino acids residues Asn, Leu, and Ala.

In this project the traditional PCR technique was used to produce both a full length and three truncated clones of SynADC protein with Asp143 mutated to three different amino acid residues (Asp143 to Asn, Asp143 to Leu and Asp143 to Ala) for further studies into the enzyme activity and the flow of electrons into the active site.

7.1 Materials and Methods

7.1.1 Site directed mutagenesis of SynADC

This work was done in collaboration with Dr. Christoph Edner at the University of Exeter. Primers were designed using QuickChange lightening (Stratagene Agilent Technologies) kit. The following three mutant enzymes were constructed according to the manufacturer's instructions.

- Asp143 ⟶ Asn (D143N): GAC ⟶ AAC
- Asp143 ⟶ Leu (D143L): GAC ⟶ CTG
- Asp143 ⟶ Ala (D143A): GAC ⟶ GCG

Control and sample reactions were prepared as shown in (Table 5.1) below. 1 µl of QuikChange Lightning enzyme was added to each control and sample

reaction, then reactions were cycled using the cycling parameters outlined in PCR program below. 2µl of *Dpn* 1 restriction enzyme was added then gently and thoroughly mixed each reaction were microcentrifuge briefly, then immediately incubated at 37ºC for 5 minutes to digest the parental dsDNA. 2µl of the *Dpn* 1-treated DNA from each reaction were transformed into separation 45µl aliquots of XL10-Gold ultracompetent cells as described in (Table 7.1) below. The transformation mixtures were spread on LB agar plates containing the appropriate antibiotic selection and incubated at 37ºC overnight. A single colony from the mutagenesis reaction was selected and grown overnight in 10 ml LB medium containing the appropriate antibiotic selection and the DNA was extracted and sent for sequencing to confirm the mutation (Source Bioscience, Nottingham, United Kingdom).

Control Reaction	**Sample Reaction**
5 µl of 10X reaction buffer	5 µl of 10X reaction buffer
5 µl (25 mg) of pWhitescript 4.5-kb control Template (5 mg/µl)	X µl (10 -100 mg) of dsDNA template
1.25 µl (125 mg) of control primer 1	X µl (125 mg) of oligonucleotide primer 1
1.25 µl (125 mg) of control primer 2	X µl (125 mg) of oligonucleotide primer 2
1 µl of dNTP mix	1 µl of dNTP mix
1.5 µl of QuikSolution reagent	1.5 µl of QuickSolution reagent
34 µl ddH_20 (for final volume of 50 µl)	ddH_2O to a final volume of 50 µl

Table 7.1 QuikChange Lightning Site-Directed Mutagenesis Kit

PCR Program

95ºC for 2 minutes

95ºC for 20 second

60ºC for 10 second

68ºC for 2.5 minutes

X 18

68ºC for 5 minutes

7.1.2 Amino acid residues for mutagenesis

D143N
Asp_{143} (GAC)-> Asn (AAC)
g715a
5'-gagggcgtagtcaagaacgaatacacccacc-3'

g715a_antisense
5'-ggtgggtgtattcgttcttgactacgccctc-3'

D143L
Asp^{143} (GAC)-> Leu (CTG)
g715c_a716t_c717g
5'-actgagggcgtagtcaagctggaatacacccacctcaac-3'
g715c_a716t_c717g_antisense
5'-gttgaggtgggtgtattccagcttgactacgccctcagt-3'

D143A
Asp^{143} (GAC)-> Ala (GCG)
a716c_c717g
5'-gagggcgtagtcaaggcggaatacacccacctca-3'
a716c_c717g_antisense
5'-tgaggtgggtgtattccgccttgactacgccctc-3'

Figure 7.1 Amino acid residues that had to be mutated are highlighted in purple

7.1.3 Expression of SynADC mutant proteins (D143N, D143L and D143A)

The expression of SynADC mutant protein was carried out as described in (section 4.2.1). The recombinant SynADC gene was transformed into *E.coli* BL21 (Merck Biosciences). The starter culture was grown under agitation at 180 rpm overnight at 37º and 10 ml used to inoculate fresh 1 L LB medium containing the appropriate antibiotic for selection, and was grown with agitation at 37ºC until the OD_{600} reached ~ 0.4 - 0.5. Then, a cold shock was applied to the culture by chilling to 15ºC for 30 min in a water-bath with ice and expression of SynADC induced with 1 mM IPTG (final concentration). Cultures were left at 15ºC for another 24 h with shaking before harvesting. The *E. coli* cells were harvested by centrifugation (8000g, 15 min, 4ºC) using Beckman JA- 25.50 rotor. The cell pellets were stored at -80ºC until further use.

7.1.4 Activity assays for SynADC mutant proteins

To test the aldehyde decarbonylase (AD) mutant proteins for activity, the following assays were set-up: 500 µl reactions containing 250 mM sodium phosphate buffer at pH 7.2 with the following components at their respective final concentrations: 30 µM of purified ADC, 200 µM substrate (octaecanal, C_{12}, C_{13}, and C_{18}), 25 µg/ml NADH), 25 µl PMS, and 60 µl of water. Negative controls included the above reaction but either without the enzyme (SynADC) or without substrate. Each reaction was incubated at 37ºC for 2 hours before being extracted with 100 µl ethyl acetate into 2 ml glass auto-sampler vials, then tested enzyme activity.

The flow rate of the helium carrier gas in the GC-MS analysis machine was set to 1.1 ml /min, with the inlet temperature maintained at 320ºC. Injections were made in split mode with a split ratio of 5:1 and a total flow of 5.7 ml/min. The oven temperature was held at 70ºC for 2 min and then increased to 280ºC at 20ºC /min and finally maintained at 280ºC for 5 min. samples were loaded into the machine and the conversation of aldehyde to alkanes was detected by GC-MS.

Chemicals

250 µl sodium phosphate buffer a pH 7.2
10 µl ferrous ammonium sulfate
100 µl substrate (octaecanal, C_{12}, C_{13}, and C_{18})
25 µl pms
25 µl NADH
30 µl SynADC
60 µl Purite ^{TM} water

7.2 Results

The speculation was; the distribution of charge (the flow of electrons) is taking place around aspartate Asp143 located at the outside part of the SynADC protein structure (Figure 6.17). Though the Asp143 residue is not in direct coordination with the

active site compared to glutamic acid residues (Glu-115, Glu-60, Glu-32 and Glu-144) (Figure 6.16). The hypothesis followed was that SynADC has a closed active site' the only way electrons can move in and out of the active site is by being picked up from the outside the active site by aspartate and then the electrons follow a route to the active site which is yet to be established.

7.2.1 Site directed mutagenesis of Asp143

Site-directed mutagenesis of Asp143 (GAC) to Asn (AAC), Leu (CTG) and Ala (GCG) was carried out as described in methods 7.1.1. The DNA was then transformed into *E. coli* expression strain BL21 (DE3) (Merck Biosciences). DNA sequencing (APPENDIX 9.7) (Source Bioscience, Nottingham, UK) was used to confirm the mutation of the aspartate residues at position 143 to Alanine, Leucine and Asparagine.

7.2.2 Expression of D143N SynADC

The D143N (Asp143 (GAC) -> Asn (AAC) construct was transformed into the *E. coli* expression strain BL21. The protein was induced by IPTG. The over-expression of SynADC was investigated by induction studies to obtain the best conditions for over-expression. The optimal over-expression was achieved as described in methods 7.1.2 and over-expression resulted in soluble protein being produced at the expected size (~29 kDa) (Figure 7.2).

