Structural Studies of *N. maritimus* Acetyl-CoA/Propionyl-CoA Carboxylase with X-Ray Crystallography Approach

by

Sabri Özkan Besler

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Molecular Biology and Genetics

October 4, 2021

Structural Studies of *N. maritimus* Acetyl-CoA/Propionyl-CoA Carboxylase with X-Ray Crystallography Approach

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Sabri Özkan Besler

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Assist. Prof. Hasan Demirci (Advisor)

_____

Assist. Prof. Serkan Kır

_____

Assist. Prof. Şeref Gül

Date: 04/10/2021

To my beloved family

ABSTRACT

Structural Studies of *N. maritimus* Acetyl-CoA/Propionyl-CoA Carboxylase with X-Ray Crystallography Approach

Sabri Özkan Besler

Master of Science in Molecular Biology and Genetics

October 4, 2021

The Archaea domain is among one of the most broadly distributed prokaryotic life forms on our planet. Most of the members of this domain are known to live in extreme environments and are involved in the fixation of inorganic compounds into organic compounds. Many ammonia-oxidizing archaea belonging to the phyla Crenarchaeota and Thaumarchaeota were found to assimilate inorganic carbon into organic carbon through a carbon fixation pathway called 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle. Interestingly, the marine species of the phylum Thaumarcheota were found to use a modified version of the HP/HB cycle. Among these organisms, the species *N. maritimus* was discovered to use the most energy-efficient carbon fixation mechanism under aerobic conditions. This feature provides *N. maritimus* several advantages to thrive in extreme oligotrophic environments. In the HP/HB cycle, the carboxylation of acetyl-CoA and propionyl-CoA is of great importance as its products are essential precursors for the maintenance of this cycle. Importantly, in *N. maritimus*, acetyl-CoA and propionyl-CoA are carboxylated by a biotin-dependent single bifunctional enzyme called Acetyl-CoA/Propionyl-CoA carboxylase. This enzyme is composed of three subunits: a carboxyltransferase Nmar_0272, carbamoyl phosphate synthase Nmar_0273, and a biotin-lipoyl attachment domain-containing protein Nmar_0274. Together, these subunits form a protein complex to become functional. In this study, we have purified and crystallized Nmar_0272 and Nmar_0274 subunits and determined the first ever structure of carboxyltransferase subunit Nmar_0272. The structure of Nmar_0272 exhibited an apo form dimer structure at 2.26 Å resolution and in P63 space group symmetry. Additionally, certain undetermined regions were reconstructed using homology and AI-based tools. Overall, our study presents a novel protein structure and serves as a starting point for elucidating the carbon fixation mechanisms in marine Thaumarchaeota.

ÖZETÇE

N. maritimus Asetil-CoA/Propiyonil-CoA Karboksilaz Enziminin X-Işını Kristalografisi ile Yapısal Çalışmaları

Sabri Özkan Besler

Moleküler Biyoloji ve Genetik, Yüksek Lisans

4 Ekim 2021

Arke alemi, gezegenimizde en geniş dağılıma sahip prokaryotik yaşam formlarını içinde barındırır. Bu alemin üyelerinin çoğu ekstrem ortamlarda yaşayarak inorganik bileşikleri organik bileşiklere dönüştürmeleri açısından büyük öneme sahiptirler. Crenarchaeota ve Thaumarchaeota şubelerine ait birçok amonyak oksitleyici arkenin, 3-hidroksipropionat/4-hidroksibutirat (HP/HB) döngüsü adı verilen bir karbon sabitleme yolağı kullanarak inorganik karbonu organik karbona asimile ettikleri bilinmektedir. Bunların yanı sıra, Thaumarcheota şubesinin denizlerde yaşayan üyelerinin HP/HB döngüsünün değiştirilmiş bir versiyonunu kullandıkları anlaşılmıştır. Bu organizmalar arasında *N. maritimus* türünün aerobik koşullar altında enerji açısından en verimli karbon sabitleme mekanizmasını kullandığı keşfedilmiştir. Bu özellik, N. maritimus'a aşırı oligotrofik ortamlarda yaşayabilmesi için çeşitli avantajlar sağlar. HP/HB döngüsünde asetil-CoA ve propiyonil-CoA moleküllerinin karboksilasyonu, bu döngüdeki öncü moleküllerin üretilmesini sağladığı için, döngünün devamlılığının sağlanmasında önemli bir yere sahiptir. Önemli olarak, *N. maritimus* asetil-CoA ve propiyonil-CoA moleküllerinin her ikisini de yalnızca bir adet çift işlevli asetil-CoA/propiyonil-CoA karboksilaz enzimi kullanarak karboksiller. Bu enzim karboksiltransferaz Nmar_0272, karbomoil fosfat sentaz Nmar_0273 ve biyotin-lipoil bağlanma alanı içeren protein Nmar_0274 olmak üzere üç adet alt birimden oluşur. Bu üç alt birim bir araya gelerek işlevsel bir protein kompleksi oluştururlar. Bu çalışmada Nmar_0272 ve Nmar_0274 alt birimlerinin saflaştırma ve kristalizasyonu yapılmış; bununla birlikte karboksiltransferaz alt birimi Nmar_0272 proteininin ilk üç boyutlu yapısı belirlenmiştir. Elde ettiğimiz bu yapı 2.26 Å çözünürlükte ve P63 uzay grubu simetrisinde bir apo-form dimer yapısı sergilemektedir. Bunlara ek olarak, modelimizdeki belirlenmemiş bölgeler, homoloji modelleme ve yapay zeka tabanlı programlar kullanılarak yeniden oluşturulmuştur. Bu çalışmamız yeni bir protein yapısı ortaya koymakla birlikte, deniz Thaumarcheota'larının karbon sabitleme mekanizmalarının daha iyi anlaşılmasına ışık tutmayı hedeflemektedir.

ACKNOWLEDGEMENTS

First of all, I would like to express my deep gratitude to my advisor Asst. Prof. Hasan Demirci, for giving me the opportunity to work in KUYBIIG-M and providing us with an outstanding scientific environment. It was a great pleasure to be one of his first students and study under the supervision of such an esteemed scientist. He always advised me to write a thesis that I can be proud to read even in the future; therefore, I tried to write my thesis worthy of his recommendation. I also would like to express my heartfelt appreciation to Professor Soichi Wakatsuki for his very valuable scientific guidance and help during my thesis preparation process. Throughout my studies, I have developed a lot thanks to their priceless support and feedback.

I would particularly like to thank my thesis jury members Asst. Prof. Serkan Kır, Asst. Prof Şeref Gül and Dr. Burak Veli Kabasakal for spending their precious time on my thesis defense process and their valued feedback. I would like to offer my special thanks to all the Graduate School of Sciences and Engineering (GSSE) and Department of Molecular Biology and Genetics members for their huge assistance. I am also very grateful to TUBITAK 2232 International Fellowship for Outstanding Researchers Program (Project No: 118C270) for providing me financial assistance for my project. It was a great honour to be a student at Koç University and I will proudly carry the privileges of studying in such a prestigious university in every moment of my life.

I would like to give my special regards to the KUYBIIG-M family for their valuable support and collaboration. I would like to especially thank Günseli Yıldırım and owe her a very important debt for her invaluable feedback, assistance, patience, and more importantly, her friendship. She has always been more than a colleague and a lifesaver when I needed support. I am deeply grateful to Betül Ertem, Jerome Johnson and Merve Yığın for their precious companionship and fruitful scientific discussions. I also received very generous support from Ebru Destan during my thesis preparation period. I consider myself very lucky to have such amazing friends and colleagues.

I am indebted to my lovely mother Sibel, my father Cem, my grandmother Semra and all my other beloved family members for their infinite support and love from the beginning of my life. Without them, this thesis would not have been materialized.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

Å	Angstrom
AI	Artificial Intelligence
α	Alpha
β	Beta
γ	Gamma
μg	Microgram
μl	Microliter
ADP	Adenosine Diphosphate
AMP	Adenosine Monophosphate
AOA	Ammonia Oxidizing Archaea
AOB	Ammonia Oxidizing Bacteria
ATP	Adenosine Triphosphate
BC	Biotin Carboxylase
BCCP	Biotin Carboxyl Carrier Protein
BTN	Biotin
BTI	5-(Hexahydro-2-Oxo-1H-Thieno[3,4-D] Imidazol-6-YL)Pentanal
C	Celsius
CFC	Chlorofluorocarbon
CFeSP	Corrinoid Iron-Sulfur Protein
CH4	Methane
CO	Carbon Monoxide
CO_2	Carbon Dioxide
CoA	Coenzyme A
CODH	Carbon Monoxide Dehydrogenase
CT	Carboxyltransferase
DHAP	Dihydroxyacetone Phosphate
DNA	Deoxyribonucleic Acid
E.coli	*Escherichia coli*
FAD	Flavin Adenine Dinucleotide

FPLC	Fast Protein Liquid Chromatography
FT	Flowthrough
g	Gram
G3P	Glyceraldehyde-3-Phosphate
GAPDH	Glyceraldehyde Phosphate Dehydrogenase
GHG	Greenhouse Gas
HCO_3^-	Bicarbonate
H_4F	Tetrahydrofolate
HFC	Hydrofluorocarbon
His-tag	Histidine tag
HP/HB	3-Hydroxypropionate/4-Hydroxybutyrate
IPTG	Isopropyl-β-D-thiogalactoside
K	Kelvin
kDa	Kilodalton
K_m	Michaelis Constant
L	Liter
LB	Lysogeny Broth
mg	Milligram
min	Minute
ml	Milliliter
M	Molar
mM	Millimolar
MMC	(S)-Malyl-CoA/ß-methylmalyl-CoA/(S)-citramalyl-CoA
NaCl	Sodium Chloride
NAD	Nicotinamide Adenine Dinucleotide
N_2O	Nitrous oxide
NO_2	Nitrogen dioxide
Ni-NTA	Nickel-Nitrilotriacetic Acid
N. maritimus	*Nitrosopumilus maritimus*
Nmar	*Nitrosopumilus maritimus* related protein
Nmar_0272	*Nitrosopumilus maritimus* carboxyltransferase subunit
Nmar_0273	*Nitrosopumilus maritimus* carbamoyl phosphate synthase subunit
Nmar_0274	*Nitrosopumilus maritimus* biotin-lipoyl attachment domain protein

OD	Optical Density
PDB	Protein Data Bank
PEP	Phosphoenolpyruvate
PFC	Perfluorocarbons
PGA	Phosphoglycerates
R5P	Ribose-5-Phosphate
RMSD	Root Mean Square Deviation
Ru5P	Ribulose-5-phosphate
RuBisCO	Ribulose-1,5-Bisphosphate Carboxylase-Oxygenase
RuBP	Ribulose-1,5-Bisphosphate
rpm	Revolutions per minute
SCM1	*Nitrosopumilus maritimus*
S7P	Sedoheptulase-7-Phosphate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SF_6	Sulfur Hexafluoride
SO_2	Sulfur Dioxide
Tris	Tris(hydroxymethyl)aminomethane
Xu5P	Xylulose-5-Phosphate

Chapter 1:
INTRODUCTION

1.1 Carbon Cycle and the Fixation of the Inorganic Carbon

Since the beginning of the industrial era, the emission of greenhouse gases (GHGs), including carbon dioxide (CO_2), nitrous oxide (N_2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), chlorofluorocarbons (CFCs), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), methane (CH_4) and sulfur hexafluoride (SF_6), has escalated significantly (Audi et al., 2019; Abeydeera et al., 2019; Khalil, 1999; Liu et al., 2019). The accumulation of these greenhouse gases in the atmosphere has caused one of today's most serious natural phenomena: global climate change. The emergence of climate change has led to significant imbalances ranging from ecological to economic issues, including global warming, the disruption of the ecological systems, technological and social problems (Liu et al., 2019; Abeydeera et al., 2019). The development of this global issue has piqued the interest of many professionals, resulting in the search for solutions to this global issue.

Among the GHGs, CO_2 was identified as the major cause of global climate change (Johnson & Derwent, 1996). Despite that, nature has evolved robust mechanisms to balance the amount of CO_2. Autotrophic carbon fixation is a natural phenomenon in which inorganic carbon is converted into organic carbon and cellular building materials by autotrophic organisms, mainly through photosynthesis and prokaryotic carbon fixation pathways. The provided organic carbon is consumed by heterotrophic organisms which convert the organic carbon back into inorganic carbon through respiration (Field et al., 1998; Hügler & Sievert, 2011). The decay of the dead organisms by decomposer microorganisms also contributes to the accumulation of the organic carbon in ground layers leading to the formation of fossil fuels in the form of coal, natural gas, and petroleum. These fossil fuels are burned for industrial purposes, resulting in the release of inorganic carbon back to the atmosphere. Additionally, the organic carbon may also be converted into inorganic carbon as a consequence of deforestation (Sellers et al., 2018;

Grace, 2004). The resulting inorganic carbon is fixed through autotrophic CO_2 fixation, maintaining the carbon cycle (**Fig. 1**).

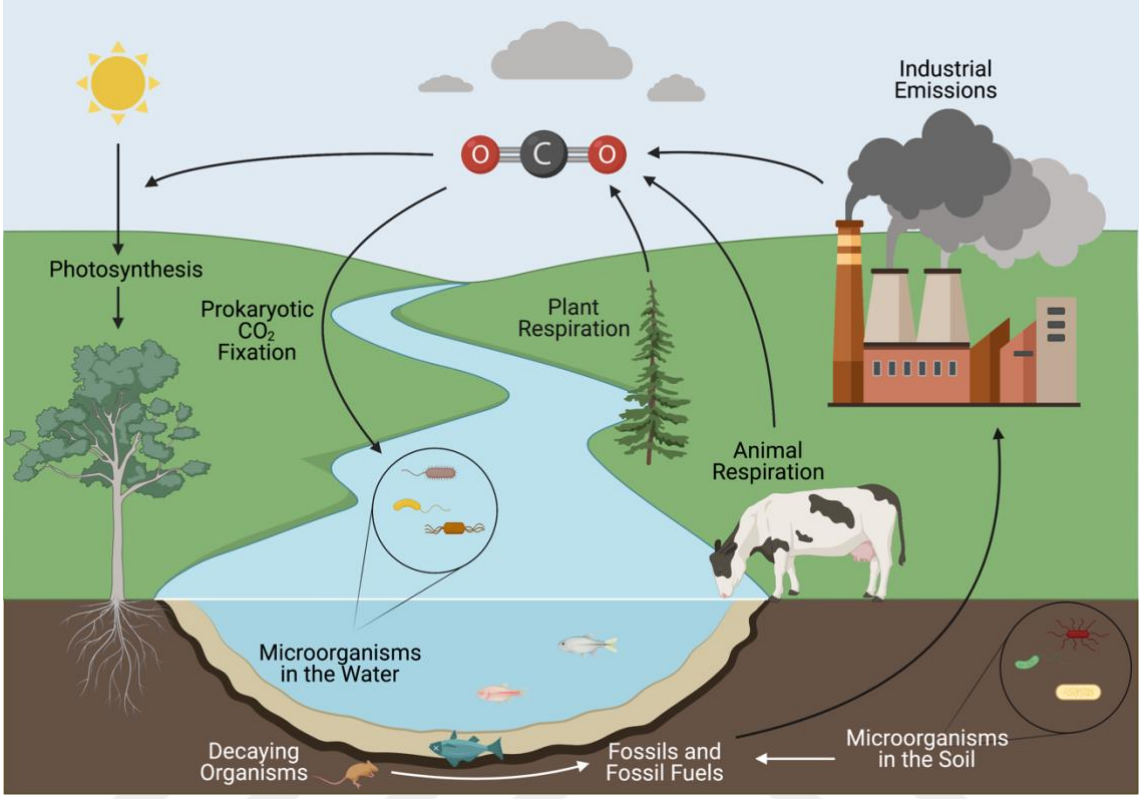


Figure 1.1: Schematic representation of the global carbon cycle. *The inorganic carbon is formed as a consequence of the respiration and combustion of fossil fuels. The emitted CO_2 is then converted into organic carbon compounds through autotrophic carbon fixation mechanisms such as photosynthesis and prokaryotic CO_2 fixation.*

In the 1970s, the domain Archaea was identified as the third domain of life along with the domains Bacteria and Eukarya (Woese & Fox, 1977). Together with the domain Bacteria, Archaea are one of the most widely distributed life forms on our planet. Many members of archaea are adapted to thrive in extreme conditions such as in extremely high or low temperatures, low pH, high salt concentration, and extremely low nutrition. These habitats include hydrothermal vents, geothermal springs, polar zones, and even the human intestine. On the other hand, many species populate in non-extreme habitats such as freshwater sources, soil, and open oceans (Chaban et al., 2006). Many species of archaea are involved in the fixation of inorganic compounds into organic compounds through autotrophic carbon fixation mechanisms (Berg et al., 2010). Up to the present, six autotrophic carbon fixation mechanisms were identified in prokaryotes: 1) reductive pentose phosphate (Calvin-Benson-Basham) cycle, 2) reductive citric acid cycle, 3) reductive acetyl-CoA pathway, 4) dicarboxylate/4-hydroxybutyrate cycle, 5) 3-

hydroxypropionate bicycle, 6) 3-hydroxypropionate/4-hydroxybutyrate cycle (Berg et al., 2010; Fuchs, 2011; Hügler & Sievert, 2011).