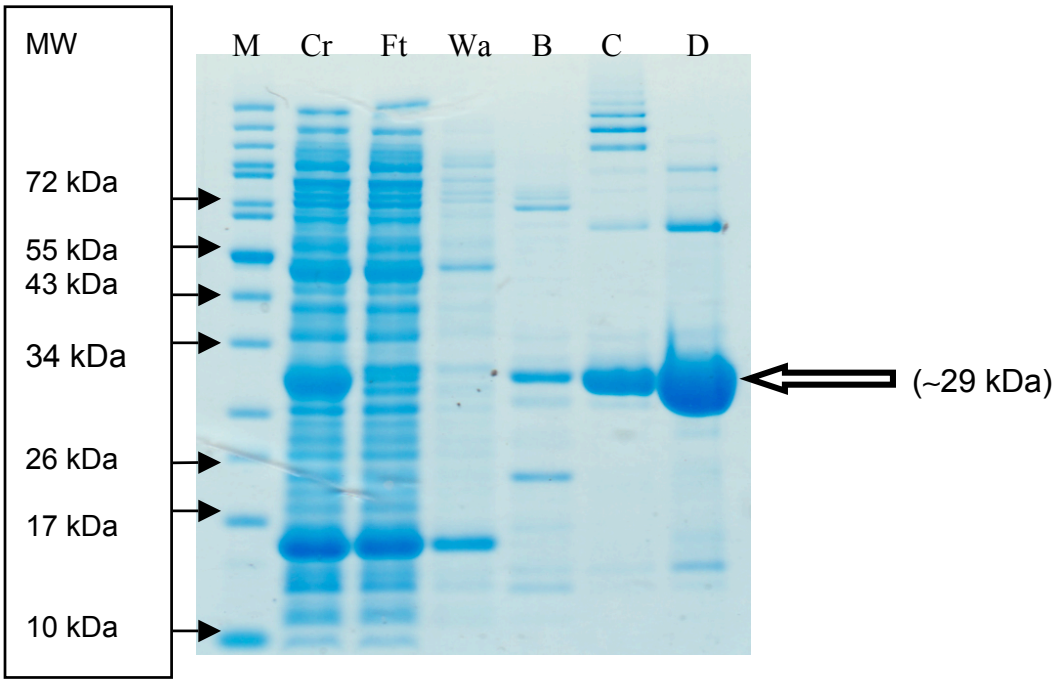


Figure 7.2. SDS-PAGE analysis of the site directed mutagenesis of Asp143

Where

(M) = Marker (Cr) = Crude extract (Ft) = Flow-through (Wa) = Wash

(B) = Elution buffer B (100 mM Imidazole) (C) =Elution buffer C (250 mM Imidazole)

(D) = Final enzyme preparation

7.2.3 Expression of D143L SynADC

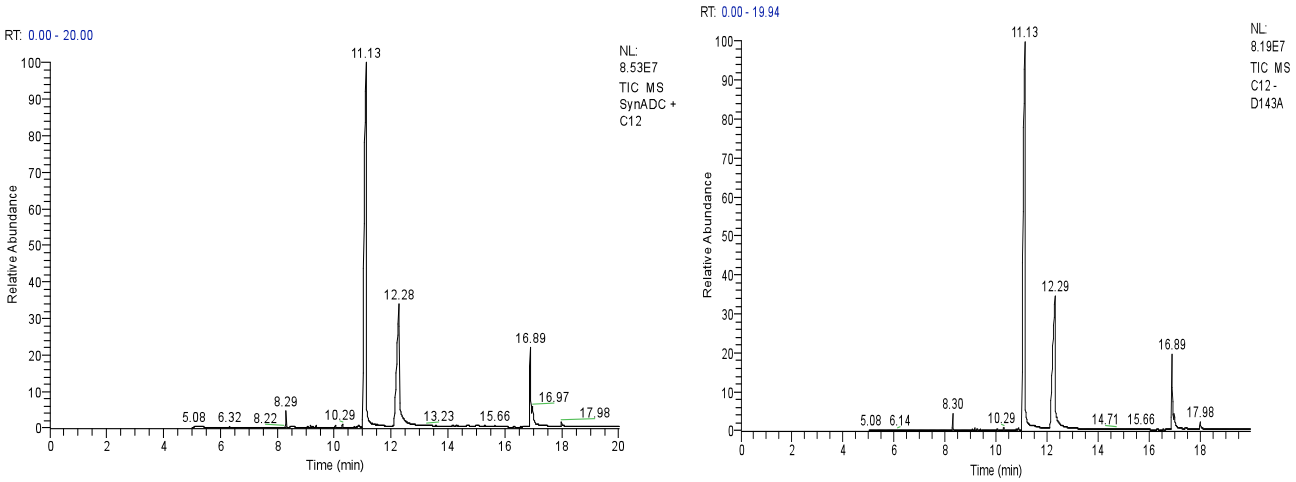
The D143L **(**Asp143 (GAC)-> Leu (CTG)) construct was transformed into the *E. coli* expression strain BL21. The protein was induced by IPTG. The over-expression of SynADC was investigated by induction studies to obtain the best conditions for over-expression. The optimal over-expression was achieved as described in methods 4.2.1 and over-expression resulted in soluble protein being produced at the expected size of ~29 kDa (Figure 7.2).

7.2.4 Expression of D143A SynADC

The D143A (Asp^{143} (GAC) -> Ala (GCG)) construct was transformed into the *E. coli* expression strain BL21. The protein was induced by IPTG. The over-expression of SynADC was investigated by induction studies to obtain the best conditions for over-expression. The optimal over-expression was achieved as described in methods 7.1.2 and over-expression resulted in soluble protein being produced (as described in 7.2.3 above) to produce a protein size of ~29 kDa.

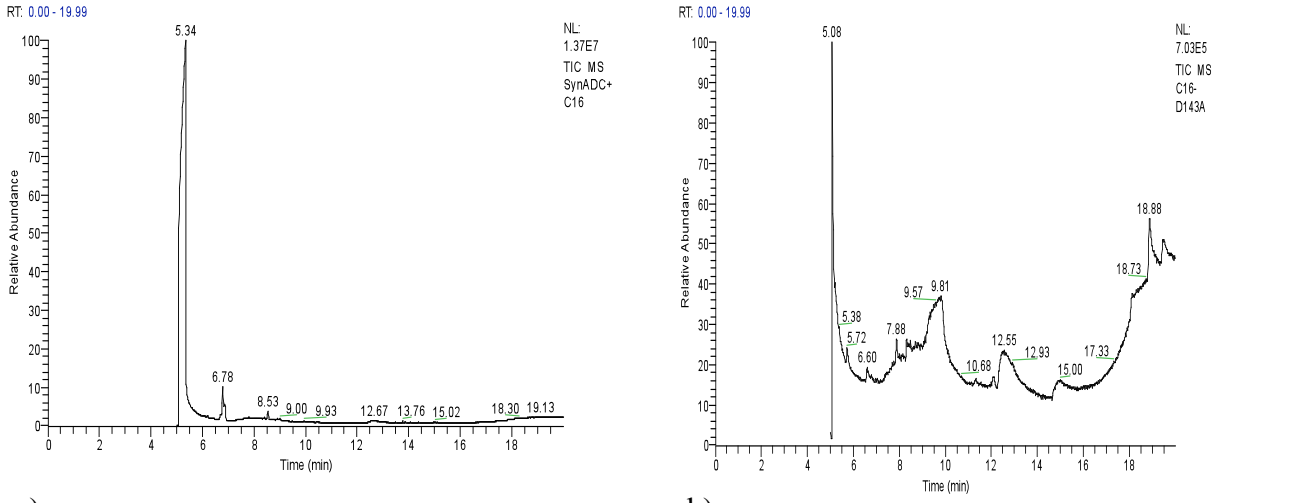
7.2.5 Activity assays for SynADC Mutant proteins

The conversation long chain aldehydes (C12, C16 and C18) to respective alkanes by aldehyde decarbonylase (SynADC) were found not to work. To determine the alkane formation, 500 µl reactions containing 250 mM sodium phosphate buffer, with the following components at their respective final concentrations (30 µM of purified SynADC, 200 µM substrate, 25 µg/ml NADH, 25 µl PMS, and 60 µl of water were prepared. The GC – MS which readily detected nothing, directly sampled the reaction products for production of alkane (Figures 7.3 – 7.5). The activity assay results reported below are for one mutagenesis amino acid residue (Asp^{143} -> Ala); others are yet to be carried out.



a) b)

Figure 7.3. GC–MS chromatograph showing conversion of 10 μl C_{12} aldehyde to alkane by SynADC mutant enzyme. (a) Mass spectrum produced when (a) 30 μl SynADC wild type protein was used to act upon 10 μl C_{12} aldehyde to alkane, (b) 30 μl SynADC mutated protein (D143A) was used to act upon 10 μl C_{12} aldehyde to alkane. No alkanes were formed.



a) b)

Figure 7.4. GC – MS chromatograph showing conversion of 10 μl C_{16} aldehyde to alkane by SynADC mutant enzyme. (a) Mass spectrum produced when (a) 30 μl SynADC wild type protein was used to act upon 10 μl C_{16} aldehyde to alkane (b) 30 μl SynADC mutated protein (D143A) was used to act upon10 μl C_{16} aldehyde. No alkanes were formed.