1.2 Reductive Pentose Phosphate (Calvin-Benson-Bassham) Cycle

In photosynthetic organisms, the fixation of the inorganic carbon into organic glucose molecules is provided through the reductive pentose phosphate (Calvin-Benson-Basham) cycle. In this pathway, a five carbon-containing carbon acceptor molecule ribulose 1,5 bisphosphate (RuBP) molecule is carboxylated via a carboxylase named RuBisCO. This reaction yields two phosphoglycerates, which are then reduced and used to form a glucose molecule (Bassham & Calvin, 1954) (**Fig. 2**).

In the first step of the cycle, a five carbon-containing ribulose-5-phosphate (Ru5P) is converted into RuBP through an enzyme called phosphoribulokinase using an ATP molecule. RuBP molecule is then carboxylated through RuBisCO (Quayle et al., 1957; Erb & Zarzycki, 2018), yielding two 3-phosphoglycerates (3-PGA). In the reduction step, the 3-PGAs are converted into 1,3-bisphosphoglycerates (1,3-PGA) by the enzyme phosphoglycerate kinase using an ATP. 1,3-PGA is then reduced to glyceraldehyde-3-phosphate (G3P) molecules by glyceraldehyde phosphate dehydrogenase (GAPDH). In the final step, the resulting G3P molecules are used to form glucose molecules or regenerate RuBP molecules to maintain the cycle. On the other hand, to form a glucose molecule and maintain the continuity of the cycle, three CO_2 molecules need to be fixed. This process requires three RuBP molecules. As the result of this reaction, six PGA molecules form and are reduced to six G3P molecules. Only one of the G3P molecule exits the cycle for the production of glucose, and the remaining five G3Ps are used to regenerate Ru5Ps (**Fig. 2**) (Bassham et al., 1954; Basham & Calvin, 1960; Basham & Calvin, 1962; Michelet et al., 2013; Sharkey, 2018).

The Ru5P regeneration requires the action of multiple specific enzymes and the consumption of ATP molecules. The first step of the regeneration is converting G3P into dihydroxyacetone phosphate (DHAP) via the enzyme triose phosphate isomerase. The resulting DHAP molecules and G3P molecules are used to produce a fructose 6-phosphate through fructose-1,6-bisphosphatase and aldolase, respectively. The formation of fructose

1,6-bisphosphate is followed by the removal of two carbons by the enzyme transketolase and yielding an erythrose 4-phosphate. The removed two carbons are then added to G3P to form xylulose-5-phosphate (Xu5P). E4P and DHAP that were formed in the previous steps are used for the formation of sedoheptulose-1,7-bisphosphate. The obtained sedoheptulose-1,7-bisphosphate is cleaved via sedoheptulose-1,7-bisphosphatase, forming sedoheptulose-7-phosphate (S7P). This step is followed by the removal of two carbons from S7P with the enzyme transketolase. The removed carbons are added to G3P molecules, yielding a ribose-5-phosphate (R5P) and another Xu5P, respectively. In the final step of the regeneration process, the previously formed R5P and Xu5P are utilized to form Ru5P using phosphopentose isomerase and phosphopentose epimerase, respectively (Bassham et al., 1954; Bassham & Calvin, 1960; Wolosiuk et al., 1993 Sharkey, 2018).

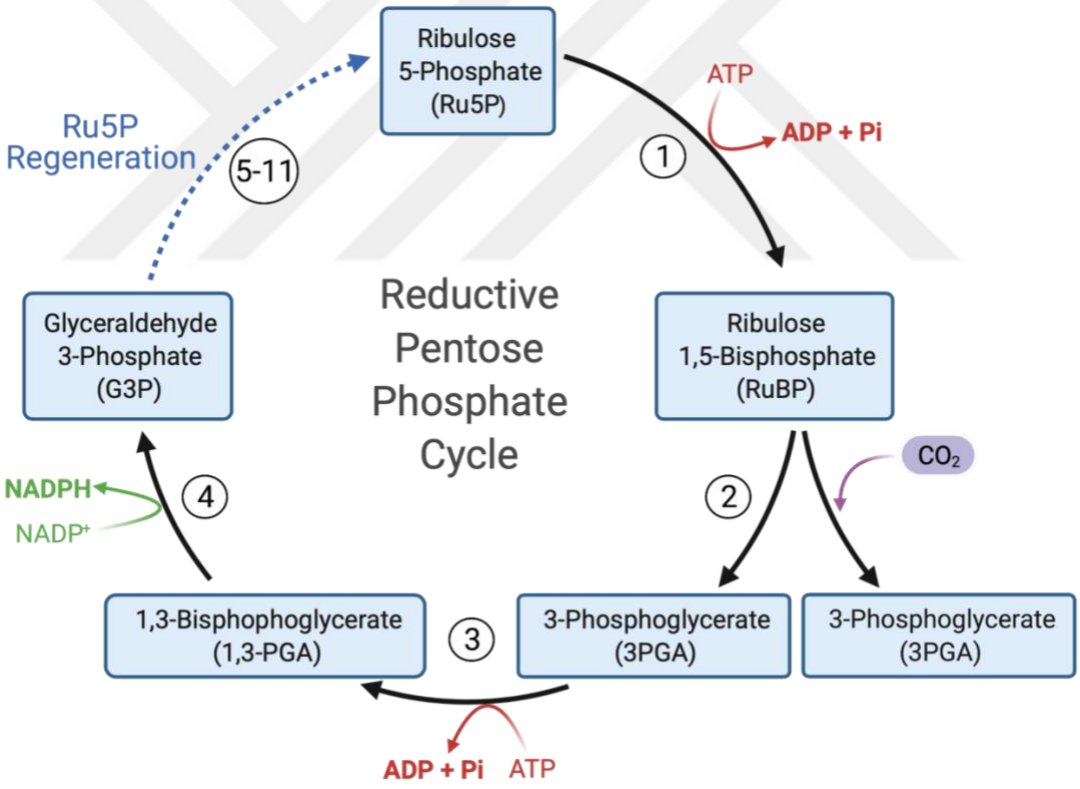


Figure 1.2: Representation of reductive pentose phosphate (Calvin-Benson-Basham) pathway*. The steps of the ribulose 5-phosphate regeneration are shown with dashed arrows. The numbers indicate the reactions catalyzed by the enzymes 1:Phosphoribulokinase, 2:Ribulose 1,5-bisphosphate carboxylase- oxygenase (RuBisCO), 3: Phosphoglycerate kinase, 4:Glyceraldehyde phosphate dehydrogenase (GAPDH), 5:Triose phosphate isomerase, 6:Aldolase, 7: Fructose 1,6-bisphosphatase, 8:Transketolase, 9:Sedoheptulase-1,7-bisphosphatase, 10: Phosphopentose isomerase, 11: Phosphopentose epimerase.*

1.3 Reductive Citric Acid Cycle

The reductive citric acid cycle is used by various anaerobic bacteria and was first identified in a green phototrophic sulfur bacterium species *C. limicola* by M.C Evans and B. B. Buchanan (Evans, Buchanan & Arnon, 1966). Later, this pathway was also found in anaerobic and microaerobic bacteria (Beh et al.,1993; Shiba et al., 1985; Schauder et al., 1987). The pathway resembles the reverse of the citric acid (Krebs) cycle, in which the precursor molecules are cycled through a reductive rather than an oxidizing direction. For instance, an acetyl-CoA is not oxidized to form a citrate molecule like in the Krebs cycle; instead, a citrate molecule is reduced to form an acetyl-CoA. Therefore the oxidizing enzymes that are used in the Krebs cycle are replaced with reductive enzymes (Fuchs, 2011).

In the first step of the cycle, a citrate molecule is cleaved into an acetyl-CoA and an oxaloacetate via the enzyme ATP citrate lyase (Ivanovsky, Sintsov & Krasilnikova, 1980). Since the oxaloacetate is a precursor molecule used in this cycle, the cycle continues from reducing the oxaloacetate molecule and resulting in a shorter cycle. Alternatively, it continues from the carboxylation of the acetyl-CoA molecule through the pyruvate synthase enzyme to yield a full cycle. The carboxylation of acetyl-CoA results in forming a pyruvate molecule (Bachofen et al., 1964). The resulting pyruvate is converted into PEP with the enzyme pyruvate: water dikinase (PEP synthetase). Subsequently, PEP is carboxylated by PEP carboxylase, forming another oxaloacetate molecule. This means that each cycle not only regenerates an oxaloacetate but also produces the same additional molecule.

The resulting oxaloacetate is then reduced using an NADH to form a malate via malate dehydrogenase, which is then converted into fumarate with fumarate hydratase. The obtained fumarate is reduced with fumarate reductase, forming a succinate. This reaction is followed by the formation of succinyl-CoA, another CO_2 acceptor molecule of the cycle through succinyl-CoA synthase. With the addition of CO_2 to the succinyl-CoA, an α-ketoglutarate molecule is formed by 2-oxoglutarate (α-ketoglutarate) synthase (Buchanan & Evans, 1965). This molecule is then reduced into isocitrate using the enzyme isocitrate dehydrogenase utilizing a CO_2 and an NADH (Kanao et al., 2002). In the last step,

isocitrate is converted into citrate via aconitate hydratase (aconitase). Thus the pathway restarts from the citrate molecule, and the maintenance of the cycle is ensured (**Fig. 3**) (Evans, Buchanan & Arnon, 1966; Buchanan & Arnon, 1990; Hügler et al., 2005; Fuchs, 2011).

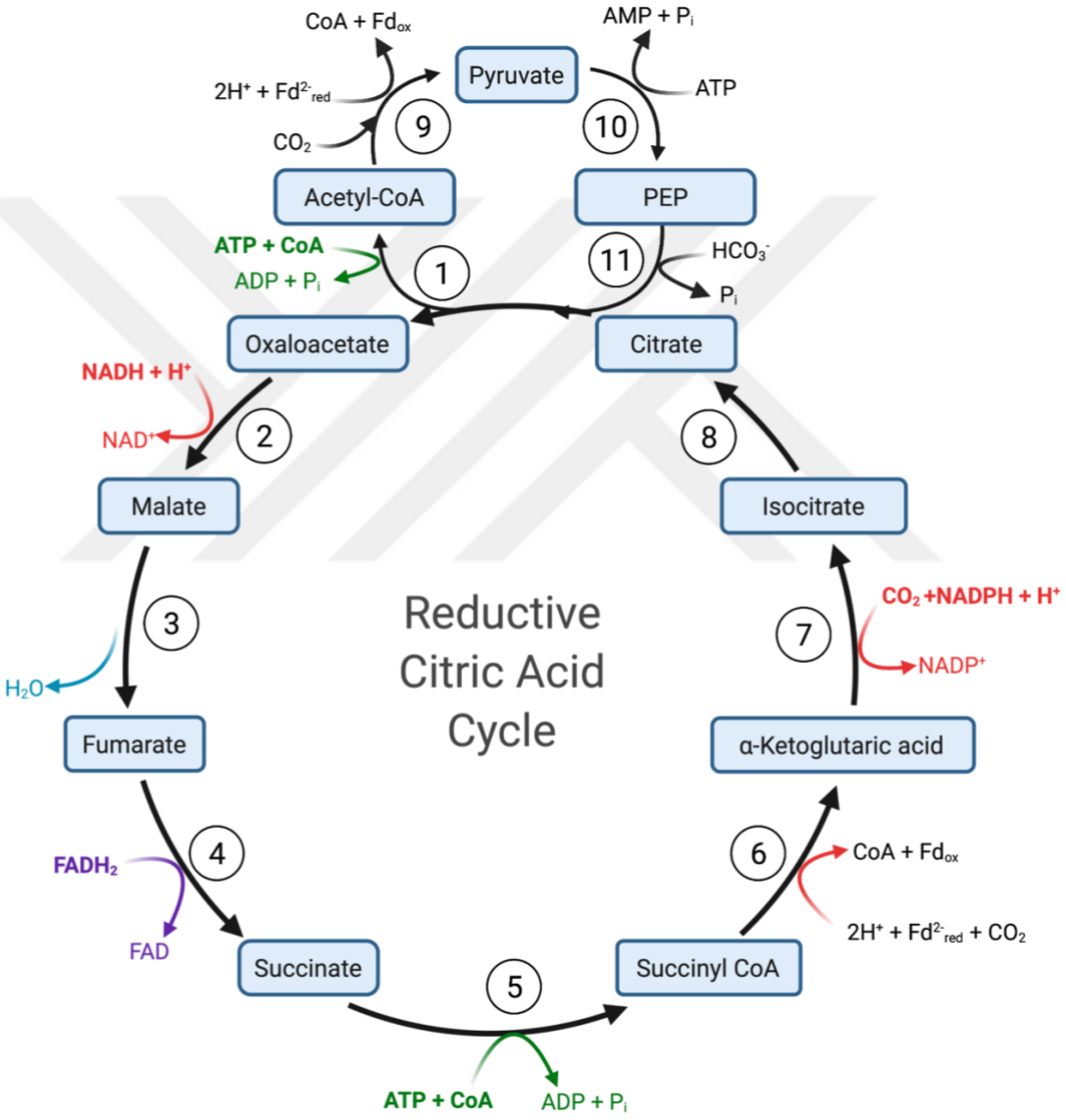


***Figure 1.3**: **The schematic representation of the reductive citric acid cycle.** The numbers indicate the reactions catalyzed by the enzymes 1: ATP citrate lyase; 2: Malate dehydrogenase; 3:Fumarate hydratase; 4: Fumarate reductase; 5:Succinyl-CoA synthase; 6: 2-oxoglutarate (α-ketoglutarate) synthase; 7:Isocitrate dehydrogenase; 8:Aconitate hydratase (aconitase); 9: Pyruvate synthase; 10:Pyruvate: water dikinase; 11:PEP carboxylase.*

1.4 Reductive Acetyl-CoA (Wood-Ljungdahl) Pathway

The reductive acetyl-CoA pathway was first identified in methanogenic archaea and strictly anaerobic acetogenic bacteria by Harland G. Wood and Lars G. Ljungdahl. At a glance, the pathway results in the formation of acetyl-CoA with the fixation of two CO_2 molecules. In this pathway, one of the CO_2 molecules is reduced to a carbonyl group attached to a metal atom located in the active site of a CO_2 accepting enzyme. This part of the reaction is called the carbonyl branch of the pathway. The second CO_2 is reduced to a methyl group attached to a carrier coenzyme, which is referred to as the methyl branch of the pathway. As a consequence of these reactions, the resulting methyl group and carbonyl group are used to form an acetyl-CoA molecule which is used later to produce acetate and biomass (Ljungdahl & Wood, 1969; Ljungdahl, 1986; Wood, 1991; Drake et al., 2008; Ljungdahl, 2009; Hügler & Sievert, 2011). Compared with most carbon fixation pathways, the reductive acetyl-CoA pathway does not require the recycling of the intermediates produced in the reaction process. Although the logic of the mechanism is the same in the organisms utilizing this pathway, the enzymes and the cofactors might differ.

According to the pathway proposed by Wood and Ljungdahl, the methyl branch of the pathway starts with the reduction of a CO_2 molecule to a formic acid (formate) molecule via formyl dehydrogenase (Thauer, 1972; Li, Ljungdahl & Wood, 1966). The resulting formate is used to convert tetrahydrofolate (H_4F) into formyl-H_4F using formyl-H_4F synthetase enzyme (Himes & Cohn, 1967; Himes, Harmony & Rabinowitz, 1973). The formyl-H_4F is then used to form a methenyl-H4F which is reduced to a methylene-H_4F by the enzymes methenyl-H_4F cyclohydrolase and methylene-H_4F dehydrogenase, respectively (Clark & Ljungdahl, 1982; Moore, O'Brien & Ljungdahl, 1974). With a further reduction of the latter molecule by methylene-H_4F reductase (Clark & Ljungdahl, 1984), a methyl-H_4F molecule is produced. The resulting methyl group is transferred via methyltransferase to a corrinoid iron-sulfur protein (CFeSP). This reaction changes the redox state of the center Co(I) atom to methyl-Co(III). Concurrently, the carbonyl branch of the reaction is initiated with the reduction of CO_2 to a carbonyl group of the bifunctional enzyme carbon monoxide dehydrogenase (CODH)/Acetyl-CoA synthase. In the final step, the resulting methyl group (CH_3) and the carbonyl group (CO) are used to

form an acetyl-CoA molecule with the same bifunctional enzyme (CODH)/Acetyl-CoA synthase (**Fig. 4**) (Ljungdahl, 1986; Wood, 1991; Drake et al., 2008; Ljungdahl, 2009; Ragsdale & Pierce, 2008).

Alternatively, in some archaeal species such as Archaeoglobus and Ferroglobus, a CO_2 molecule is reduced to a CO attached to a nickel atom located in the reaction center of the enzyme CO dehydrogenase. Meanwhile, the methyl branch of the pathway is initiated with the fixation of another CO_2 molecule, converting a methanofuran molecule into a formylmethanofuran molecule via formylmethanofuran dehydrogenase. This reaction is followed by the transfer of the formyl group of formylmethanofuran to the coenzyme tetrahydromethanopterin. This reaction is catalyzed by the enzyme formylmethanofuran: tetrahydromethanopterin formyltransferase, yielding a 5-formyltetrahydromethanopterin. The resulting 5-formyltetrahydromethanopterin is then converted to 5,10-methenyltetrahydromethanopterin with the enzyme methenyl-tetrahydromethanopterin cyclohydralase. The obtained molecule is used to form a 5,10-methylene-tetrahydromethanopterin molecule via methylene-tetrahydromethanopterin dehydratase. Subsequently, the previous molecule is utilized to synthesize a methyltetrahydromethanopterin by methylene-tetrahydromethanopterin reductase. The reaction is concluded by forming an acetyl-CoA molecule, using the methyl group of the methyltetrahydromethanopterin and the previously formed nickel-bound CO through CO dehydrogenase-acetyl-CoA-synthase enzyme (Vorholt et al., 1995; Vorholt et al., 1997; Berg et al., 2010).

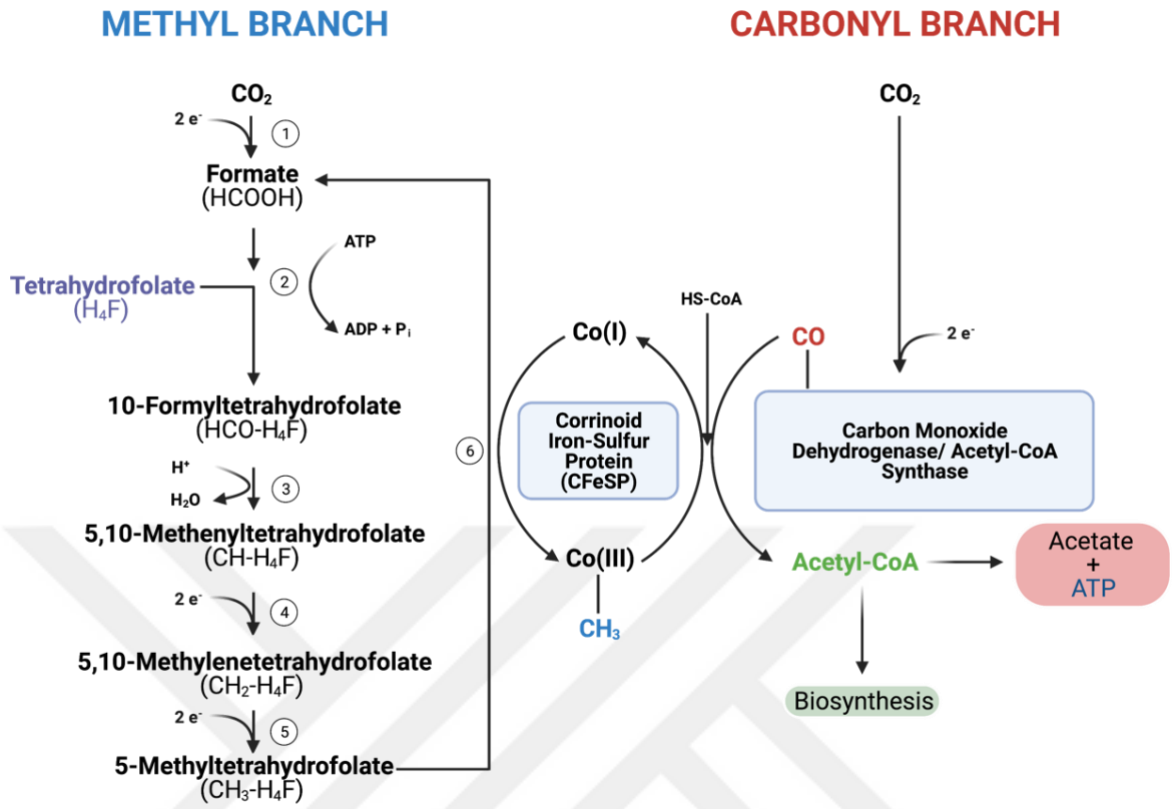


Figure 1.4: The schematic representation of Reductive Acetyl-CoA (Wood-Ljungdahl) Pathway*. The numbers indicate the reactions catalyzed by the enzymes 1: Formyl dehydrogenase; 2: Formyl-H4F synthetase; 3: Methenyl-H4F cyclohydrolase; 4: Methylene-H4F dehydrogenase; 5: Methylene-H4F reductase; 6: Methyltransferase.*

1.5 Dicarboxylate/4-Hydroxybutyrate Cycle

The dicarboxylate/4-hydroxybutyrate cycle is one of the main carbon fixation mechanisms utilized by autotrophic anaerobic and microaerobic Cranarchaeal species, including the orders Desulfurococcales (Huber et al., 2008) and Thermoproteales (Ramos-Vera, Berg & Fuchs, 2009). The pathway involves the enzymes used in the reductive citric acid cycle and the 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle. At a glance, in the first half of the reaction, an acetyl-CoA molecule is converted into a succinyl-CoA molecule, which requires the enzymes involved in the reductive citric acid cycle. The second half of the reaction uses the enzymes that are involved in HP/HB cycle. In this part of the cycle, the succinyl-CoA is first converted into a 4-hydroxybutyrate molecule, an intermediate required to produce two acetyl-CoA molecules. Once the acetyl-CoA molecules are produced, one is used for biosynthesis

while the second is recycled to fix a CO_2 molecule and yield a pyruvate molecule (**Fig. 5**). Therefore the continuity of this cycle is maintained (Huber et al., 2008; Berg et al., 2010; Hügler & Sievert, 2011).

The first half of the reaction starts with converting an acetyl-CoA molecule to a pyruvate molecule via the pyruvate synthase enzyme. The obtained pyruvate is then used to produce a PEP, which is carboxylated to yield an oxaloacetate molecule with the enzymes pyruvate: water dikinase (Ramos-Vera, Berg & Fuchs, 2009) and PEP carboxylase (Patel, Kraszewski & Mukhopadhyay, 2004; Ettema et al., 2004) respectively. The reaction follows the steps, including the reduction of the oxaloacetate into a malate and conversion of the resulting malate into a fumarate by malate dehydrogenase and fumarate hydratase. The latter intermediate is reduced into a succinate, which is, in turn, converted into a succinyl-CoA via fumarate reductase and succinate thiokinase enzymes, respectively. The second half of the cycle is initiated with the reduction of succinyl-CoA into a succinic semialdehyde, the key intermediate required to yield a 4-hydroxybutyrate molecule. These two reactions take place with the help of the enzymes succinyl-CoA reductase and succinate semialdehyde reductase. Subsequently, through the action of 4-hydroxybutyryl-CoA synthetase enzyme, a coenzyme A molecule is attached to the 4-hydroxybutyrate forming a 4-hydroxybutyryl-CoA (Huber et al., 2008; Ramos-Vera, Berg & Fuchs, 2009). Following the formation of 4-hydroxybutyryl-CoA, a crotonyl-CoA molecule is produced, which is further converted to an (S)-3-hydroxybutyryl-CoA. These reactions necessitate the usage of [4F-4S] and FAD containing 4-hydroxybutyryl-CoA dehydratase (Buckel & Golding, 2006) and crotonyl CoA hydratase (Huber et al., 2008). The latter molecule is used to generate an acetoacetyl-CoA, the precursor of acetyl-CoA via 3-hydroxybutyryl-CoA dehydratase. The last step of the reaction is catalyzed by the enzyme acetoacetyl-CoA ß-ketothiolase, resulting in the formation of two acetyl-CoA molecules, one of which is consumed for biosynthesis and the second is used for the maintenance of the cycle (Huber et al., 2008; Ramos-Vera, Berg & Fuchs, 2009; Fuchs, 2011).

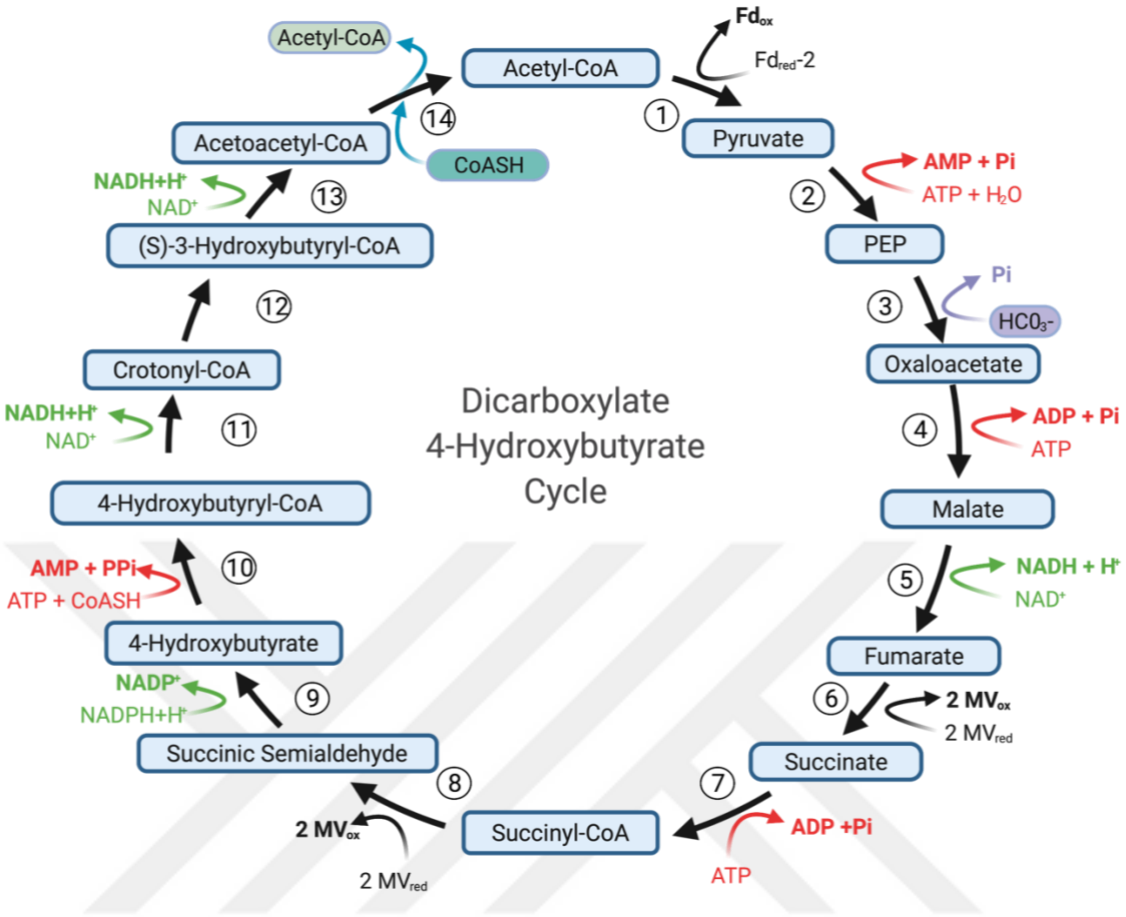


Figure 1.5: The schematic representation of the dicarboxylate/4-hydroxybutyrate cycle. *The numbers indicate the reactions catalyzed by the enzymes 1: Pyruvate synthase 2: Pyruvate: water dikinase 3:PEP carboxylase 4: Malate dehydrogenase 5: Fumarate hydratase 6: Fumarate reductase 7: Succinate thiokinase 8: Succinyl-CoA reductase 9: Succinate semialdehyde reductase 10: 4-Hydroxybutyryl-CoA synthetase 11: 4-Hydroxybutyryl-CoA dehydratase 12: Crotonyl-CoA hydratase 13: 3-Hydroxybutyryl-CoA dehydrogenase 14: Acetoacetyl-CoA ß-ketothiolase.*

1.6 3-Hydroxypropionate Bicycle

3-Hydroxyproiponate bicycle is an autotrophic carbon fixation pathway which operates in the various members of the family Chloroflexaceae and was initially discovered in a phototrophic bacteria species *C. aurantiacus* (Strauss et al., 1992; Eisenreich et al. 1993; Strauss & Fuchs, 1993; Herter et al., 2002). Rather than the other autotrophic carbon fixation cycles, this pathway necessitates two separate cycles. At a glance, the first cycle converts two bicarbonate (HCO_3^-) and one acetyl-CoA molecules to an (S)-malyl-CoA, and its cleavage results in the formation of an acetyl-CoA and a glyoxylate. In the second cycle, the resulting glyoxylate and a propionyl-CoA are used to produce a pyruvate and an acetyl-CoA. The resulting pyruvate is used for biosynthesis precursor, and the regenerated acetyl-CoA contributes to the continuity of the cycle (Zarzycki et al., 2009).

The first cycle is initiated with the carboxylation of an acetyl-CoA molecule via acetyl-CoA carboxylase, which fixes a bicarbonate to yield a malonyl-CoA. The reduction of the malonyl-CoA via malonyl-CoA reductase (Hügler et al., 2002) leads to the production of 3-hydroxypropionate, the precursor required to form a propionyl-CoA by the enzyme propionyl-CoA synthase (Holo, 1989; Alber & Fuchs, 2002). The resulting propionyl-CoA is carboxylated via propionyl-CoA carboxylase, leading to the formation of a methylmalonyl-CoA through fixation of the second bicarbonate molecule. This reaction is followed by the conversion of methylmalonyl-CoA to succinyl-CoA through the enzymes methylmalonyl-CoA epimerase and methylmalonyl-CoA mutase, respectively. Subsequently, the succinyl-CoA is converted into succinate via succinyl-CoA: malate CoA transferase and the succinate is in turn converted to malate by succinate dehydrogenase. Through the transfer of coenzyme-A obtained in the previous reaction to the malate, a malyl-CoA molecule is formed. The obtained molecule is then cleaved into an acetyl-CoA and a glyoxylate via an enzyme called (S)-Malyl-CoA/ß-methylmalyl-CoA/(S)-citramalyl-CoA (MMC) lyase (Herter et al., 2001; Herter, Busch & Fuchs, 2002). The acetyl-CoA is used for the continuity of the first cycle, and the glyoxylate is utilized in the second cycle (Hügler & Fuchs, 2005).

In the second cycle, the glyoxylate obtained in the first cycle condensates with a propionyl-CoA molecule through the MMC lyase enzyme used in the previous reaction. A methylmalonyl-CoA is obtained as a result. Following this reaction, the methylmalonyl-CoA is used to form a mesaconyl-C1-CoA, which is, in turn, converted into a mesaconyl-C4-CoA by the enzymes mesaconyl-C1 CoA hydratase and mesaconyl-C1-C4 CoA transferase respectively. Afterwards, the previous product is used to form a citramalyl-CoA via mesaconyl-C4 CoA hydratase, which is in turn disproportionated into an acetyl-CoA and a pyruvate via MMC lyase. The obtained acetyl-CoA is carboxylated using a bicarbonate molecule via acetyl-CoA carboxylase to yield a malonyl-CoA, where the pyruvate is used for biosynthesis and leaves the reaction cycle. This reaction follows the reduction of malonyl-CoA via malonyl-CoA reductase, and a 3-hydroxypropionate is obtained. Eventually, through a reaction catalyzed by propionyl-CoA synthase, a coenzyme A molecule attaches to the 3-hydroxypropionate, yielding the first intermediate of the second reaction cycle, a propionyl-CoA (Strauss & Fuchs, 1993; Herter et al., 2002; Zarzycki et al., 2009; Hügler & Sievert, 2011).