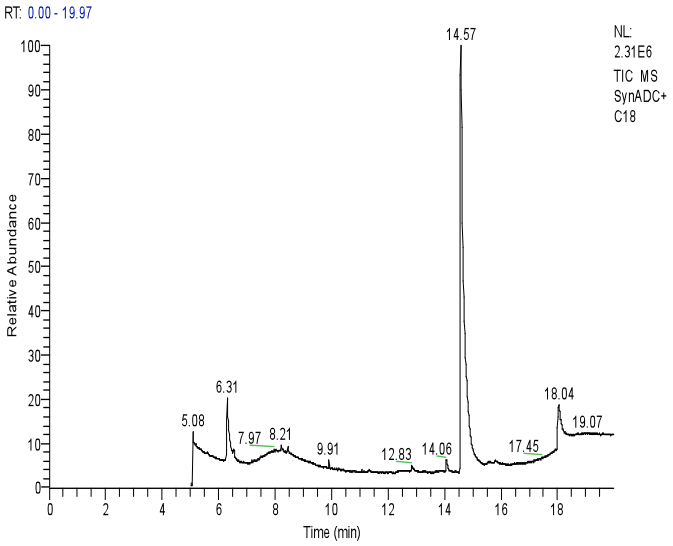


Figure 7.5. GC – MS chromatograph showing conversion of 10 μl C_{18} aldehyde to alkane by SynADC mutant enzyme. Mass spectrum produced when 30 μl SynADC mutated protein (D143A) was used to act upon 10 μl C_{18} aldehyde substrate. No alkanes were formed.

7.3 Overview

The site-direct mutagenesis of Asp143 to Asn, Leu and Ala were carried out successfully as described in method in methods 7.1.1. Mutant genes (D143N, D143L and D143A) expressed well in *E.coli* and the proteins were induced by the addition of 1 mM IPTG to the final concentration to produce a soluble protein of ~29 kDa as before the mutation. DNA sequencing (APPENDIX 9.7) (Source Bioscience, Nottingham, UK) confirmed the mutation of the Aspartate residues at position 143 to Asparagine, Leucine and Alanine residues.

However, results from activity assays using SynADC mutant protein were inconclusive. Neither the wild type AD nor the mutant AD had activity when used to convert long-chain aldehyde (C_{13}, and C_{18}) to their respective short chain alkanes. However in the first activity assays carried out (chapter 4 Figures 4.5 and 4.6) had activity. But since the two assays were set-up differently, NADH replaced Ferredoxin used during the first attempt assays. The activity assays for the mutant proteins will be repeated using Ferredoxin as an electron donor, which proved to give positive results (Figure 4.5 and Figure 4.6) to complete this preliminary phase.

CHAPTER 8:

CONCLUDING COMMENTS AND FUTURE WORK

8.1 Summary and Concluding Comments

The aim of this research was to solve the protein structure of Cyanobacterial aldehyde decarbonylases. The work presented in this thesis carried out by me describes; over expression, protein purification, site directed mutagenesis, spectroscopy studies and crystallisation, co-crystallisation with substrates, crystal soaking with both metals and substrates. The native molecular weight of the protein was found to be 28 kDa; a definitive elution profile was obtained on the gel filtration column. Protein crystallisation of SynADC has been achieved with crystals from various crystallisation conditions, protein crystals with ligands added in the active site.

Dr. Edner Christoph – a postdoc researcher at Exeter University, cloned the gene encoding SynADC protein isolated from *Synechocystis sp*. into pET160/GW/D-TOPO and this was overexpressed in *E.coli*. The protein formed some inclusion bodies and a percentage present in a soluble form.

These results have three consequences. The first is that SynADC is now available readily overexpressed in *E.coli,* the protein easily purified in just two chromatography steps and ready for X-ray crystallography. The protein is cloned in a cost-effective vector with a small tag attached; now the protein can be crystollised without cleaving off the His-tag. SynADC–Tag (without a tag) protein samples were hard to purify, had to go through many purification steps (ammonium sulphate precipitation, phenylsepharose column, ion exchange (FFQ) column, and then GF), lots of protein get lost during purification steps. With the large number of conditions tried during this screen a certain amount of confidence has been gained that the best vector system has been optimised and near optimal conditions for the crystallisation process of SynADC have been identified.

The second consequence is that SynADC structure has been solved by Dr. Misha Isupov with substrates (C_4 aldehyde, C_8 aldehyde, Hexanoic acid and valeric acid) added in the active site. It was originally hoped that the data would provide some defined understanding of how electrons are transported through the system but with low metal occupancy within the protein structure this has made it impossible to understand the mechanism. Structures obtained are from co-crystallisation protein soaking techniques. Addition of metals to co-crystallized plates and soaking crystals in substrates containing excess metal ions improved the metal occupancy in the structure and improves the active site, we are yet to confirm and quantify if it recovers activity as well. The second approach was to try to strip off all metals and simultaneously identify the features of SynADC metals responsible for its activity and crystallisation. This proteins needs iron (Fe) for activity and maybe it requires all metals (Fe, Zn, Ni and Mn) as we failed to get the apo-structure of SynADC.

The third consequence is that a successful site directed mutagenesis of Asp143 to Asn, Asp to Leu and Asp to Ala amino acids of SynADC protein has been achieved and this is the first, directed mutagenesis study of Asp143 residue of on this protein. Characterisation of these mutant proteins are yet to be carried out, Asp143 is thought to be the source electrons to the enzyme active site and it will be interesting to know if this has improved or has lowered enzyme activity. Further work needs to be carried out to find out this unknown.

8.2 Future Work

There are number of ways in which the work contained within this thesis can be taken forward with most of the further work focusing on the metal occupancy in active site and to solve the hypothesis of electron transfer mechanism for the structure.

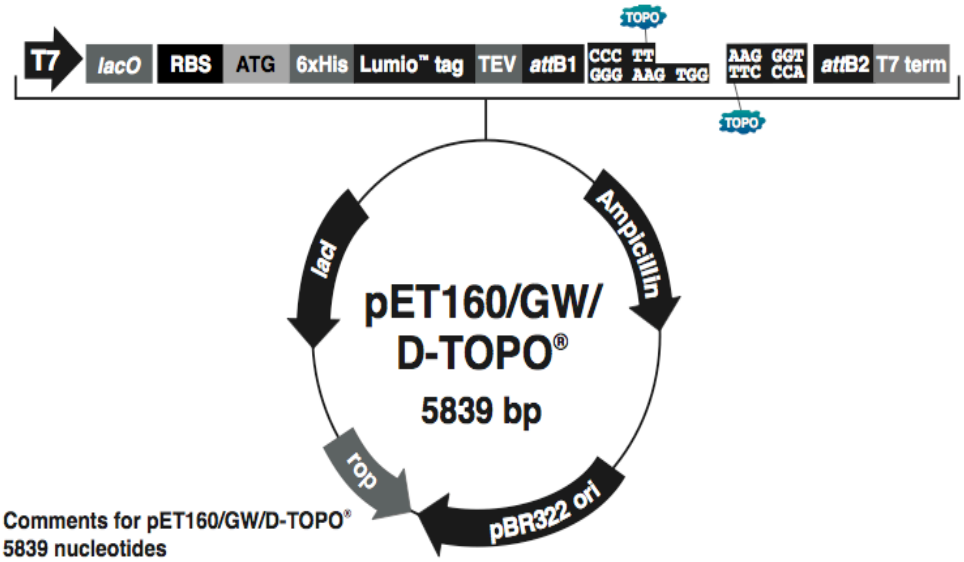
Now that large quantities of SynADC are available many avenues are open for extending the research carried out as part of this work. A more comprehensive study into the mechanism of electron transfer within the structure is now possible since site-direct mutagenesis of Asp143 into various amino acids has been achieved. As can be seen from the introduction in Chapter 1 the true mechanism of the flow of electrons and the role of metals attained with this enzyme is still not fully understood and more work is needed in this field, which has previously been hampered by the lack of protein available. The substrate specificity of SynADC is still need to be full investigated.