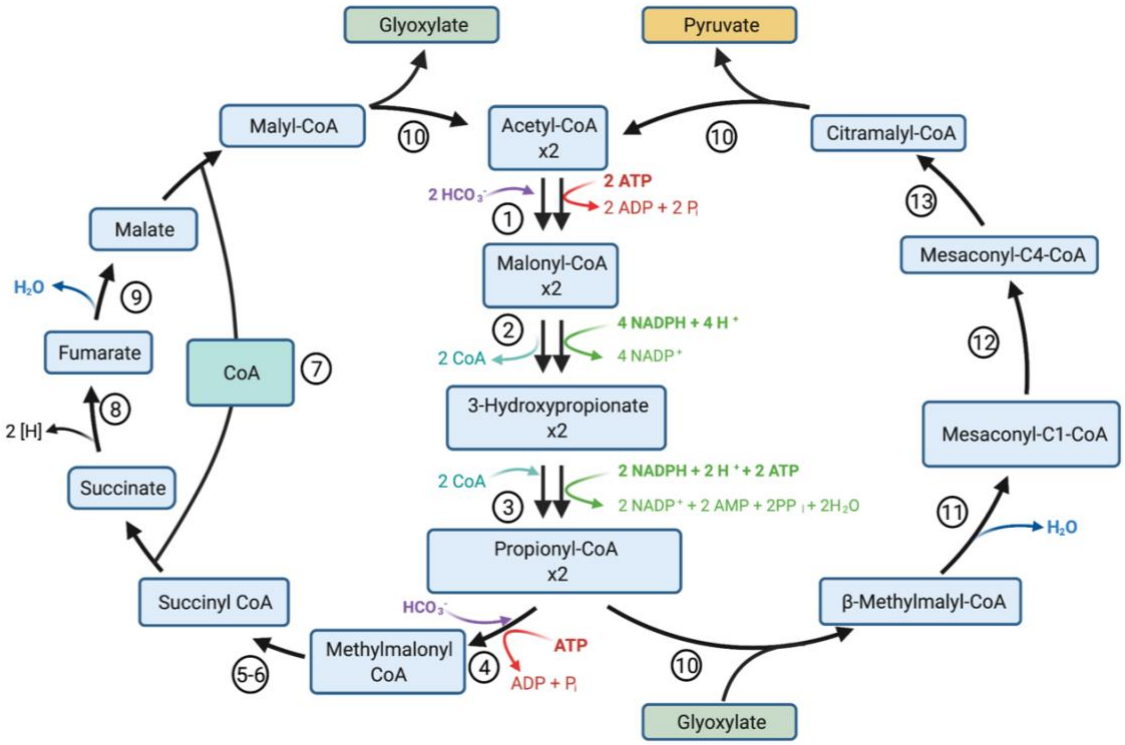


Figure 1.6: Schematic representation of 3-hydroxypropionate bicycle *(proposed by Zarzycki et al.). The numbers indicate the reactions catalyzed by the enzymes 1: Acetyl-CoA carboxylase 2: Malonyl-CoA reductase 3:Propionyl-CoA synthase 4: Propionyl-CoA carboxylase 5: Methylmalonyl-CoA epimerase 6: Methylmalonyl-CoA mutase 7: Succinyl-CoA:(S)-malate-CoA transferase 8: Succinate dehydrogenase 9: Fumarate hydratase 10: (S)-Malyl-CoA/ß-methylmalyl-CoA/(S)-citramalyl-CoA (MMC) lyase 11: Mesaconyl-C1 CoA hydratase 12: Mesaconyl-C1-C4 CoA transferase 13: Mesaconyl-C4 CoA hydratase. The reaction diagram's left and right sides represent the first and the second cycles, respectively.*

1.7 3-Hydroxypropionate/4-Hydroxybutyrate (HP/HB) Cycle

The operation of the 3-hydroxypropionate/4-hydroxybutyrate cycle was initially identified in the Crenarchaeal order *Sulfolobales* (Berg et al., 2007; Berg et al. 2010). Lately, it was understood that some of the members of the phylum Thaumarchaeota, such as *N. maritimus* (Walker et al., 2010; Könneke et al., 2014) and *C. symbiosum* (Hallam et al., 2006), also operates a modified version of this pathway. In this pathway, two acetyl-CoA are formed, which requires the assimilation of two bicarbonate molecules. In the first half of the cycle, a molecule of acetyl-CoA is progressively converted into a succinyl-CoA, which necessitates the usage of 3-hydroxypropionate as an intermediate. In the second half, the succinyl-CoA is also incrementally converted to two acetyl-CoA molecules through a 4-hydroxybutyrate intermediate (***Fig. 1.7***) (Berg et al., 2007).

The first part of the cycle is initiated through the carboxylation of acetyl-CoA using a bicarbonate molecule via a bifunctional enzyme acetyl-CoA/propionyl-CoA carboxylase. This reaction results in the formation of a malonyl-CoA molecule (Chuakrut et al., 2003; Hügler et al., 2003). The resulting malonyl-CoA is then reduced to a malonic semialdehyde, where the resulting product is further reduced to a 3-hydroxypropionate using the enzymes malonyl-CoA reductase (Alber et al., 2006) and malonic semialdehyde reductase, respectively (Kockelkorn & Fuchs, 2009). By the attachment of a coenzyme A molecule to the 3-hydroxypropionate via 3-hydroxypropionyl-CoA synthetase enzyme (Alber, Kung & Fuchs, 2008), a molecule of 3-hydroxypropionyl-CoA is formed. The resulting intermediate is converted to an acryloyl-CoA molecule through a dehydration reaction catalyzed by 3-hydroxypropionyl-CoA dehydratase. The latter molecule is then reduced to a propionyl-CoA, the intermediate required to form a methylmalonyl-CoA via acryloyl-CoA reductase (Teufel et al., 2009). The carboxylation of propionyl-CoA to methylmalonyl-CoA is catalyzed by the previously mentioned bifunctional acetyl-CoA/propionyl-CoA carboxylase (Chuakrut et al., 2003; Hügler et al., 2003). Therefore the second bicarbonate molecule is also fixed through this carboxylation reaction. With the action of methylmalonyl-CoA epimerase and methylmalonyl-CoA mutase enzymes, the obtained methylmalonyl-CoA is converted into a succinyl-CoA (Berg et al. 2010b). The obtained succinyl-CoA is either used for sustaining the second half of the cycle or for biosynthesis, which requires an extra half-turn of the cycle. For instance, in *M. sedula* (Huber et al., 1989), the succinyl-CoA is converted into oxaloacetate or malate, which are then decarboxylated to yield PEP or pyruvate (Estelmann, 2011; Könneke et al., 2014).

The second half of the reaction is initiated with the reduction of succinyl-CoA via succinyl-CoA reductase, yielding a molecule of succinic semialdehyde. With the reduction of succinic semialdehyde via succinate semialdehyde reductase, a molecule of 4-hydroxybutyrate forms (Kockelkorn & Fuchs, 2009). This reaction follows the attachment of a CoA molecule to the 4-hydroxybutyrate, forming a 4-hydroxybutyryl-CoA through the catalysis done by 4-hydroxybutyryl-CoA synthetase (Berg et al., 2010a; Könneke et al., 2014). The resulting intermediate is then converted into a crotonyl-CoA through a dehydration reaction via 4-hydroxybutyryl-CoA dehydratase (Martins et al., 2004). Subsequently, the crotonyl-CoA is converted to a 3-hydroxybutyryl-CoA, which

is then dehydrogenated to yield an acetoacetyl-CoA by crotonyl-CoA hydratase and 3-hydroxybutyryl-CoA dehydrogenase enzymes, respectively. The last reaction is catalyzed by acetoacetyl-CoA β-ketothiolase, converting the acetoacetyl-CoA into two molecules of acetyl-CoA, where one of them leaves the reaction cycle and participates in biosynthesis, and the second sustains the cycle (Berg et al., 2007; Berg et al., 2010ab; Fuchs, 2011; Hügler & Sievert, 2011; Könneke et al., 2014).

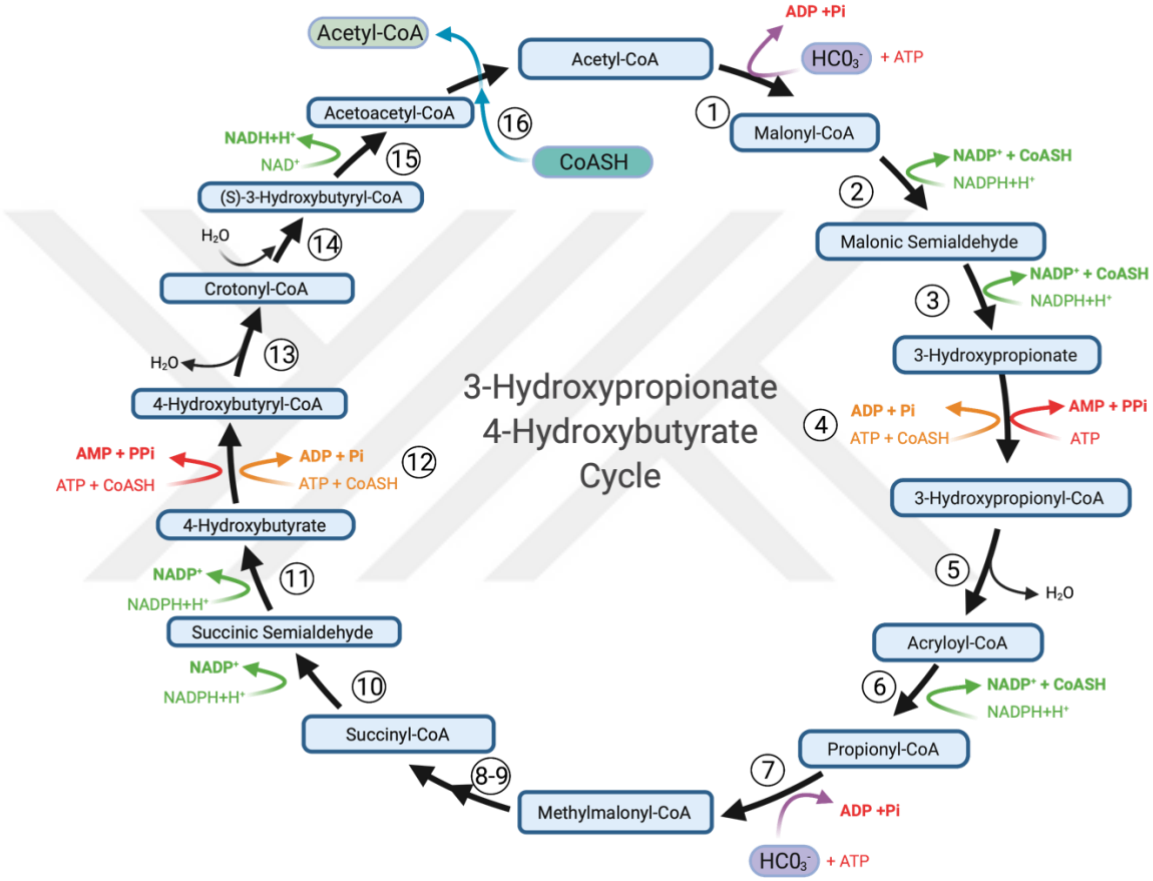


Figure 1.7: Representation of the 3-Hydroxypropionate/4-Hydroxybutyrate cycle*. The numbers indicate the reaction catalyzed by the enzymes 1: Acetyl-CoA/Propionyl-CoA carboxylase, 2:Malonyl-CoA reductase, 3:Malonic semialdehyde reductase, 4:3-Hydroxypropionate-CoA ligase(Red: Crenarchaeota, Orange: Thaumarchaeota), 5: 3-Hydroxypropionyl-CoA dehydratase, 6:Acryloyl-CoA reductase, 7:Acetyl-CoA/Propionyl-CoA carboxylase, 8: Malonyl-CoA epimerase, 9: Malonyl-CoA mutase, 10:Succinyl-CoA reductase, 11:Succinic semialdehyde reductase, 12:4-Hydroxybutyrate-CoA ligase (Red: Crenarchaeota, Orange: Thaumarchaeota), 13:4-Hydroxybutyryl-CoA dehydratase, 14:Crotonyl-CoA hydratase, 15:(S)-3-Hydroxybuyryl-CoA dehydrogenase, 16:Acetoacetyl-CoA ß-ketothiolase*

Similar to the other carbon fixation pathways, acetyl-CoA is one of the main precursors required for biosynthesis. On the other hand, rather than in the anaerobic organisms, the production of the intermediates required for the biosynthesis is not provided with the reductive carboxylation of the acetyl-CoA (Könneke et al., 2014). Therefore, the

organisms that use the HP/HB cycle utilize biosynthesis through a different mechanism. As described earlier, the succinyl-CoA is a significant intermediate for biosynthesis as it is used to produce pyruvate and PEP, which necessitates another half turn of the cycle (Estelmann et al., 2011).

During this process, the succinyl-CoA is converted into succinate via succinyl-CoA synthetase. Alternatively, a succinate can also be obtained through a dehydrogenation reaction of a succinic semialdehyde by succinic semialdehyde dehydrogenase. This reaction follows the formation of a malate molecule through succinate dehydrogenase and fumarate hydratase enzymes respectively. After this step, the malate is either decarboxylated to yield a pyruvate molecule through the malic enzyme, or converted into an oxaloacetate by malate dehydrogenase. The pyruvate produced in this process is utilized to form a PEP through the catalysis facilitated by pyruvate: water dikinase. The formation of the oxaloacetate follows a decarboxylation reaction and results in the production of another PEP by the enzyme PEP carboxykinase (Fukuda et al., 2004). Alternatively, the previous oxaloacetate may also undergo a condensation reaction with acetyl-CoA, yielding a citrate molecule facilitated by (si)-citrate synthase. Following this step, the citrate molecule is used to produce a 2-oxoglutarate molecule via aconitase and isocitrate dehydrogenase enzymes. As a result, 2-oxoglutarate and PEP molecules are used for biosynthesis. For instance, in *M. sedula*, the PEP molecules are used to synthesize triosephosphates through a set of reactions catalyzed by the enzymes: enolase, phosphoglycerate mutase, phosphoglycerate kinase, glyceraldehyde 3-phosphate dehydrogenase, and triosephosphate isomerase, respectively. In the final step, the triosephosphates are converted into fructose 6-phosphate through fructose 1,6 bisphosphate aldolase/phosphatase (***Fig 1.8***) (Estelmann et al., 2011; Könneke et al., 2014).

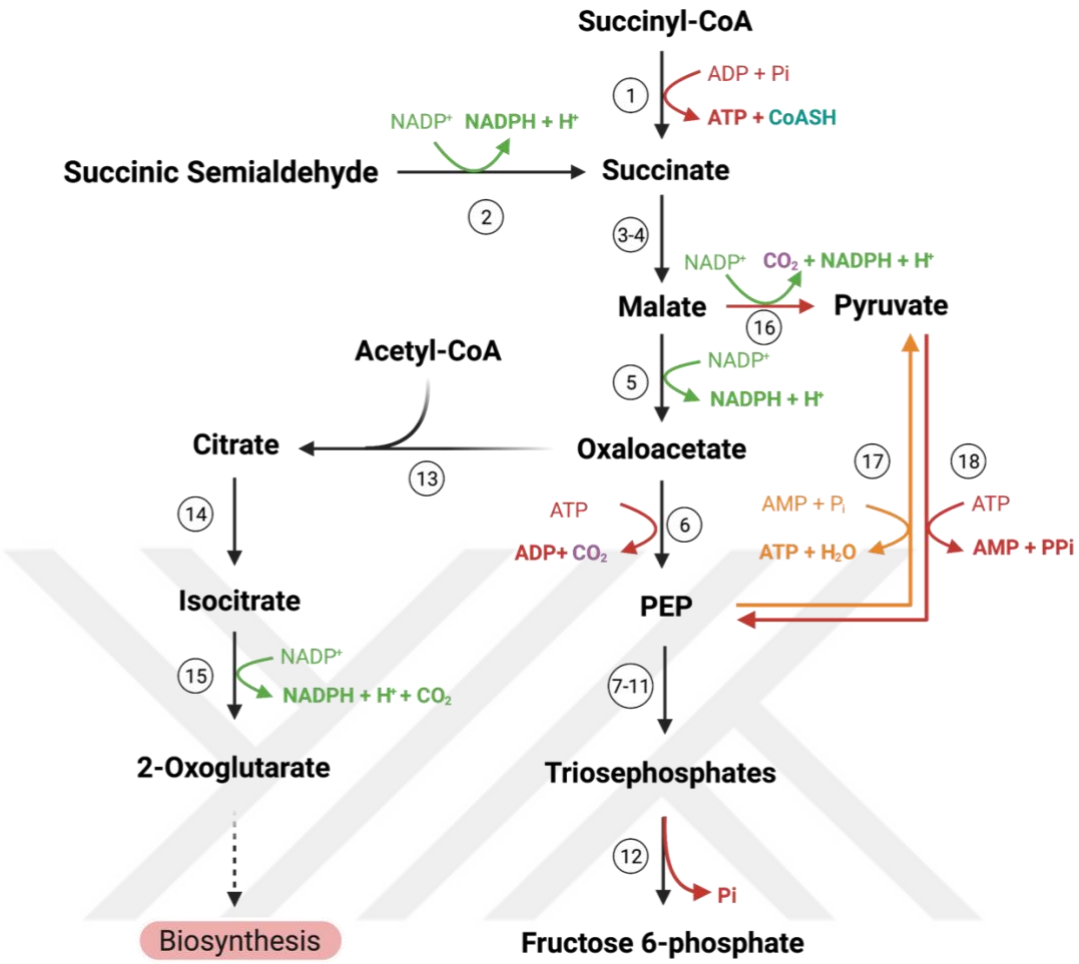


Figure 1.8: The schematic representation of the biosynthesis pathway proposed by Estelmann et al. *The numbers indicate the reactions catalyzed by the enzymes 1: Succinyl-CoA synthetase, 2: Succinic semialdehyde dehydrogenase, 3:Succinate dehydrogenase, 4:Fumarate hydratase, 5: Malate dehydrogenase, 6:PEP carboxykinase, 7: Enolase, 8: Phosphoglycerate mutase, 9: Phosphoglycerate kinase, 10: Glyceraldehyde 3-phosphate dehydrogenase, 11:Triosephosphate isomerase, 12: fructose 1,6-bisphosphate aldolase/phosphatase, 13: (si)-citrate synthase, 14: Aconitase, 15: Isocitrate dehydrogenase 16: Malic enzyme, 17: Pyruvate: phosphate dikinase, 18:Pyruvate: water dikinase. (**Red:** M. sedula, **Orange:** N. maritimus)*

1.8 Thaumarchaeota Uses a Modified HP/HB Cycle

The phylum Thaumarchaeota constitutes the third branch of the domain Archaea together with Euryarchaeota and Crenarchaeota. Before the proposal of this third phylum, the Thaumarchaeota was affiliated with the mesophilic members of marine and terrestrial Crenarchaeota. However, the genomic and phylogenetic studies revealed that this phylum significantly differs from the phylum Crenarchaeota as many of the Crenarchaeal species are composed of hyperthermophilic members (Brochier-Armanet et al., 2008; Spang et al., 2010).

Within few years after the discovery of the HP/HB cycle in Crenarcheaota (Berg et al., 2007), the same pathway was also identified in the Thaumarchaeal species *N. maritimus* SCM1 (Könneke et al., 2014). *Nitrosopumilus maritimus* (SCM1) is a mesophilic marine chemolithotroph which grows in aerobic conditions. Interestingly, this species was considered the first ammonia-oxidizing species in the domain Archaea (Könneke et al., 2005). Kinetic studies show that *N. maritimus* possess an exceedingly high affinity to ammonia (Martens-Habbena et al., 2009). Furthermore, when compared with the Crenarchaeal species, *N. maritimus* was found to utilize a modified version of the HP/HB cycle. Currently, the HP/HB cycle utilized by *N. maritimus* and the other Thaumarchaeal species is considered the most energy-efficient aerobic carbon fixation mechanism (Könneke et al., 2014). These characteristics give these species various advantages to thrive in extreme oligotrophic environments where the energy supplies and ammonia concentrations are exceedingly low (Stahl & de la Torre, 2012). Moreover, it was discovered that the ammonia-oxidizing archaea (AOA) are more abundant in the oceans and the soil when compared with the abundance of ammonia-oxidizing bacteria (AOB), making them the key players of the carbon cycle and the nitrification in these habitats (Leininger et al., 2006; Wuchter et al., 2006; Yakimov et al., 2011).

Compared with the phylum Crenarchaeota, the Thaumarchaeal species *Nitrosopumilus maritimus* utilize the HP/HB cycle in a much more energy-efficient way. This high energy efficiency is achieved through three properties: the use of ADP rather than AMP forming enzymes, possessing multifunctional enzymes, and high oxygen tolerance (Könneke et al., 2014). In Crenarchaeota, the synthesis of 3-hydroxypropionyl-CoA and 4-hydroxybutyryl-CoA requires the activity of an AMP forming 3-hydroxypropionyl-CoA synthetase and another AMP forming 4-hydroxybutyryl-CoA synthetase. However, *N. maritimus* uses an alternative ADP-forming variant of these two enzymes. This means that, in one turn of the reaction cycle, Thaumarchaea only spends half of the high energy bonds that Crenarchaea consume (Könneke et al., 2014).

Furthermore, the carboxylation of acetyl-CoA and propionyl-CoA in Crenarchaeal and Thaumarchaeal HP/HB cycle requires a single bifunctional biotin-dependent enzyme, Acetyl-CoA/Propionyl-CoA carboxylase (Nmar_0272/0273/0274 in *N. maritimus)*. However, in the 3-Hydroxypropionate bicycle, the carboxylation of these two

intermediates requires two distinct acetyl-CoA and propionyl-CoA carboxylases (Zarzycki et al., 2009). Similarly, the dehydration of 3-hydroxypropionyl-CoA and the hydration of crotonyl-CoA is also promoted by a single bifunctional enzyme named 3-hydroxypropionyl-CoA dehydratase/crotonyl-CoA hydratase (Nmar_1308) in *N. maritimus*. Thus, the energy consumption required for enzyme biosynthesis is significantly reduced (Könneke et al., 2014; Liu et al., 2021).

Compared to the closely related highly oxygen-sensitive 4-hydroxybutyryl-CoA dehydratases from *Clostridium aminobutyricum* and *Clostridium kluyveri* (Gerhardt et al., 2000; Scherf et al., 1994); *N. maritimus* 4-hydroxybutyryl-CoA dehydratase (Nmar_0207) was observed to be oxygen insensitive. This oxygen tolerance validates that *N. maritimus* operates the HP/HB cycle in aerobic conditions. A similar oxygen tolerance was observed in the 4-hydroxybutyryl-CoA dehydratase of Crenarchaeal species *M. Sedula* (Hawkins et al., 2014). The oxygen tolerance increases the half-life of the 4-hydroxybutryryl-CoA dehydratase significantly, therefore dramatically decreases the turnover and maintenance costs of this enzyme (Könneke et al., 2014).

1.9 ATP Requirements of the Aerobic Carbon Fixation Pathways

Each carbon fixation pathway contributes to the synthesis of central metabolic precursors. These precursors are required for the production of complex carbon compounds with a process called biosynthesis. Such metabolic precursors include phosphoenolpyruvate (PEP), pyruvate, oxaloacetate, 2-oxoglutarate, and acetyl-CoA. On the other hand, each of these pathways requires different energy needs in terms of ATP consumption (***Table 1.1***).

1.9.1 Calvin-Benson-Basham cycle

In order to form a G3P molecule, 9 ATP molecules need to be consumed. The conversion of G3P to PEP through glycolysis yields 1 ATP molecule, meaning that net 8 ATP is consumed to form one molecule of PEP. Through carboxylation of PEP via PEP carboxylase, an oxaloacetate molecule is obtained. However, this reaction does not require to use of ATP. Hence net 8 ATP is required to form one molecule of oxaloacetate. The further conversion of PEP to pyruvate releases an extra ATP molecule. Therefore the

net consumption required to form a pyruvate is 7 ATP. The dehydrogenation of pyruvate by pyruvate dehydrogenase yields an acetyl-CoA molecule. Similarly, no ATP molecule is used in this process, making the net consumption 7 ATP. The formation of 2-oxoglutarate from acetyl-CoA and oxaloacetate is a multi-step reaction and required for glutamate synthesis. The catalysis necessitates the action of citrate synthase, aconitase, and isocitrate dehydrogenase enzymes and spends a net of 15 ATP molecules (Basham et al., 1954; Könneke et al., 2014).

1.9.2 3-Hydroxypropionate bicycle

Pyruvate is the main product of this cycle, and its synthesis requires the consumption of 7 ATP molecules (Strauss & Fuchs, 1993; Herter et al., 2002). The obtained pyruvate is converted to an acetyl-CoA molecule through pyruvate dehydrogenase. In this reaction, no ATP molecule is spent, meaning that acetyl-CoA synthesis requires 7 ATP molecules. Alternatively, pyruvate can be converted into PEP using the enzyme pyruvate phosphate dikinase (Holo & Sirevag, 1986; Ivanovsky, Krasilnikova & Fal, 1993) by using net 9 ATPs. Similarly, oxaloacetate synthesis from PEP via PEP carboxylase does not consume ATP molecules and requires net 9 ATP. The oxoglutarate synthesis necessitates the usage of acetyl-CoA and oxaloacetate and the net usage of 16 ATP molecules. This process utilizes multiple reactions catalyzed by citrate synthase, aconitase, and isocitrate dehydrogenase (Könneke et al., 2014).

1.9.3 HP/HB Cycle (M. sedula)

The synthesis of acetyl-CoA requires 6 ATP molecules. Since the production of succinyl-CoA for biosynthesis necessitates another half-turn of the cycle, an additional 4 ATP is consumed. Hence, succinyl-CoA synthesis costs 10 ATP molecules. The conversion of succinyl-CoA to oxaloacetate by succinyl-CoA synthetase, succinate dehydrogenase, fumarate hydratase, and malate dehydrogenase releases 1 ATP. Thus the net ATP consumed to generate an oxaloacetate is 9. Alternatively, an oxaloacetate can be synthesized through succinic semialdehyde dehydrogenase, succinate dehydrogenase, fumarate hydratase, and malate dehydrogenase without using an ATP (Estelmann et al., 2011). In this case, net 10 ATP is used for this conversion. Depending on these two pathways, the conversion of oxaloacetate to PEP via PEP carboxykinase requires a net of

10 or 11 ATPs. Similarly, depending on the synthesis pathway of malate, the pyruvate can be either formed using 9 or 10 ATPs via the malic enzyme. Through a set of reactions involving the condensation of acetyl-CoA and oxaloacetate, 2-oxoglutarate is formed with the net cost of 15 or 16 ATP (Könneke et al., 2014).

1.9.4 HP/HB Cycle (N. maritimus)

By using two ADP-forming synthetases, acetyl-CoA formation costs only 4 ATPs. Therefore, the additional half-turn of the cycle for succinyl-CoA synthesis requires 7 ATPs. Since the conversion of succinyl-CoA to oxaloacetate produces 1 ATP, the net cost is 6 ATPs. The malic enzyme and PEP carboxylase have not been identified in *N. maritimus* to convert C_4 intermediates to C_3 (Könneke et al., 2014). Thus, through the catalysis of PEP carboxykinase (Nmar_0292) using an ATP, net 7 ATP is used to convert oxaloacetate to PEP. In *N. maritimus*, PEP is converted to pyruvate via pyruvate: phosphate dikinase (Nmar_0951) and yields 2 ATP molecules, resulting in the consumption of net 5 ATPs. The synthesis of 2-oxoglutarate requires the condensation of oxaloacetate and acetyl-CoA, which necessitates the net usage of 10 ATP molecules (Könneke et al., 2014).

Table 1.1: Comparison of the moles of high-energy bonds of ATP required to form central metabolic precursors. *The numbers indicate the moles of high energy bonds of ATP required to form 1 mol of central metabolic precursor for each aerobic carbon fixation cycle (Könneke et al., 2014).*

Precursor (1 mol)	Calvin Benson Cycle	3-Hydroxypropionate Bicycle	HP/HB Cycle (Crenarchaeota)	HP/HB Cycle (Thaumarchaeota)
Acetyl-CoA	7	7	6	4
Oxaloacetate	8	9	9 or 10	6
PEP	8	9	10 or 11	7
Pyruvate	7	7	9 or 10	5
2-Oxoglutarate	15	16	15 or 16	10

1.10 Nitrosopumilus maritimus Possesses a Single Biotin Dependent Bifunctional Acetyl-CoA/Propionyl-CoA Carboxylase

The HP/HB cycle operates the carboxylation reactions of acetyl-CoA and propionyl-CoA using a single bifunctional biotin-dependent acetyl-CoA/propionyl-CoA carboxylase enzyme. The operation of this enzyme was previously identified in many of the members

of the Sulfolobacae family belonging to the phylum Crenarchaeoata (Chuakrut et al., 2003; Hügler et al., 2003). Later, a highly similar bifunctional carboxylase activity was also discovered in the species *N. maritimus* of the phylum Thaumarchaeota (Könneke et al., 2014). In Bacteria and Eukaryotes, the presence of acetyl-CoA carboxylases is associated with fatty acid synthesis (Moss & Lane, 1971). Since the Archaea do not possess fatty acids, it was later elucidated that this enzyme is involved in an autotrophic carbon fixation mechanism (Berg et al., 2010b). This was also validated with the enzyme assays, which demonstrated that the enzymatic activity of the acetyl-CoA/propionyl-CoA carboxylase increases when the cells are grown in autotrophic conditions. However, this activity dramatically decreases in heterotrophic growth conditions (Hügler et al., 2003; Chuakrut et al., 2003).

In the biotin-dependent carboxylases, the carboxylation of the substrates is a two-step reaction. The first reaction is initiated through a carboxylation reaction of a biotin cofactor via a biotin carboxylase (BC) subunit. This reaction requires a carbon-dioxide donor such as bicarbonate and the use of an ATP. A biotin carrier protein (BCCP) covalently links a biotin cofactor using a lysine residue and provides a reaction environment by translocating the biotin to a carboxyltransferase (CT) subunit. This process also requires the translocation of the BCCP, as the distance between the subunits is not sufficient for the lysine-linked biotin to reach the CT. In the second step, the carboxyl group is transferred to an acceptor molecule such as an acetyl-CoA and a propionyl-CoA (***Fig 1.9***) (Knowles, 1989; Atwood & Wallace, 2001; Tong, 2013).

The bifunctional acetyl-CoA/propionyl-CoA carboxylase is well characterized in the species *Metallosphaera sedula* and *Acidianus brierleyi.* Sequence alignment studies of these carboxylases indicate the highest sequence similarities with the equivalent enzymes from the species *S. tokadaii, S. metallicus, and S. solfataricus.* This single promiscuous enzyme catalyzes the conversion of both acetyl-CoA and propionyl-CoA into malonyl-CoA and methylmalonyl-CoA, respectively. Before the identification of this bifunctional carboxylase, two separate acetyl-CoA and propionyl-CoA carboxylase enzymes were thought to be involved in the catalysis of these reactions. However, the subsequent kinetic studies showed highly comparable enzymatic activities and K_m values to both of these substrates. Moreover, the Southern Blot analyses also revealed that only a set of genes is

responsible for the carboxylation of acetyl-CoA and propionyl-CoA (Hügler et al., 2003; Chuakrut et al., 2003).

Genomic studies show that the archaeal and bacterial biotin-dependent carboxylases share functional similarities; however, their structural features might differ. For instance, in *E. coli*, the acetyl-CoA carboxylase consists of four subunits encoded by the genes accAD and accBC (Cronan & Rock, 2008). Despite that, the archaeal acetyl-CoA/propionyl-CoA carboxylase contains three subunits: a biotin carboxylase (α), a carboxyltransferase (γ), and a biotin carrier protein (β). At the DNA level, these subunits are coded by the genes accC, accB, and pccB and are generally clustered together. On the other hand, some exceptions are present such as, the pccB gene is clustered separately in *M. sedula* from accC and accB genes (Hügler et al., 2003). Similar to the Crenarchaeal species, the Thaumarchaeal *N. maritimus* acetyl-CoA/propionyl-CoA carboxylase is composed of three subunits: the carboxyltransferase Nmar_0272, the carbamoyl phosphate synthase Nmar_0273 and the biotin/lipoyl attachment domain-containing protein Nmar_0274. These subunits form a holoenzyme structure to become a functional protein complex (Könneke et al., 2014). Importantly, there are no studies on the functional and structural characterization of N. maritimus acetyl-CoA/propionyl-CoA carboxylase, making this protein a significant target for structural studies.

The aim of this thesis is to unravel the structure of the carboxyltransferase subunit of *N. maritimus* Acetyl-CoA carboxylase/Propionyl-CoA carboxylase (Nmar_0272), to understand how this subunit forms a complex with the other subunits and interact with its substrates. Analyzing the structure of the carboxyltransferase subunit will shed light on the mechanism of the acetyl-CoA and propionyl-CoA carboxylation in *N. maritimus*. Additionally, understanding the mechanism of such energy-efficient carbon-fixation in *N. maritimus* will reveal how the single cellular marine organisms can maintain their life-cycle and the fixation of the inorganic compounds in extremely nutrient-poor environments such as open oceans. Furthermore, thanks to the advancement of today's synthetic biology, the enzymes that are used in this mechanism can be used for the synthesis of industrial and pharmaceutical compounds in a way that requires less energy (Erb et al.,2011; Yuzawa et al., 2012; Keller et al. 2013). For instance, the products of the carboxylation reactions are involved in the biosynthesis of polyketides. Some of these

products are known to be used for the synthesis of pharmaceutical compounds such as antibiotics, anti-cancer drugs, immunosuppressants, and blood-pressure-lowering compounds (Cane et al., 1998; Diacovich et al., 2004). Therefore, studying the structure and characteristics of this protein is of great importance for future clinical studies as well.