Failure in produce the apo-structure of this enzyme is also interesting and an understanding of the metal content needed for activity needs to be established. Use of other techniques such as circular dichroism could be used to look at any structural changes in mutant protein and in apo and holo structures of the enzyme.

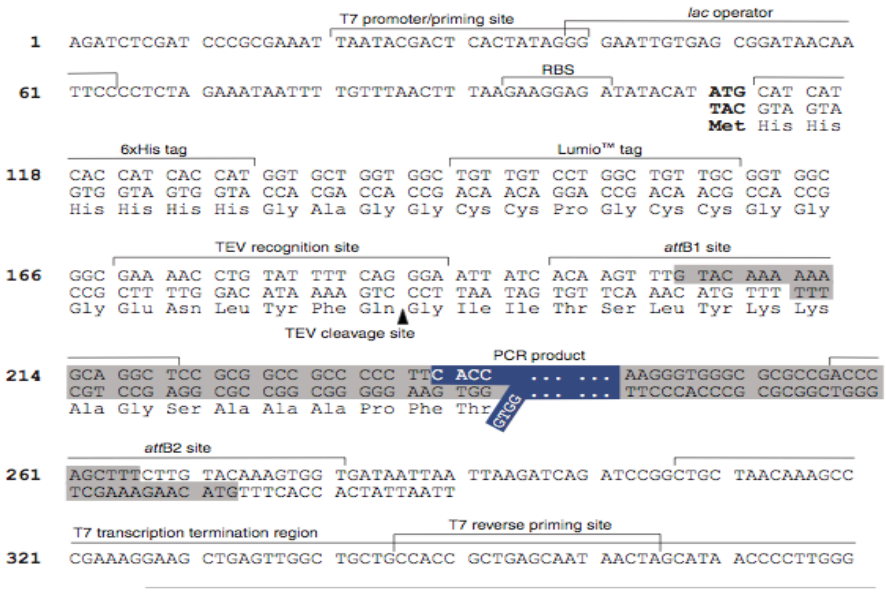
Now that SynADC protein from *Synechocystis* can now easily be over-expressed in *E.coli* and purified without use of expensive proteases, producing the protein samples to further tweaking with the X-ray structure and activity assays can be carried out.

9. Appendix

9.1 Vector Map of pET160/ GW/D-TOPO

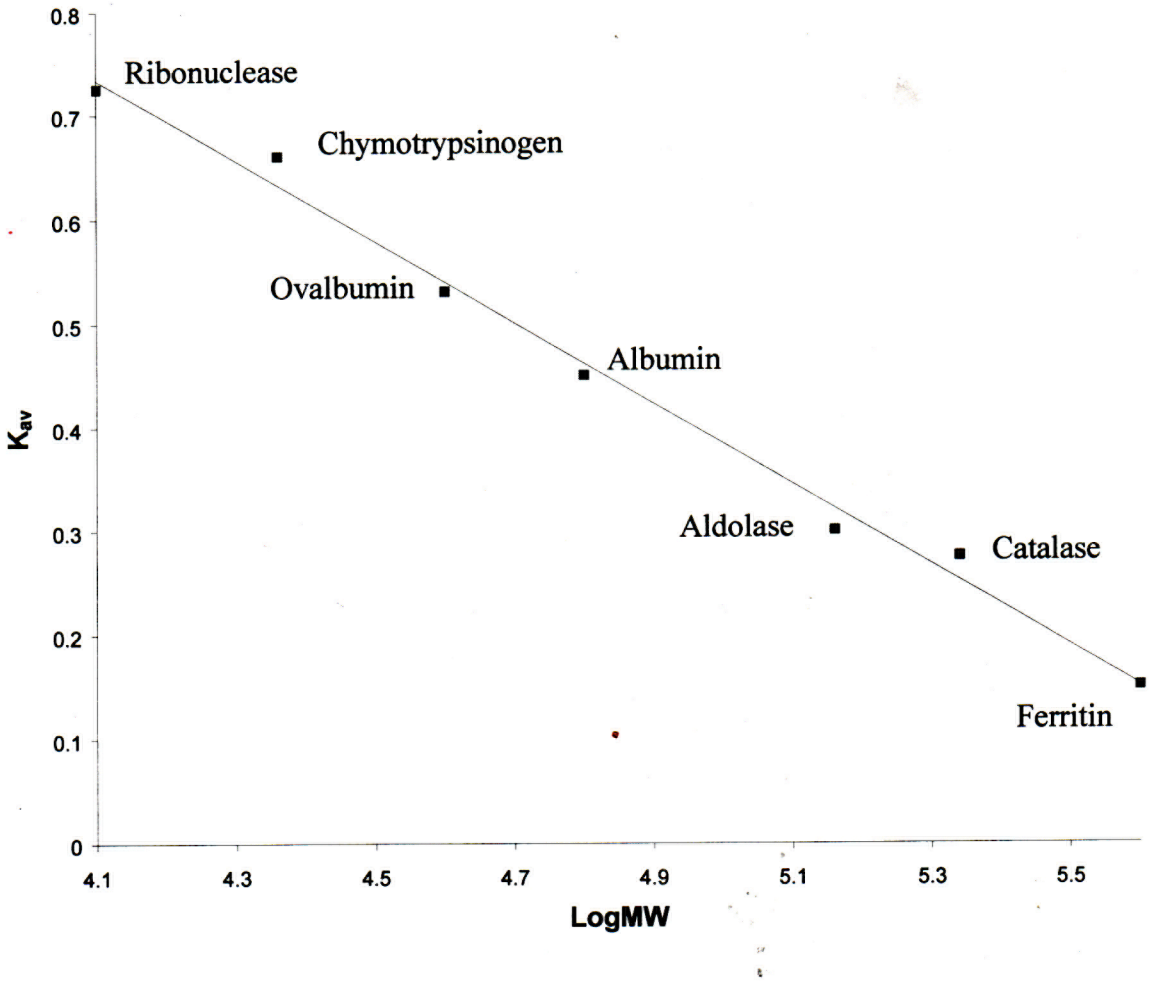


9.2 Cloning site of PET160/GW/D-TOPO

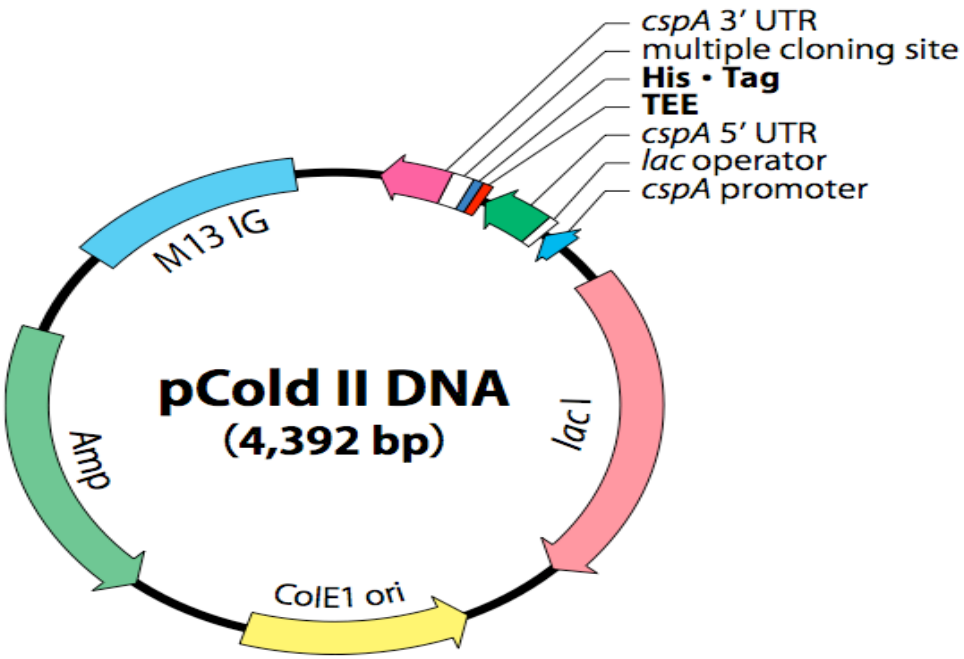


9.3 Superdex 200 gel filtration column calibration

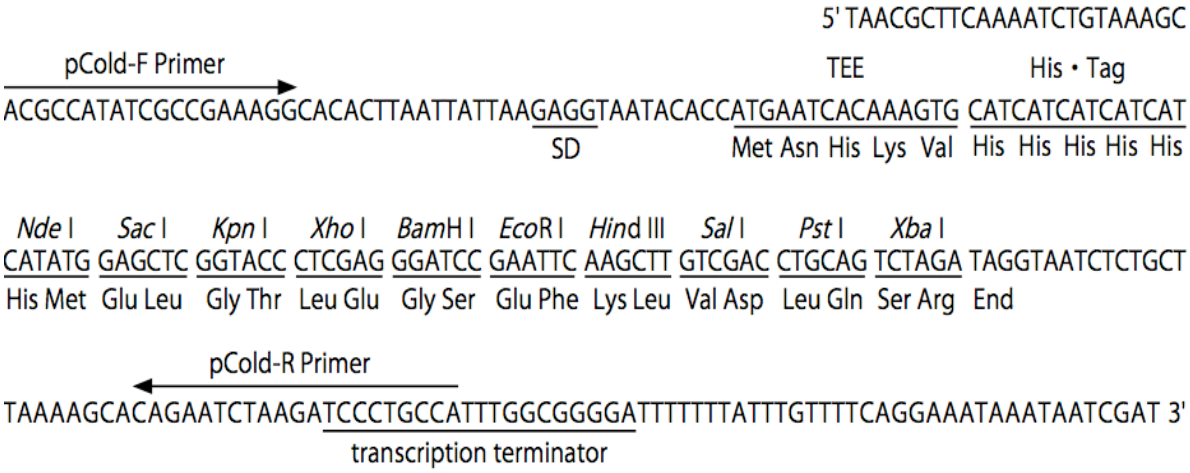
Superdex 200 Gel Filtration Column Calibration



9.4 pCold II DNA (Vector Map of pCold II DNA)



9.5 Cloning site of pCold II DNA



9.6 pCOLD_SynADC.ape

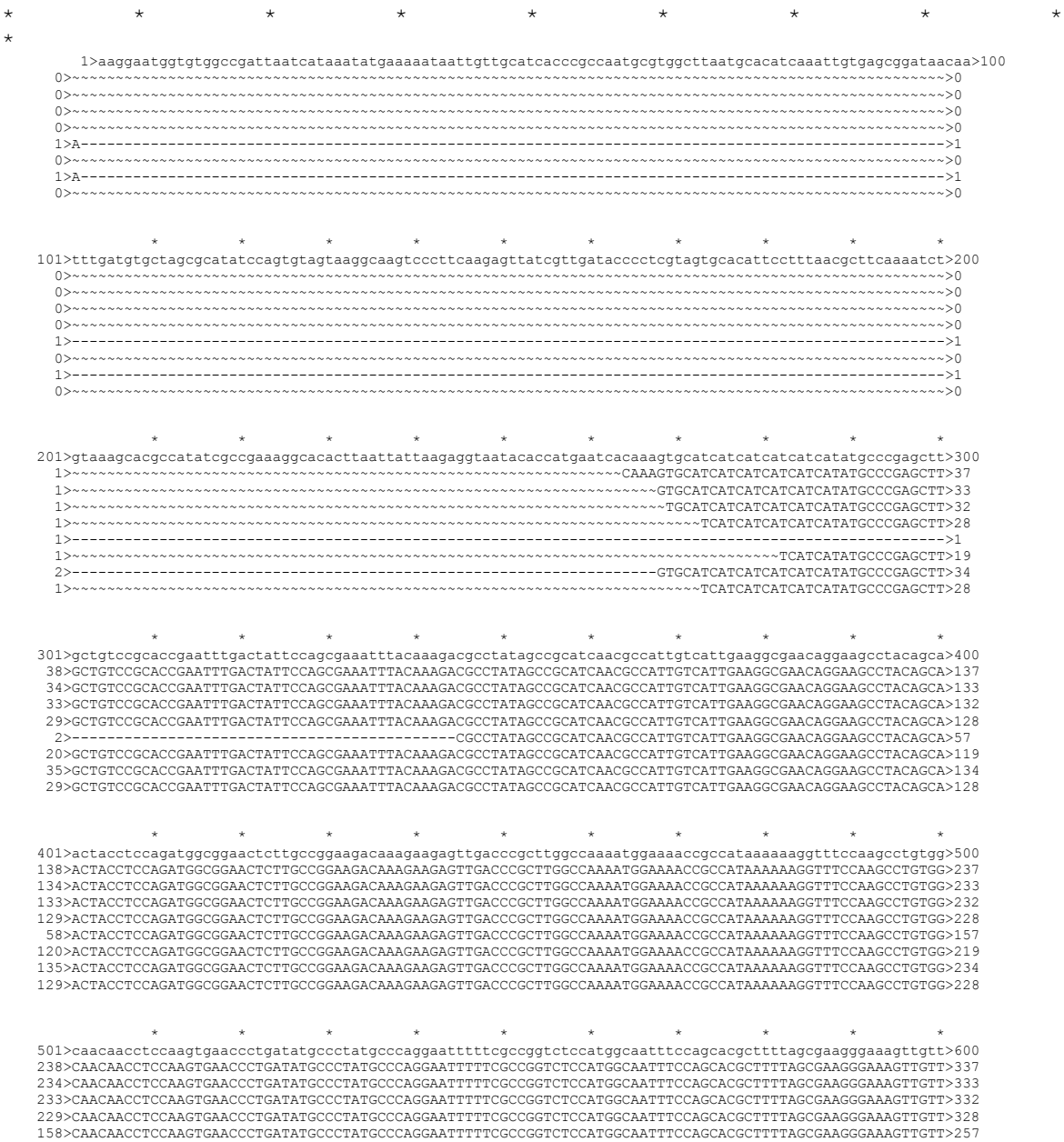
Translation 23 amino acids (MW=2697.36)

```
  1  K  E  W  C  G  R  L  I  I  N  M  K  N  N  C  C  I  T  R  Q  C  V  A  *  C
  1  aaggaatggtgtggccgattaatcataaatatgaaaaataattgttgcatcacccgccaatgcgtggcttaatgc

 26  T  S  N  C  E  R  I  T  I  *  C  A  S  A  Y  P  V  *  *  G  K  S  L  Q  E
 76  acatcaaattgtgagcggataacaatttgatgtgctagcgcatatccagtgtagtaaggcaagtcccttcaagag

 51  L  S  L  I  P  L  V  V  H  I  P  L  T  L  Q  N  L  *  S  T  P  Y  R  R  K
151  ttatcgttgataccctcgtagtgcacattcctttaacgcttcaaaatctgtaaagcacgccatatcgccgaaag

 76  A  H  L  I  I  K  R  *  Y  T  M  N  H  K  V  H  H  H  H  H  H  M  P  E  L
226  gcacacttaattattaagaggtaatacaccatgaatcacaaagtgcatcatcatcatcatcatatgCCCGAGCTT

101  A  V  R  T  E  F  D  Y  S  S  E  I  Y  K  D  A  Y  S  R  I  N  A  I  V  I
301  GCTGTCCGCACCGAATTTGACTATTCCAGCGAAATTTACAAAGACGCCTATAGCCGCATCAACGCCATTGTCATT

126  E  G  E  Q  E  A  Y  S  N  Y  L  Q  M  A  E  L  L  P  E  D  K  E  E  L  T
376  GAAGGCGAACAGGAAGCCTACAGCAACTACCTCCAGATGGCGGAACTCTTGCCGGAAGACAAAGAAGAGTTGACC

151  R  L  A  K  M  E  N  R  H  K  K  G  F  Q  A  C  G  N  N  L  Q  V  N  P  D
451  CGCTTGGCCAAAATGGAAAACCGCCATAAAAAAGGTTTCCAAGCCTGTGGCAACAACCTCCAAGTGAACCCTGAT

176  M  P  Y  A  Q  E  F  F  A  G  L  H  G  N  F  Q  H  A  F  S  E  G  K  V  V
526  ATGCCCTATGCCCAGGAATTTTTCGCCGGTCTCCATGGCAATTTCCAGCACGCTTTTAGCGAAGGGAAAGTTGTT

201  T  C  L  L  I  Q  A  L  I  I  E  A  F  A  I  A  A  Y  N  I  Y  I  P  V  A
601  ACCTGTTTATTGATCCAGGCTTTGATTATCGAAGCTTTTGCGATCGCCGCCTATAACATATATATCCCTGTGGCG

226  D  D  F  A  R  K  I  T  E  G  V  V  K  D  E  Y  T  H  L  N  Y  G  E  E  W
676  GACGACTTTGCTCGGAAAATCACTGAGGGCGTAGTCAAGGACGAATACACCCACCTCAACTACGGGGAAGAATGG

251  L  K  A  N  F  A  T  A  K  E  E  L  E  Q  A  N  K  E  N  L  P  L  V  W  K
751  CTAAAGGCCAACTTTGCCACCGCTAAGGAAGAACTGGAGCAGGCCAACAAAGAAAACCTACCCTTAGTGTGGAAA

276  M  L  N  Q  V  Q  G  D  A  K  V  L  G  M  E  K  E  A  L  V  E  D  F  M  I
826  ATGCTCAACCAAGTGCAGGGGGACGCCAAGGTATTGGGCATGGAAAAAGAAGCCCTAGTGGAAGATTTTATGATC

301  S  Y  G  E  A  L  S  N  I  G  F  S  T  R  E  I  M  R  M  S  S  Y  G  L  A
901  AGCTACGGCGAAGCCCTCAGTAACATCGGCTTCAGCACCAGGGAAATTATGCGTATGTCTTCCTACGGTTTGGCC

326  G  V  *  E  F  K  L  V  D  L  Q  S  R  *  V  I  S  A  *  K  H  R  I  *  D
976  GGAGTCTAGgaattcaagcttgtcgacctgcagtctagataggtaatctctgcttaaaagcacagaatctaagat
```