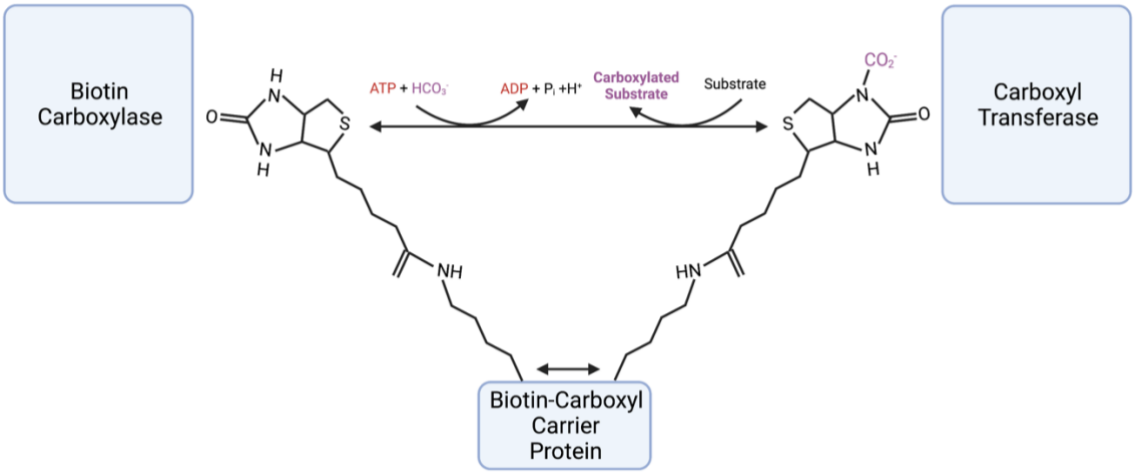


Figure 1.9: Representation of the general mechanism of biotin-dependent carboxylases. *The top reciprocal arrow indicates the interconversion between biotin and carboxybiotin. The bottom reciprocal arrow represents the transfer of biotin and carboxybiotin between the subunits through biotin-carboxyl carrier protein. The blue boxes represent the subunits.*

1.11 Protein Crystallization

The initial step of X-ray Crystallography is the crystallization of the proteins. In order to obtain high-quality protein crystals, a soluble protein with high purity and concentration is required. The main objective of this method is to force the proteins to crystallize by providing different crystallization conditions. These conditions are highly variable as each protein has distinct characteristics. Moreover, the crystallization process is affected by many parameters such as the concentration of the protein solution, the type of the buffer and the precipitant, the pH of the buffer, the temperature of the experimental environment, and the method used for crystallization. Additionally, the presence of flexible and motile regions and some of the post-translational modifications such as glycosylation negatively affect the crystal formation (Smyth & Martin, 2000; Song & Gouaux, 1997). In order to identify the best conditions for crystallization, special agents called the sparse matrix screens containing a variety of buffers, precipitants, salts, and pH are used (Smyth & Martin, 2000).

Finding a suitable crystallization technique is also of great importance for protein crystallization. Microbatch and vapor diffusion techniques are the most widely used methods used to obtain protein crystals. In the microbatch method, small drops of the protein solution and the crystallization screening condition are mixed in equal volumes in microbatch well plates and sealed with paraffin oil droplets. This process is repeated for each screening solution. Since the density of the paraffin oil is low, the protein solution stays under the paraffin oil, preventing rapid evaporation and possible contamination.

There are two vapor diffusion techniques called the sitting drop and the hanging drop. In both of these techniques, a drop containing equal volumes of the crystallization agent and the protein solution is placed in a hanging or sitting position over a reservoir solution containing the crystallization agent. The wells containing the reservoir solutions and the sample drops are then sealed with a coverslip, and the drops are incubated inside these wells. Through the diffusion of the water from the droplets, the concentrations of the protein solution and the crystallization agent change in time. This concentration change forces the protein molecules to come out of the solution, leading to crystallization (***Fig 1.10***). (Smyth & Martin, 2000; Chayen & Saridakis, 2008; McPherson, 2009)

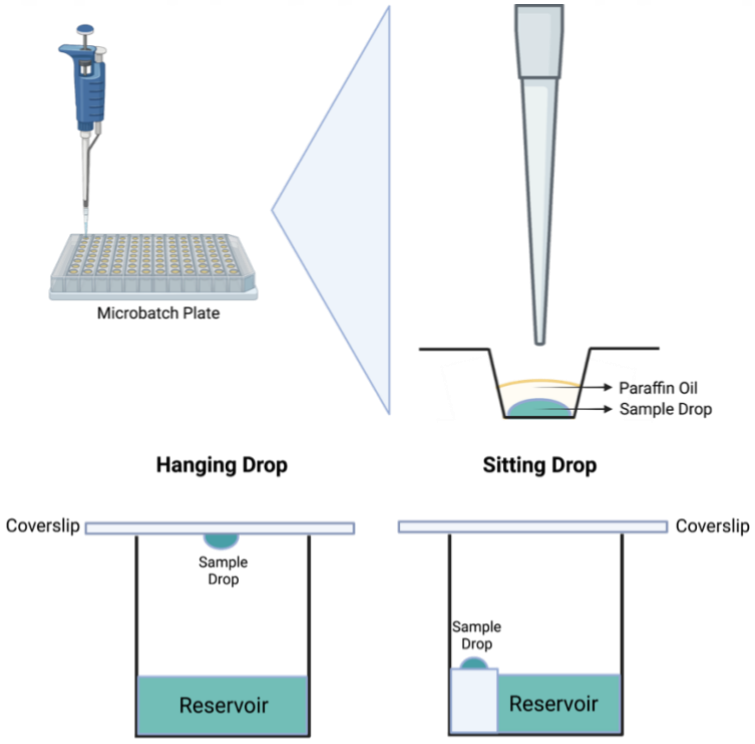


Figure 1.10: Representation of the microbatch (top) and the vapor diffusion (bottom) crystallization techniques. *Sample drop is composed of equal volumes of the protein solution and the crystallization screening agent.*

1.12 Structural Determination of the Proteins with X-Ray Crystallography

X-Ray crystallography is one of the most common methods used to determine the three-dimensional structures of the macromolecules and the proteins. This method is based on the ability of the X-rays to interact with the atoms and the electrons within the molecules thanks to their small wavelengths. However, for a single molecule, this interaction is quite destructive as the intensity of the ionic radiation can easily cause damage to the molecule. Crystals are composed of vast number of copies of identical molecules packed in repetitive unit cells. The unit cells are the smallest elements of the crystals, and their combination forms a highly ordered crystal lattice. This repeating feature makes them great candidates for X-ray diffraction studies. In the X-ray crystallography approach, the crystals are hit by monochromatic X-rays, and as a result, specific diffraction patterns called the diffraction images are obtained by a detector. The pattern of the diffraction depends on the arrangements of the atoms within the crystal lattice. Therefore the data obtained from these diffractions can be processed to interpret the electron densities and the structural properties of these molecules. (Smyth & Martin, 2000; McPherson, 2009)

In order to determine a crystal structure, the data obtained from the diffractions are processed by specific software and algorithms (***Fig. 1.11***). The information acquired from the diffraction images is utilized to determine the resolution, the space group, unit cell dimensions, and the crystal system. The resolution is positively correlated with the number of diffraction spots that are located near the edge of the diffraction image, and it indicates the data quality. If the majority of the diffraction data are consistent, a high resolution protein structure is obtained. In some cases, the flexible and the motile regions of the proteins result in less consistent diffractions, thereby decreasing the resolution. According to the shape of the unit cells, seven crystal systems have been identified: monoclinic, triclinic, trigonal, tetragonal, cubic, orthorhombic, and hexagonal. The length of the unit cells is indicated with the letters a, b, c, and the angles are shown with α, β, and γ. The space groups are used to represent the molecular packing and the symmetry within the crystals. There are a total of 230 space groups; however, due to the chirality of the amino acids, fewer space groups are present for the proteins. (Smyth & Martin, 2000; McPherson, 2009)

Once the crystals are exposed to the monochromatic X-rays, the detector measures the diffracted wavelength and the intensity (amplitude) of the radiation. Some of the X-rays are scattered constructively once they are in identical phases and result in higher amplitudes. Despite that, some of them are scattered with opposite destructive phases and leading to lower intensities. Unfortunately, the detectors are not able to evaluate the phases of the diffracted X-rays. This gives rise to one of the most fundamental problems of X-ray crystallography, called the phase problem. The phase problem can be solved by using the isomorphous replacement or the molecular replacement methods. Isomorphous replacement is applied when there is no comparable structure model is present. In this method, the protein crystals are soaked in a heavy atom-containing solution and compared with the native protein crystals. The differences between the two data sets are used to determine the structure. In contrast, the new protein model is constructed using a comparable protein model or a protein with a high sequence identity in the molecular replacement method. (Smyth & Martin, 2000; Cowtan, 2003)

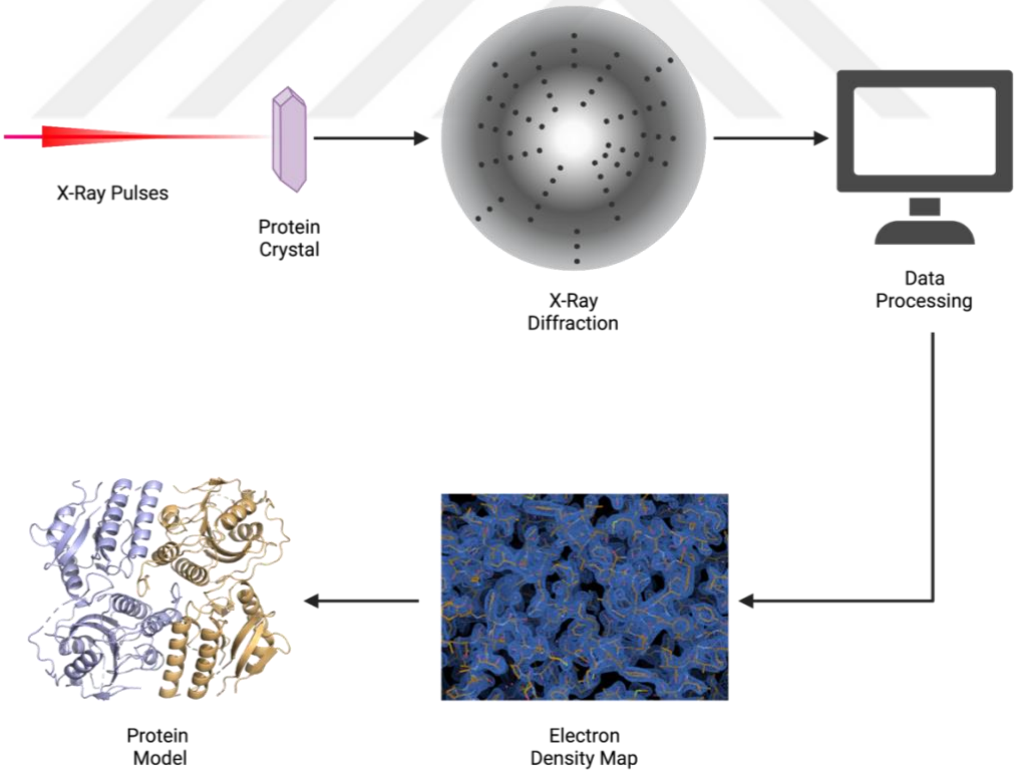


Figure 1.11: Schematic representation of the general steps of X-ray crystallography. *The diffraction data collected from the crystals is processed by certain algorithms. The resulting electron density map is used for the determination of the protein model.*

Chapter 2:
MATERIALS AND METHODS

2.1 Construct Design and Cloning

The following amino acid sequences were used to design the constructs for native *Nitrosopumulus maritimus* Acetyl-CoA/Propionyl-CoA carboxylase subunits Nmar_0272, Nmar_273 and Nmar_274. The synthesis of the genes was provided by Genscript, USA, and the insert sequences were cloned into separate pET28a(+) vectors containing a hexahistidine tag and a thrombin cleavage site. The cloning was performed by using NdeI and BamHI restriction endonucleases from 5' to 3' ends. In the following amino acid sequences, the hexahistidine tags and the thrombin cleavage sites were indicated with orange and blue colors, respectively. The stop codon is indicated with the red asterisk.

Nmar_0272:

MGSSHHHHHHSSGLVPR(thrombin_cut_site)GSHMMHSEKLENYNNKHKTSQQG
GGQDRIKAQHDKGKLTARERIDLLLDEGSFTEIDPMVTHHYHEYDMQKKKFFT
DGVVGGYGNVNGRQIFVFAYDFTVLGGTLSQMGAKKITKLMDHAVRTGCPVI
GIMDSGGARIQEGIMSLDGFADIFYHNQLASGVVPQITASIGPSAGGSVYSPAMT
DFVVMVEKAGSMFVTGPDVVKTVLGEEISMDDLGGAMTHGSKSGVAHFVAQ
NEYECMDYIKKLISYIPQNNSEEPPKIKTDDDPNRLDNNLINVIPENPLQPYDMK
EIINSIVDNHEFFEVHELFAPNIVVGYARMDGQVVGIIANNPMHLAGALDIDSSN
KSARFIRFCDAFNIPIITLVDTPGYMPGSNQEHNGIIRHGSKLLYAYCEATVPRIT
LVIGKAYGGAYIAMGSKNLRTDINYAWPTARCAVLGGEAAVKIMNRKDLADA
DNPEELKKKLIDEFTEKFENPYVAASHGTVDNVIDPAETRPMLIKALKMLANKR
EKQLPRKHGNINL*

Nmar_0273:

MGSSHHHHHHSSGLVPR(thrombin_cut_site)GSHMMIEKVLIANRGEIALRVIRTC
NALGIKTVAVYSDEDYNSLHVKKADESYHIGEAAPAKSYLNQEKILEVMLSSG
ADAVHPGYGFLSENDDFARLCEKNKINFIGPSADSMNLCGDKMECKAAMLKA
QVPTVPGSPGLVDTAEEAEKIANEIGYPVLLKSVYGGGGRGIRLVTTDQELREG
FETVTSESIAAVGKSAIIVEKFLEKTRHIEYQMCRDHHGNAVHLFERECSIQRRN
QKLIEQTPSPVVDEAKREEIGELVVKAAEAVNYTNLGTAEFLRADNGEFYFIEIN

ARLQVEHPISEMVSGLDFVKLQIDIANGETLPFKQKDLKMNGYAIECRINAEDT
FLDFAPSTGPVPDVTIPAGPNVRCDTYLYPGCTVSPFYDSLMAKLCTWGPTFEE
SRTRMLTALNDMYVQGVETSIPLYKTILNSEEYKNGELSTDFLKRYGMIDKLSE
DLKKEKEDKSEAALAAAIIHSEYFKNRVQNDNASSATWKNKLD*

Nmar_0274:

MGSSHHHHHHSSGLVPR(thrombin_cut_site)GSHMMDYKIADVEKSFEGKITENL
GNNDYVIKINDKEHQLKILSMNAKGIEFILDQQYHKAKYLETATNEMNLVIDN
VPVTLNMNTHFDEIVYKNSGGGGAGGAQVALKSQIPGKVVSIAVAEGDSVKK
GDVVCTLESMKMQVGIKAHKDGEVKNLKIKEGATVAKGDVIADLE*

2.2 Transformation and the Preparation of Glycerol Stocks

The transformation of the Nmar_0272, Nmar_0273, and Nmar_0274 constructs was performed by using the *E. coli* strain BL21 Rosetta 2. The transformed colonies were taken and inoculated in separate 150 ml LB-Miller media (30 g/L) supplemented with 150 µl of both chloramphenicol (35 µg/ml) and kanamycin (50 µg/ml). The cells were then incubated overnight in New Brunswick Innova 4430R incubator at 37 °C and 110 rpm. 750 µl of each of the overnight grown cultures were transferred in separate 2 ml Wuxi NEST Biotechnology cryovials and mixed with 750 µl 80% glycerol. The samples were placed and stored in a -80 °C freezer.