9.7 Alignment of SynADC_pCOLD and sequencing data from mutagenesis

(Confirmation of mutagenesis)

```
SynADC_pCOLDII from 1 to 5061
Alignment to
D143A1  Matches:927; Mismatches:0; Gaps:4136; Unattempted:0
D143A2  Matches:814; Mismatches:0; Gaps:4248; Unattempted:0
D143A3  Matches:902; Mismatches:0; Gaps:4160; Unattempted:0
D143N1  Matches:913; Mismatches:1; Gaps:4147; Unattempted:0
D143N2  Matches:658; Mismatches:2; Gaps:4401; Unattempted:0
D143N3  Matches:897; Mismatches:1; Gaps:4164; Unattempted:0
D143L1  Matches:760; Mismatches:0; Gaps:4303; Unattempted:0
D143L3  Matches:811; Mismatches:0; Gaps:4252; Unattempted:0
```



```
220>CAACAACCTCCAAGTGAACCCTGATATGCCCTATGCCCAGGAATTTTTCGCCGGTCTCCATGGCAATTTCCAGCACGCTTTTAGCGAAGGGAAAGTTGTT>319
235>CAACAACCTCCAAGTGAACCCTGATATGCCCTATGCCCAGGAATTTTTCGCCGGTCTCCATGGCAATTTCCAGCACGCTTTTAGCGAAGGGAAAGTTGTT>334
229>CAACAACCTCCAAGTGAACCCTGATATGCCCTATGCCCAGGAATTTTTCGCCGGTCTCCATGGCAATTTCCAGCACGCTTTTAGCGAAGGGAAAGTTGTT>328


           *         *         *         *         *         *         *         *         *         *
601>acctgtttattgatccaggctttgattatcgaagcttttgcgatcgccgcctataacatatatatccctgtggcggacgactttgctcggaaaatcactg>700
338>ACCTGTTTATTGATCCAGGCTTTGATTATCGAAGCTTTTGCGATCGCCGCCTATAACATATATATCCCTGTGGCGGACGACTTTGCTCGGAAAATCACTG>437
334>ACCTGTTTATTGATCCAGGCTTTGATTATCGAAGCTTTTGCGATCGCCGCCTATAACATATATATCCCTGTGGCGGACGACTTTGCTCGGAAAATCACTG>433
333>ACCTGTTTATTGATCCAGGCTTTGATTATCGAAGCTTTTGCGATCGCCGCCTATAACATATATATCCCTGTGGCGGACGACTTTGCTCGGAAAATCACTG>432
329>ACCTGTTTATTGATCCAGGCTTTGATTATCGAAGCTTTTGCGATCGCCGCCTATAACATATATATCCCTGTGGCGGACGACTTTGCTCGGAAAATCACTG>428
258>ACCTGTTTATTGATCCAGGCTTTGATTATCGAAGCTTTTGCGATCGCCGCCTATAACATATATATCCCTGTGGCGGACGACTTTGCTCGGAAAATCACTG>357
320>ACCTGTTTATTGATCCAGGCTTTGATTATCGAAGCTTTTGCGATCGCCGCCTATAACATATATATCCCTGTGGCGGACGACTTTGCTCGGAAAATCACTG>419
335>ACCTGTTTATTGATCCAGGCTTTGATTATCGAAGCTTTTGCGATCGCCGCCTATAACATATATATCCCTGTGGCGGACGACTTTGCTCGGAAAATCACTG>434
329>ACCTGTTTATTGATCCAGGCTTTGATTATCGAAGCTTTTGCGATCGCCGCCTATAACATATATATCCCTGTGGCGGACGACTTTGCTCGGAAAATCACTG>428


           *         *         *         *         *         *         *         *         *
701>agggcgtagtcaaggac--gaatacacccacctcaactacggggaagaatggctaaaggccaactttgccaccgctaaggaagaactggagcaggccaac>798
438>AGGGCGTAGTCAAGG-CG-GAATACACCCACCTCAACTACGGGGAAGAATGGCTAAAGGCCAACTTTGCCACCGCTAAGGAAGAACTGGAGCAGGCCAAC>535
434>AGGGCGTAGTCAAGG-CG-GAATACACCCACCTCAACTACGGGGAAGAATGGCTAAAGGCCAACTTTGCCACCGCTAAGGAAGAACTGGAGCAGGCCAAC>531
433>AGGGCGTAGTCAAGG-CG-GAATACACCCACCTCAACTACGGGGAAGAATGGCTAAAGGCCAACTTTGCCACCGCTAAGGAAGAACTGGAGCAGGCCAAC>530
429>AGGGCGTAGTCAAGAAC--GAATACACCCACCTCAACTACGGGGAAGAATGGCTAAAGGCCAACTTTGCCACCGCTAAGGAAGAACTGGAGCAGGCCAAC>526
358>AGGGCGTAGTCAAGAAC--GAATACACCCACCTCAACTACGGGGAAGAATGGCTAAAGGCCAACTTTGCCACCGCTAAGGAAGAACTGGAGCAGGCCAAC>455
420>AGGGCGTAGTCAAGAAC--GAATACACCCACCTCAACTACGGGGAAGAATGGCTAAAGGCCAACTTTGCCACCGCTAAGGAAGAACTGGAGCAGGCCAAC>517
435>AGGGCGTAGTCAAG--CTGGAATACACCCACCTCAACTACGGGGAAGAATGGCTAAAGGCCAACTTTGCCACCGCTAAGGAAGAACTGGAGCAGGCCAAC>532
429>AGGGCGTAGTCAAG--CTGGAATACACCCACCTCAACTACGGGGAAGAATGGCTAAAGGCCAACTTTGCCACCGCTAAGGAAGAACTGGAGCAGGCCAAC>526


    *         *         *         *         *         *         *         *         *         *
799>aaagaaaacctacccttagtgtggaaaatgctcaaccaagtgcaggggacgccaaggtattgggcatggaaaaagaagccctagtggaagattttatga>898
536>AAAGAAAACCTACCCTTAGTGTGGAAAATGCTCAACCAAGTGCAGGGGGACGCCAAGGTATTGGGCATGGAAAAAGAAGCCCTAGTGGAAGATTTTATGA>635
532>AAAGAAAACCTACCCTTAGTGTGGAAAATGCTCAACCAAGTGCAGGGGGACGCCAAGGTATTGGGCATGGAAAAAGAAGCCCTAGTGGAAGATTTTATGA>631
531>AAAGAAAACCTACCCTTAGTGTGGAAAATGCTCAACCAAGTGCAGGGGGACGCCAAGGTATTGGGCATGGAAAAAGAAGCCCTAGTGGAAGATTTTATGA>630
527>AAAGAAAACCTACCCTTAGTGTGGAAAATGCTCAACCAAGTGCAGGGGGACGCCAAGGTATTGGGCATGGAAAAAGAAGCCCTAGTGGAAGATTTTATGA>626
456>AAAGAAAACCTACCCTTAGTGTGGAAAATGCTCAACCAAGTGCAGGGGGACGCCAAGGTATTGGGCATGGAAAAAGAAGCCCTAGTGGAAGATTTTATGA>555
518>AAAGAAAACCTACCCTTAGTGTGGAAAATGCTCAACCAAGTGCAGGGGGACGCCAAGGTATTGGGCATGGAAAAAGAAGCCCTAGTGGAAGATTTTATGA>617
533>AAAGAAAACCTACCCTTAGTGTGGAAAATGCTCAACCAAGTGCAGGGGGACGCCAAGGTATTGGGCATGGAAAAAGAAGCCCTAGTGGAAGATTTTATGA>632
527>AAAGAAAACCTACCCTTAGTGTGGAAAATGCTCAACCAAGTGCAGGGGGACGCCAAGGTATTGGGCATGGAAAAAGAAGCCCTAGTGGAAGATTTTATGA>626


    *         *         *         *         *         *         *         *         *         *
899>tcagctacggcgaagccctcagtaacatcggcttcagcaccagggaaattatgcgtatgtcttcctacggtttggccggagtctaggaattcaagcttgt>998
636>TCAGCTACGGCGAAGCCCTCAGTAACATCGGCTTCAGCACCAGGGAAATTATGCGTATGTCTTCCTACGGTTTGGCCGGAGTCTAGGAATTCAAGCTTGT>735
632>TCAGCTACGGCGAAGCCCTCAGTAACATCGGCTTCAGCACCAGGGAAATTATGCGTATGTCTTCCTACGGTTTGGCCGGAGTCTAGGAATTCAAGCTTGT>731
631>TCAGCTACGGCGAAGCCCTCAGTAACATCGGCTTCAGCACCAGGGAAATTATGCGTATGTCTTCCTACGGTTTGGCCGGAGTCTAGGAATTCAAGCTTGT>730
627>TCAGCTACGGCGAAGCCCTCAGTAACATCGGCTTCAGCACCAGGGAAATTATGCGTATGTCTTCCTACGGTTTGGCCGGAGTCTAGGAATTCAAGCTTGT>726
556>TCAGCTACGGCGAAGCCCTCAGTAACATCGGCTTCAGCACCANGGAAATTATGCGTATGTCTTCCTACGGTTTGGCCGGAGTCTAGGAATTCAAGCTTGT>655
618>TCAGCTACGGCGAAGCCCTCAGTAACATCGGCTTCAGCACCAGGGAAATTATGCGTATGTCTTCCTACGGTTTGGCCGGAGTCTAGGAATTCAAGCTTGT>717
633>TCAGCTACGGCGAAGCCCTCAGTAACATCGGCTTCAGCACCAGGGAAATTATGCGTATGTCTTCCTACGGTTTGGCCGGAGTCTAGGAATTCAAGCTTGT>732
627>TCAGCTACGGCGAAGCCCTCAGTAACATCGGCTTCAGCACCAGGGAAATTATGCGTATGTCTTCCTACGGTTTGGCCGGAGTCTAGGAATTCAAGCTTGT>726


    *         *         *         *         *         *         *         *         *         *
999>cgacctgcagtctagataggtaatctctgcttaaaagcacagaatctaagatccctgccatttggcggggatttttttt-atttgttttcaggaaataaat>1097
736>CGACCTGCAGTCTAGATAGGTAATCTCTGCTTAAAAGCACAGAATCTAAGATCCCTGCCATTTGGCGGGGATTTTTTTTATTTGTTTTCAGGAAATAAAT>835
732>CGACCTGCAGTCTAGATAGGTAATCTCTGCTTAAAAGCACAGAATCTAAGATCCCTGCCATTTGGCGGGGATTTTTTT-ATTTGT~~~~~~~~~~~~~~~>815
731>CGACCTGCAGTCTAGATAGGTAATCTCTGCTTAAAAGCACAGAATCTAAGATCCCTGCCATTTGGCGGGGATTTTTTT-ATTTGTTTTCAGGAAATAAAT>829
727>CGACCTGCAGTCTAGATAGGTAATCTCTGCTTAAAAGCACAGAATCTAAGATCCCTGCCATTTGGCGGGGATTTTTTT-ATTTGTTTTCAGGAAATAAAT>825
656>CGACC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>660
718>CGACCTGCAGTCTAGATAGGTAATCTCTGCTTAAAAGCACAGAATCTAAGATCCCTGCCATTTGGCGGGGATTTTTTTTATTTGTTTTCAGGAAATAAAT>817
733>CGACCTGCAGTCTAGATAGGTAATCTCTGC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>762
727>CGACCTGCAGTCTAGATAGGTAATCTCTGCTTAAAAGCACAGAATCTAAGATCCCTGCCATTTGGCGGGGATTTTTTT--T-------------AT---T>808


     *         *         *         *         *         *         *         *         *         *
1098>aatcgatcgcgtaataaaatctattattatttttgtgaagaataaatttgggtgcaatgagaatgcgcaggccctttcgtctcgcgcgtttcggtgatga>1197
836>AATCGATCGCGTAATAAAATCTATTATTATTTTTTGTGAAGAATAAATTTGGGTGCAATGAGAATGCGCAGGCCCTTTCGTCTCGCGCGTTTCGG~~~~~~>929
815>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>815
830>AATCGATCGCGTAATAAAATCTATTATTATTTTTTGTGAAGAATAAATTTGGGTGCAATGAGAATGCGCAGGCCC~~~~~~~~~~~~~~~~~~~~~~~>903
826>AATCGATCGCGTAATAAAATCTATTATTATTTTTTGTGAAGAATAAATTTGGGTGCAATGAGAATGCGCAGGCCCTTTCGTCTCGCGCGT~~~~~~~~~~~>914
660>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>660
818>AATCGATCGCGTAATAAAATCTATTATTATTTTTTGTGAAGAATAAATTTGGGTGCAATGAGAATGCGCAGGCCCTTTCGTCT~~~~~~~~~~~~~~~~~>899
762>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>762
809>--T-G-T----T--T~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>813
```