2.3 Small Scale Protein Expression and Pulldown Assay

The glycerol stocks containing the NMAR_0272, NMAR_0273, and NMAR_0274 constructs were inoculated in independent 100 ml LB-Miller media (30 g/L) which contains 35 µg/ml chloramphenicol and 50 µg/ml kanamycin. The same procedure was applied to NMAR_0274 with a rich LB-Miller medium (50 g/L). The cells were incubated at 37°C and 110 rpm in the New Brunswick Innova 4430R incubator. The protein expression of each culture was induced with IPTG with a final concentration of 0.4 mM once the OD600 reached approximately 0.8. After the addition of IPTG, the cultures were incubated at 18°C overnight. The next day, the cells were centrifuged with Beckman Allegra 15R centrifuge at 3500 rpm for 20 minutes, and the pellets were collected. The pellets were resuspended with a lysis buffer containing 500 mM NaCl, 150 mM Tris pH 7.5, 0.1% Triton X-100 and 5% glycerol and sonicated with Branson W250 Sonifier. The

lysate was centrifuged at 15000 rpm for 30 minutes using Beckman Avanti J-26S centrifuge. The supernatant was collected, and the pellets were resuspended in 20 mL lysis buffer (500 mM NaCl, 150 mM Tris pH 7.5, 0.1% Triton X-100, 5% glycerol). 300 µl Ni-NTA agarose resin (QIAGEN, USA) was added to each supernatant, and the samples were incubated overnight at 4 °C. The next day the bead containing supernatant was centrifuged at 1000 rpm for 2 minutes with Beckman Allegra 15R centrifuge, and the supernatant was discarded. The beads were washed with a washing buffer containing 150 mM NaCl and 20 mM Tris pH 7.5. The washed resin was centrifuged at 1000 rpm for 2 minutes with Beckman Allegra 15R centrifuge, and the supernatant was discarded. The beads were eluted with 1 mL HisB elution buffer containing 150 mM NaCl, 20 mM Tris pH 7.5, and 250 mM imidazole. The expression of the proteins was validated with BioRad TGX-mini protean gradient SDS-PAGE.

2.4 Large Scale Protein Expression and Purification

The glycerol stocks containing the target genes NMAR_0272 and NMAR_0274 were inoculated in separate 150 ml LB-Miller media (30 g/L) supplemented with 35 µg/ml chloramphenicol and 50 µg/ml kanamycin. The cultures were incubated at 37°C and 110 rpm overnight in the New Brunswick Innova 4430R incubator. On the following day, the NMAR_0272 culture was split into equal volumes of six 2 L LB-Miller media (30g/L), which contains the same concentrations of chloramphenicol (35 µg/ml) and the kanamycin (50 µg/ml) (Total 12 L). For NMAR_0274, a rich LB- Miller medium (50 g/L) was used instead. Once the OD600 reached between 0.8 and 1.2, the protein expression was induced by adding IPTG with the final concentration of 0.4 mM. Following the induction, the induced cells were incubated at 18 °C overnight. The cultures were centrifuged at 3500 rpm and 4°C for 30 minutes using Beckman Allegra 15R centrifuge. The supernatant was discarded, and the pellet samples were collected and resuspended with lysis buffer containing 500 mM NaCl, 150 mM Tris pH 7.5, 0.1% Triton X-100 and 5% glycerol. The cells were lysed via sonication using Branson W250 Sonifier, and the lysate was centrifuged at 40000 rpm and 4 °C by Beckman OptimaTM L-80 XP Ultracentrifuge for 1 hour with Beckman Ti45 rotor. The supernatant was collected, and the pellet was resuspended with the lysis buffer. For protein purification, the nickel affinity chromatography procedure was applied using the QIAGEN Ni-NTA

agarose resin column. The column was equilibrated with a 50% HisA (150 mM NaCl and 20 mM Tris pH 7.5) and 50% HisB (150 mM NaCl, 20 mM Tris pH 7.5, 250 mM imidazole) buffers. Following this step, the supernatant was loaded into AKTA Prime FPLC with a 2.5 ml/min flow rate. After the loading step, the column was washed with HisA buffer (150 mM NaCl and 20 mM Tris pH 7.5). For Nmar_0272, in order to prevent non-specific binding, the column was washed with 5% HisB buffer (150 mM NaCl, 20 mM Tris pH 7.5, 25 mM imidazole) and 10% HisB buffer (150 mM NaCl, 20 mM Tris pH 7.5, 50 mM imidazole) respectively. This step was followed by the elution of the proteins from the column using 100% HisB buffer (150 mM NaCl, 20 mM Tris pH 7.5, 500 mM imidazole). The fractions were collected, and the expression of the proteins was approved via BioRad TGX-mini protean gradient SDS-PAGE. The same procedure was applied to Nmar_0274 using a HisB buffer with different imidazole concentrations (150 mM NaCl, 20 mM Tris pH 7.5, 250 mM imidazole), and only 10% HisB buffer was used to prevent the non-specific binding. Following the purification, the purified Nmar_0274 was dialyzed using a 3 kDa dialysis membrane in HisA buffer (150 mM NaCl, 20 mM Tris pH 7.5) to remove the excess imidazole. The purified samples were aliquoted in 1.5 ml Eppendorf tubes and stored at -80°C. Half of the Nmar_0274 sample was reserved for reverse Ni-NTA chromatography.

2.5 Reverse Ni-NTA Chromatography

The N-terminal hexahistidine tag of Nmar_0274 was cleaved by using thrombin protease (Sigma, USA). Following the thrombin addition, the protein sample was incubated for 5 days at 4 °C. In order to remove the hexahistidine-tags, the cleaved Nmar_0274 solution was loaded into AKTA Prime FPLC with the flow rate of 2.5 ml/min, and the reverse Ni-NTA procedure was applied using QIAGEN Ni-NTA agarose resin column. The flowthrough containing the unbound and hexahistidine tag-free Nmar_0274 fraction was collected. The bound fraction was eluted using HisB buffer containing 150 mM NaCl, 20 mM Tris pH 7.5, and 250 mM imidazole and collected in a separate 50 ml Falcon tube. The presence of the cleaved Nmar_0274 protein was validated by using BioRad TGX-mini protean gradient SDS-PAGE. Following the validation, the untagged Nmar_0274 protein samples were aliquoted in 1.5 ml Eppendorf tubes, and the samples were stored at -80°C.

2.6 Protein Crystallization

The crystallization experiments were performed by applying microbatch crystallization under the paraffin oil method at ambient temperature. The non-cleaved Nmar_0272 (7.01 mg/ml), Nmar_0274 (4.64 mg/ml) and the thrombin cleaved Nmar_0274 (0.793 mg/ml) samples were crystallized independently. Using 72 well Terasaki crystallization plates (Greiner-Bio, Germany), equal volumes (0.83µl) of the purified protein samples and approximately 3500 sparse matrix crystal screening agents were mixed in separate wells. The mixtures were covered with 16 µl 100% paraffin oil for each well. For non-cleaved Nmar_0272, the crystallization experiments were replicated for the conditions in which the best crystals were obtained. According to the quality of the crystals, NeXtal Procomplex I (NeXtal Biotechnologies, USA), condition #24, Clear Strategy Screen pH 4.5 (Molecular Dimensions, UK), condition # 16, and Index II (Hampton Research, USA) condition #96 crystallization screens were chosen for larger-scale crystallization. For the large scale crystallization 50 µl of each condition were mixed with 50 µl of protein solution in separate 1.5 ml Falcon tubes. Additionally, the crystals obtained in the replication plates were collected and transferred into separate 1.5 ml Eppendorf tubes. All the crystal samples were stored at ambient temperature.

2.7 Gel Filtration

The gel filtration experiments were performed for the non-cleaved Nmar_0272 (7.01 mg/ml) and non-cleaved Nmar_0274 (4.64 mg/ml) protein solutions. The Nmar_0272 and Nmar_0274 solutions were mixed in equal volume (100µL each) and incubated overnight at 4°C. The following day, gel filtration procedure was applied to Nmar_0272, Nmar_0274, and the Nmar_0272/Nmar_0274 mixture independently. The gel filtration of each sample was performed by using AKTA go chromatography system. The individual samples were loaded into Superdex 200 Increase 10/300 GL (Cytiva, USA) column and run with 0.5 ml/min flow rate.

2.8 Data Collection and Processing

The X-ray diffraction experiments were performed in Stanford Synchrotron Radiation Lightsource (SSRL) at SLAC National Accelerator Laboratory in Menlo Park, CA, USA.

The diffraction data of Apo-Nmar_0272 was collected at the beamline BL12-2 using the Pilatus 6M detector at 2.26 Å resolution. The unit cell dimensions were a=113.46 Å, b= 113.46 Å, c= 144.64 Å, α= 90, β= 90 γ= 90 with P63 space group. The experiments were conducted with 0.9794 Å wavelength at 100 °K. The diffraction data's indexing and scaling were achieved by using the XDS and XSCALE software packages (Kabsh, 2010).

2.9 Structure Determination, Refinements, Visualization and Modeling

The phasing of the apo-Nmar_0272 dimer was performed with the molecular replacement extension PHASER (McCoy et al., 2007) of the PHENIX software (Adams et al., 2010). A previously determined structure (PDB:1X0U) was used as a template search model. Based on the model, the initial rigid body and simulated annealing refinements were performed with PHENIX software. The refinement of the TLS parameters and the individual coordinates were also conducted with the PHENIX software. The altered side chains and the missing water molecules were controlled and fixed. Also, the residues and water molecules with no electron density were removed manually with the program COOT (Emsley & Cowtan, 2004). A template model based on 1X0U was generated to fill the missing residues of the apo-Nmar_0272 structure using the webserver Swiss-Model (www.swissmodel.expasy.org). The common residues with the original model were deleted from the template model, and the two PDB files were merged using COOT. The Ramachandran statistics (favored/ allowed/outliers) of the apo-Nmar_0272 structure are 96.39 / 3.29 / 0.31 % respectively. The obtained R_{work} and R_{free} values after the final refinement are 0.22 and 0.28, respectively. The structural alignments were conducted using the program PyMOL (www.schrodinger.com/pymol). The root means square deviation (RMSD) values were determined according to Cα atoms. The multiple sequence alignments of the Nmar_0272 and its related homologs were performed with the Jalview software (Clamp et al., 2004) and its ClustalW algorithm (Thompson et al., 2003). The Nmar_0272 structure predictions were performed using the Deep Learning based structural prediction tools AlphaFold and RoseTTAFold (Jumper et al., 2021; Baek et al., 2021). All the figures were generated by the molecular visualization program PyMOL (www. schrodinger.com/pymol) and the webserver BioRender (www.biorender.com).

Chapter 3:
RESULTS

3.1 SDS-PAGE Results

3.1.1 Nmar_0272 Large Scale Expression

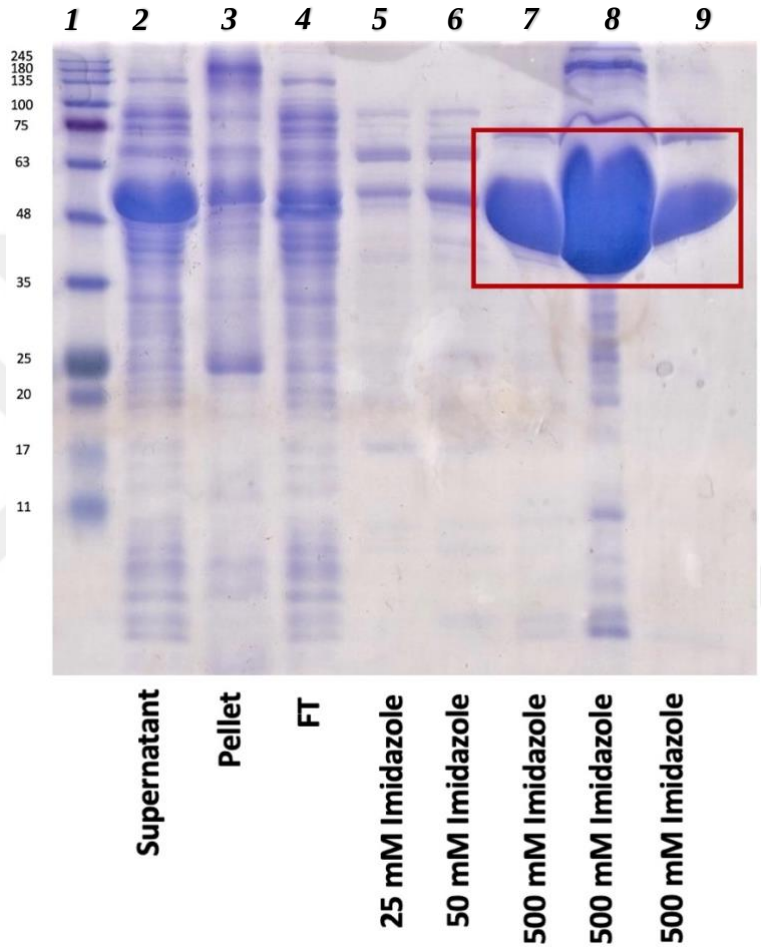


Figure 3.1.1: The SDS-PAGE gel image of hexahistidine tagged Nmar_0272 protein soluble fraction. *The protein bands were visualized with Coomassie Brilliant Blue. The first lane is the protein ladder, where the numbers next to the first lane indicate the molecular weights as kDa. The 2nd lane is the supernatant sample of the whole cell lysate, the 3rd lane is the pellet sample resuspended with lysis buffer, and the 4th lane is the flowthrough sample. Lanes 5, 6, 7, 8, and 9 are the elution samples, eluted with 25mM, 50 mM, and 500mM imidazole containing HisB buffers, respectively. The 56 kDa Nmar_0272 protein elution bands are indicated with the red frame.*

The purification of the Nmar_0272 protein containing a hexahistidine tag was conducted with Ni-NTA affinity chromatography. Nmar_0272 has an estimated size of 56 kDa. Therefore the bands between the 48 and 63 kDa markers indicate the presence of Nmar_0272. The supernatant sample contained the proteins in the whole cell lysate after ultracentrifugation; therefore, multiple protein bands, including the Nmar_0272, were observed in lane 2 as expected. To understand whether the protein was soluble, the pellet

sample was used as a control. A low amount of protein in the pellet sample and a significantly high amount of protein in the elution bands clearly indicated that the Nmar_0272 was highly soluble. The flowthrough sample was used to validate the binding of the proteins to the Ni-NTA column. Only a small amount of protein in the flowthrough sample validated the proper binding of Nmar_0272 to the Ni-NTA column. The 25 mM and 50 mM imidazole containing HisB buffers were used to prevent non-specific binding. The HisB buffer containing 500 mM imidazole was used to elute the protein sample from the Ni-NTA column. As indicated with the red frame, a high amount of Nmar_0272 was observed between the 48 and 63 kDa bands in lanes 7, 8, and 9, validating the presence of Nmar_0272 (***Fig 3.1.1***).