Reference List

Aarts MGM, Keijzer CJ, Stiekema WJ, and Pereira A (1995). Molecular characterization of the CER1 gene of Arabidopsis involved in epicuticular wax biosynthesis and pollen fertility. *Plant Cell* **7**: 2115 - 2127.

Albro PW, and Dittmer JC (1970). Bacterial hydrocarbons: occurrence, structure and metabolism. *Lipids* **5**: 320 - 325

Altschul, S.F, Gish, W., Miller, W., Myers, E.W. and Lipman, D.J, (1990). Basic Local alignment search tool. *J. Mol. Bio*. **215**: 339 - 410

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

Andreas Schirmer, Mathew A. Rude, Xuezhi Li, Emanuela Povova, Stephen B. Del Cordayre (2010). Microbial Biosynthesis of Alkanes. Science **329**: 559

Bagaeva, T.V (1998). Sulfate-reducing bacteria, hydrocarbon producers. Thesis Doctoral *(Biol*) Dissertation. Russia: Kazan State University

Barton, G.J, (1996). Protein sequence alignment and database screening in protein structure predication Oxford University Press, Oxford

Belyaeva, M.I, Zolotukhina, L.M and Bagaeva, T.V. (1995). Method for the production of liquid hydrocarbons. *Invention Certificate*.

Berman, H.M., Westbrook, J., Feng, Z,Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000). The protein Data Bank. *Nucleic Acids Research,* **28**: 235 - 242.

Bradford. M.M (1976). A rapid and sensitive method for the quantitation of microgram

quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*. **72**:248 - 54

Bragg,W.L, (1913). The structure of crystals as indicated by their X-ray diffraction patterns. *Proc. Roy. Soc.* **A89**: 248 - 277.

Brocks JJ, Buick R, Summons RE and Logan GA (2003). A reconstruction of Archean biological diversity based on molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Super group. *Geochim Cosmochim Acta*; **67**: 4321 - 35.

Brodsky, L.I., Ivanov, V.V., Kalai, D., Ya.L., Leontovich, A.M., Nikolaev, V.K., Feranchuk, S.I and Drachev V.A. (1995). GeneBee-NET: Internet-based server for analyzing biopolymers structures, *Biochemistry*, **60**: 923 - 928.

Brodsky, L.I., Vasiliev, A.V., Kalaidzidis, Ya.L., Osipov, Yu.S., Tatuzov, R.L. and Feranchuk, S.I. (1992). GeneBee: the program package for biopolymer structure analysis. *Dimacs*, **8**: 127 - 139.

Brumshtein, B., H.Greenblat, A.Futerman, I.Silman, and J.Sussman., (2008). Control of the rate of evapouration in protein crystallisation by the “microbatch under oil” method. *J.Appl. Cryst.* **41**: 969 - 971.

Buist. P. H (2007). Exotic biomodification of fatty acids. *Nat Prod Rep*. **5**: 1110 - 27

Chayen, N.E, P.D. Shaw Stewart, and D.M.Blow., (1992). Microbatch crystallisation under oil - a new technique allowing many small-volume crystallisation trials. *Journal of Crystal Growth*, **122**: 176 - 180.

Cheesbrough T.M and P. E. Kolattukudy (1984). Alkane biosynthesis by decarbonylation of aldehydes catalyzed by a particulate preparation from Pisum sativum. *Biochem,* **21**: 6613 - 6617.

David L. Nelson, Michael M. Cox, 5th edition “lehninger Principles of Biochemistry, W.H Freeman Publishers.

Debasis Das, Bekir E. Eser, Jaehong Han, Aaron Sciore and Neil G. Marsh, (2011). Oxygen-independent decarbonylation of aldehydes by Cyano-bacterial aldehyde decarbonylase: a new reaction of di-iron enzymes. *Angewandte chemie*. **123**: 7286 - 7290.

Dennis. M and P. E. Kolattukudy (1992). A cobalt-porphyrin enzyme converts a fatty aldehyde to a hydrocarbon and CO. *Biochem*. **12**: 5306 - 5310.

Douglas M. Warui, Ning Li, Hanne Norgaard, Carsten Kreb, J. Martin Bollinger, Jr, and Squire J. Booker, (2011). Detection of Formate, Rather than Carbon Monoxide, As the Stoichiometric Co-product in Conversion of Fatty Aldehydes to Alkanes by a Cyanobacterial Aldehyde Decarbonylase *J. Am. Chem. Soc.* **10** 3316 - 3331

Esnouf RM (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J Mol Graph*, **15**:133 -138

Fukuda, H., Ogawa, T and Fujii, T (1987). Method for producing hydrocarbon mixtures. *US Patent* **4:** 698 - 304.

Gill S.C, P. H. von Hippel (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*, **182**: 319 - 326.

Guengerich FP (2008). "Cytochrome p450 and chemical toxicology". *Chem. Res. Toxicol.* **1**: 70 - 83.

Hall, T. (1999) Bioedit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic acids symposia*, **41**: 95 - 98.

Higgins, D., Thompson, J., Gibson, T., Thompson, J.D, Higgins, D.G. and Gibson T.J (1994) CLUSTAL W: improving the sensitive of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. **22**: 4673 - 4680.

Jackson L.L and Blomquist G.J (1976). Cuticicular lipids of insects. *Biochem* **11**: 77 - 79

Jones JG (1969). Studies on lipids of soil microorganisms with particular reference to hydrocarbons. *J Gen Microbiol* **59**:145 - 52.

Julie A. Kovacs,(2004). Synthetic Analogues of Cysteinate-Ligated Non-Heme Iron and Non-Corrinoid Cobalt Enzymes. *Chem. Rev* **104**: 825 - 848.

Karlsson, A.; Parales, J. V.; Parales, R. E.; Gibson, D.T.;Eklund,H.; Ramaswamy, S. (2003). *Science* **299**: 1039.
Kolattukudy P.E, Croteau R. and Buckner J.S (1976). Biochemistry of plant waxes. In *Biochemistry of Natural waxes*. Pp 289 - 347.

Kunst. L, A.L. and Samuels (2003). Biosynthesis and secretion of plant cuticular wax. *Plant Physiology*. 42: 51 - 58.

Ladygina. N, E.G. Dedyukhina and M.B. Vainshtein (2006). A review on microbial synthesis of hydrocarbons. *Process Biochemistry* **41**: 1001 - 1014.

Laemmli UK (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4". *Nature* **227** (5259): 680 - 685.

McNicholas . S, E. Potterton, K.S. Wilson and M.E. Noble (2011). The CCP4MG Molecular-graphics software. *Acta Cryst* **D67**: 386 - 394.

Merritt E. A and Wever, R. (1997) Raster 3D Version 2: photorealistic molecular graphics. *Methods in Enzymeology*, **277**: 505 - 524.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbour Symposia in Quantitative Biology*. **51**: 263 - 73.

Naccarato WF, Gilbertson JR, and Gelman RA (1974). Effects of different

culture media and oxygen upon lipids of Escherichia coli K-12. *Lipids* **9**: 322 - 9.

Paul Brian Charles James, University of Exeter PhD thesis (2010). Investigation into peroxiredoxin and interactions in the peroxiredoxin peroxide scavenging system

Paul Emsley, Bernhard Lohkamp, William G. Scott, and Kevin Cowtan (2010). Features and Development of Coot. *Acta Crystallographica Section D - Biological Crystallography.* **66**: 486 - 501.

Reed JR, Quilici DR, Blomquist GJ, Reitz RC, (1995). Proposed mechanism for cytochrome P450-catalyzed conversion of aldehyde to hydrocarbons in house fly, *musca domestica*. *Biochemistry*, **34**: 16221 - 16227

Richard I Gumport, Roger E Koeppe, Lubert Stryer (1995). Student companion for Stryer's Biochemistry, New York: W.H. Freeman: 4th Edition

Rohde, J-U.; In, J.-H.; Lim, M. H.; Brennessel, W. W.; Bukowski M. R.; Stubna, A.; Munck, E.; Nam, W.; Que, L.,(2003). Jr. *Science* **299**, 1037.

S. J. Lippard, J. M. Berg "Principles of Bioinorganic Chemistry" University Science Books: Mill Valley, CA; 1994.

Schirmer A, Rude MA, LiX, Popova E, del Cardayre SB (2010). Microbial biosynthesis of alkanes. Science. **329**: 559 - 562.

Schuler, M.A. and Werck-Reichhart, D. (2003). Functional Genomics of P450s. *Annu*. *Rev*. *Plant Biol*. **54**: 629-637.

Stone and Zobell, (1952). Biodegradation and Bioremediation (Metabolism of alkylbenzens, alkanes and other hydrocarbons in anaerobic bacteria)

Stookey .L. L (1970). Ferrozine-A New Spectrophotometric Reagent for Iron *Analytical Chemistry*, **42**: 779 - 781.

Stubbe J. and P. Riggs-Gelasco (1998). Harnessing free radicals: formation and function of the tyrosyl radical in ribonucleotide reductase *Trends Biochem*. *Sci*, **23**: 438.

Summons RE, Jahnke LL, Hope JM and Logan GA (1999). 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature*; **400**: 554 - 7.

Tanford C 1964. Cohesive forces and disruptive reagents. *Brookhaven Symp Biol*. **17**: 154 -183.

Terese Bergfors (2003). 'Seeds to Crystals'. *Journal of Structural Biology*, **142**: 66 - 76.

Terron, M.C, Verhagen, F.J.M, Franssen, M.C.R., Field, J.A, (1998). Chemical bromination of phenol red by hydrogen peroxide is possible in the absence of haloperoxidases, *Chemosphere* **36**: 1445-1452.

Thompson, J.D, Higgins, D.G and Gibson, T.J, (1994). ClustalW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids*. **22**: 4673-4680.

Tornabene TG, Morrison SJ, and Kloos WE (1970). Aliphatic hydrocarbon contents of various members of the family *Micrococcaceae*. *Lipids* **5**: 929 - 34.

Unpublished, structure solved by Joint Center of Structural Genomics (PDB 2OC5A)
Warburg, O, and W. Christian (1941). Isolation and crystallization of enolase. Biochem **310**: 384 - 421.