3.1.2 Nmar_0274 Large Scale Expression

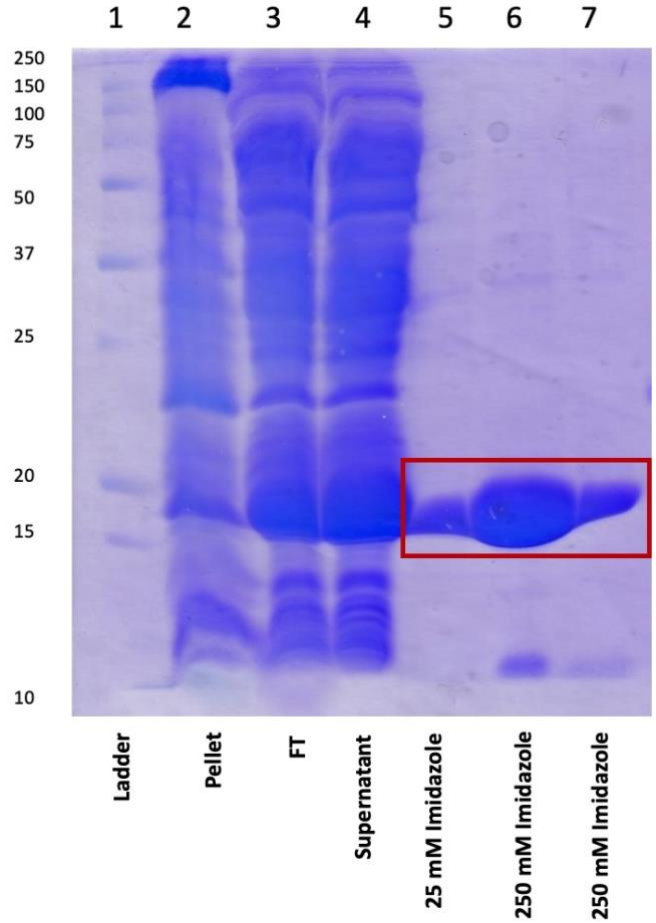


Figure 3.1.2: The SDS-PAGE image of hexahistidine tagged Nmar_0274 soluble fraction. *For the visualization of the protein bands, Coomassie Brilliant Blue was used. The first lane and the numbers on its left indicate the molecular weights in kDa. The 2nd lane is the pellet sample resuspended with lysis buffer. The 3rd lane belongs to the flowthrough sample, and the 4th lane is the supernatant of the whole cell lysate. Lanes 5, 6, and 7 are the elution samples, in which 25mM and 250mM imidazole containing HisB buffers were used, respectively. The obtained protein bands are indicated with a red frame.*

The hexahistidine tagged Nmar_0274 protein was purified using the Ni-NTA affinity chromatography method. Nmar_0274 has an estimated size of 18.5 kDa. Hence the protein bands between the 15 and 20 kDa ladder markers correspond to Nmar_0274 protein. In the pellet sample, only a small amount of Nmar_0274 was observed compared to the elution samples. This validates that the protein sample is highly soluble. The flowthrough sample was used to detect if the proteins were properly bound to the Ni-NTA resin. The thick band between the 15 and 20 kDa bands in lane 3 indicates that a portion of the Nmar_0274 was lost during the purification process. Since the supernatant sample contains the solution of the whole cell lysate after ultracentrifugation, multiple protein bands, including the Nmar_0274 were detected as expected. The first elution sample containing the 25 mM imidazole in lane 5 was used to prevent the non-specific binding of the other proteins to the resin. Instead of the elution of non-specific proteins, a small amount of Nmar_0274 protein was observed in this elution sample. The complete elution samples were obtained from the HisB buffer containing 250 mM imidazole as indicated in lanes 6 and 7. The thick bands indicated with the red frame show that a sufficient amount of soluble His-tagged Nmar_0274 was obtained from these fractions (***Fig 3.1.2***).

3.1.3 Nmar_0274 Reverse Ni-NTA Chromatography

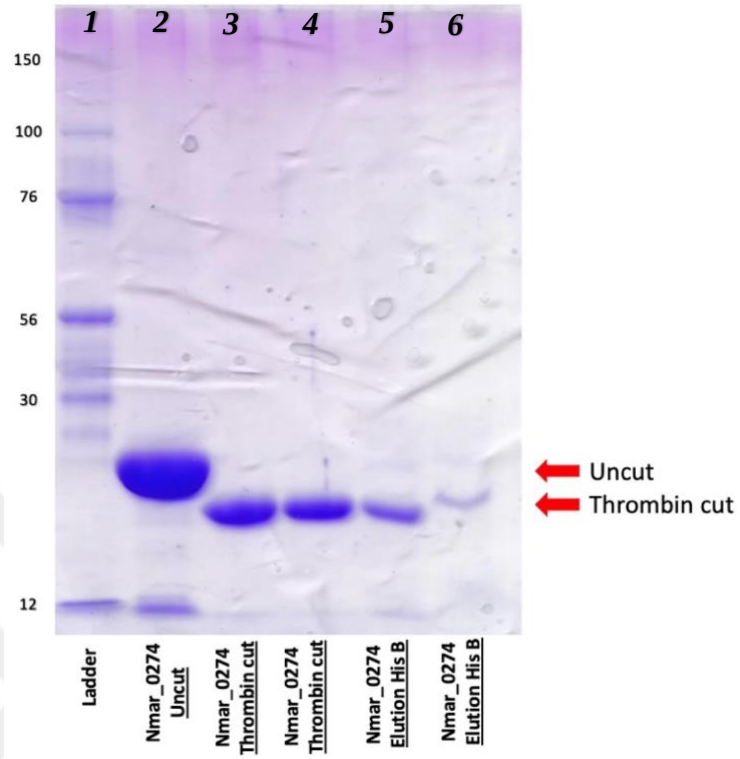


Figure 3.1.3: SDS-PAGE image of the reverse Ni-NTA chromatography of Nmar_0274. *Coomassie Brilliant Blue was used for the visualization of the bands. Lane 1 and the numbers left to the lane are the protein ladder and their molecular weights in kDa, respectively. The 2nd lane is the non-cleaved native Nmar_0274. Lanes 3 and 4 are the thrombin cleaved Nmar_0274 proteins. Lane 5 and 6 are the elution fractions of the Nmar_0274 samples with the HisB buffer containing 250 mM imidazole. The shift of the uncut and the thrombin cut protein bands are indicated with red arrows.*

Reverse Ni-NTA chromatography was performed to cleave the hexahistidine tag from the Nmar_0274 solution. Following the incubation of the Nmar_0274 containing hexahistidine tag with thrombin protease for 5 days, the protein solution was loaded into Ni-NTA resin. The unbound thrombin cut fraction was collected for crystallization experiments. The 2nd lane belongs to the Nmar_0274 uncut fraction obtained from the large-scale purification and used as control. Lanes 3 and 4 belong to the samples that are collected after 5 days of thrombin incubation. After collecting cleaved proteins, the resin was washed with HisB buffer containing 250 mM imidazole to elute possible remaining hexahistidine tagged Nmar_0274 bound to the Ni-NTA resin. The gel picture indicates a clear band shift between the uncut Nmar_0274 and thrombin cut Nmar_0274. The band shift is indicated with the red arrows in **Fig. 3.1.3**. In lanes 5 and 6, a low amount of remaining cleaved Nmar_0274 was observed, whereas no hexahistidine tagged protein was detected. These results validate that the Nmar_0274 protein was successfully cleaved by thrombin protease.

3.2 Crystallization Results

3.2.1 Nmar_0272 Crystals

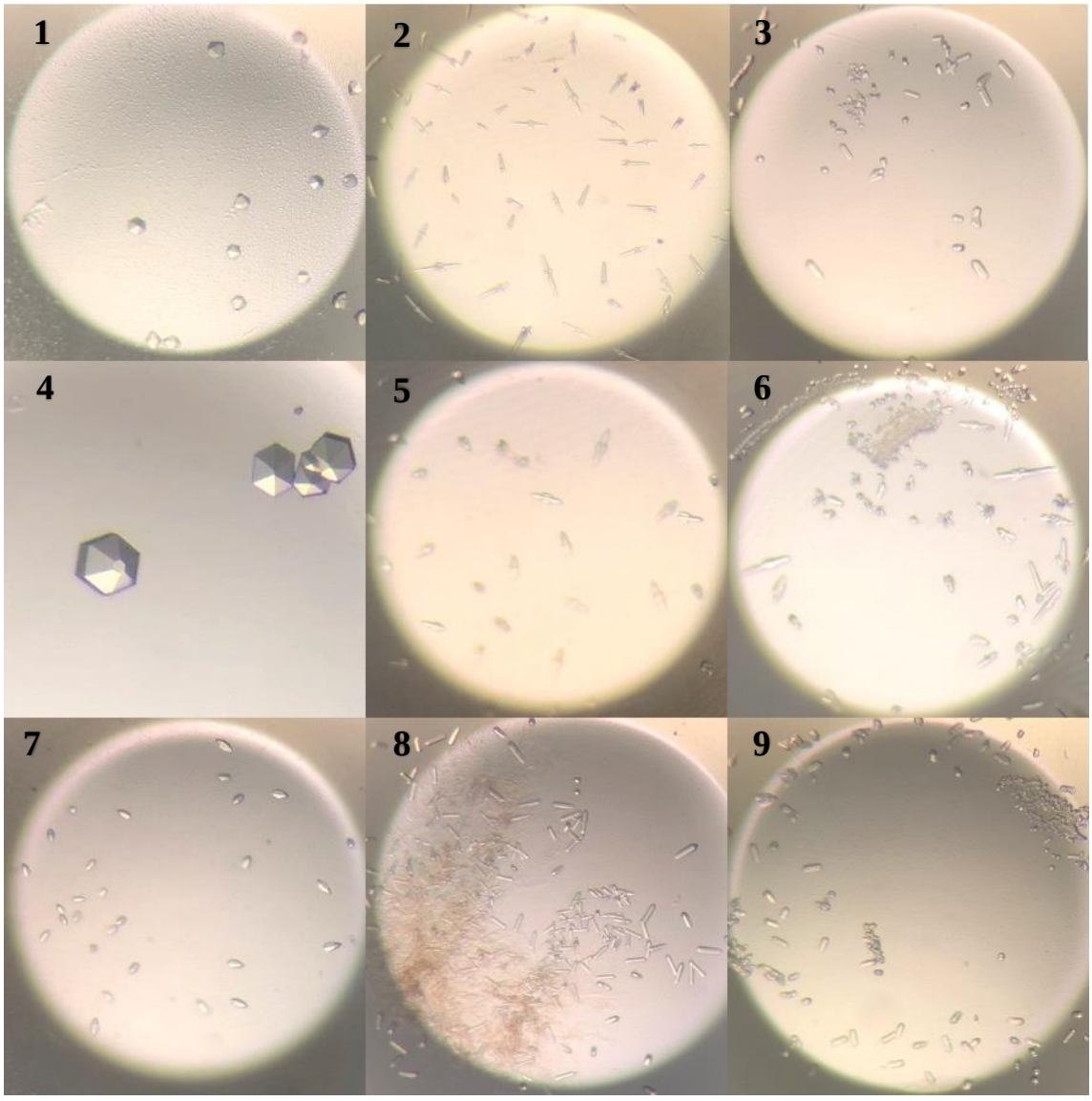


Figure 3.2.1: The microscope picture of His-tagged Nmar_0272 protein crystals obtained from microbatch under oil crystallization at ambient temperature. *The crystals were obtained from the sparse matrix crystallization screening conditions indicated in* ***Table 3.2.1****. The numbers correspond to the sparse matrix crystallization conditions that are shown in* ***Table 3.2.1****.*

The crystallization experiment of hexahistidine tagged Nmar_0272 proteins was carried out with microbatch crystallization under the paraffin oil method. For the experiment, about 3500 sparse matrix crystallization screening conditions were applied to the protein solutions. All the experiments were performed at ambient temperature. Once the crystals were obtained, they were placed in a 4°C cold room. The best crystals were obtained from

9 of the following crystallization screening conditions, respectively. The condition numbers are indicated with the symbol (#): Index II (Hampton Research) # 48, NR-LBD I (Molecular Dimensions) #5, NR-LBD I # 16, NR-LBD II # 6, Nextal Procomplex I (NeXtal Biotechnologies) #16, Nextal Procomplex I #24, Nextal Procomplex II #3, Clear Strategy Screen (Molecular Dimensions) pH 4.5 #16, Clear Strategy Screen pH 6.5 #15. The salt, buffer, and precipitant ingredients of these crystallization screening conditions are listed in **Table 3.2.1** according to crystal numbers indicated in **Fig. 3.2.1**.

Table 3.2.1: The salt, buffer, and precipitant ingredients of the sparse matrix crystallization conditions. *The ingredients in which the best Nmar_0272 crystals were obtained. The number column corresponds to the crystal numbers that are indicated in Fig. 3.2.1.*

Number	Crystallization Condition	Salt	Buffer	Precipitant
1	Index II #48	0.2M Calcium chloride dihydrate	0.1M Bis-Tris pH 5.5	45% v/v (+/-)-2-Methyl-2,4-pentanediol
2	NR-LBD I #5	None	0.1 M Tris pH 8.0	9% w/v PEG 8000
3	NR-LBD I #16	0.4 M Sodium chloride	0.1 M Tris pH 8.0	19% w/v PEG 4000
4	NR-LBD II #6	0.2 M Sodium Chloride	0.1 M Bis-Tris pH 6.5	19% w/v PEG 2000 MME
5	Nextal Procomplex I #16	0.2 M Sodium chloride	0.1 M MES pH 6.5	10% (w/v) PEG 4000
6	Nextal Procomplex I #24	0.1 M Magnesium chloride	0.1 M HEPES pH 7	15% (w/v) PEG 4000
7	Nextal Procomplex II #3	0.2 M Sodium chloride	0.1 M Sodium cacodylate pH 6.0	8% (w/v) PEG 8000
8	Clear Strategy pH 4.5 #16	0.2 M Potassium bromide	10% PEG 8000+	10% (w/v) PEG 1000
9	Clear Strategy pH 6.5 # 15	0.2 M Magnesium chloride	10% PEG 8000+	10% (w/v) PEG 1000

3.2.2 Nmar_0274 Crystals

Figure 3.2.2: The microscope picture of hexahistidine tagged Nmar_0274 protein crystals. *The crystals were obtained by microbatch under paraffin oil crystallization method at ambient temperature.* ***Table 3.2.2*** *indicates the sparse matrix crystallization condition used to obtain Nmar_0274 protein crystals.*

The crystallization experiments of purified hexahistidine tagged Nmar_0274 protein were performed with microbatch under paraffin oil crystallization procedure. For the crystallization of the thrombin cleaved Nmar_0274, half of the (~35 ml) purified fraction of hexahistidine tagged Nmar_0274 protein solution was incubated with thrombin protease. Following the incubation, reverse Ni-NTA chromatography was applied, and the cleaved fraction of Nmar_0274 proteins was collected and crystallized with microbatch under paraffin oil method. Approximately 3500 sparse matrix crystallization conditions were applied to both hexahistidine tagged and thrombin cleaved Nmar_0274 solutions, and all the crystallization process was carried out at ambient temperature. The obtained crystals were stored at 4°C. Hexa-histidine tagged Nmar_0274 protein crystals were obtained from a single crystallization screening condition: Wizard Cryo II #9 (***Fig. 3.2.2***). However, no protein crystals were obtained from thrombin cleaved Nmar_0274 samples. The salt, buffer, and precipitant ingredients of the condition are listed in **Table 3.2.2**.

Table 3.2.2: The salt, buffer, and precipitant ingredients of the sparse matrix crystallization screening condition *in which the thrombin cut Nmar_0274 crystals are obtained.*

Crystallization Condition	Salt	Buffer	Precipitant
Wizard Cryo II #9	Zinc Acetate 50 mM	Imidazole/HCl	Propanol

3.3 Gel Filtration of Nmar_0272 and Nmar_0274

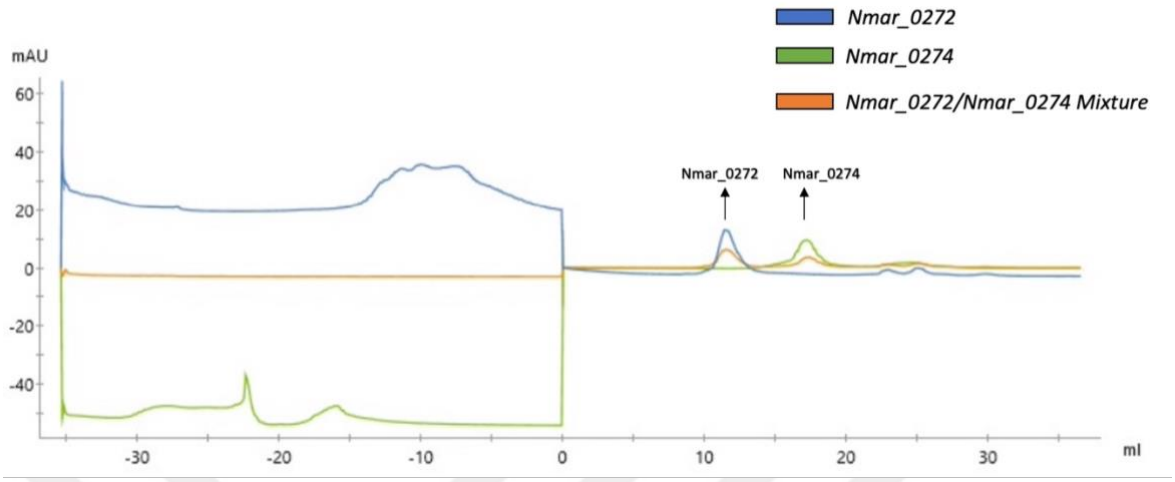


Figure 3.3.1: The gel filtration graph of Nmar_0272 (blue), Nmar_0274 (green) and Nmar_0272/Nmar_0274 mixture (orange). *The peaks that belong to the Nmar_0272 and Nmar_0274 proteins are indicated with arrows.*

In order to detect possible protein-protein interactions between the purified non-cleaved Nmar_0272 and Nmar_0274 subunits, gel filtration was performed. For the experiment, Nmar_0272, Nmar_0274, and Nmar_0272/Nmar_0274 mixture were loaded onto a Superdex 200 gel filtration column and run independently. The blue and the orange graph indicate the first and the second runs of the experiment in which only the Nmar_0272 protein and Nmar_0274 were loaded respectively. The third run was performed to show whether the two subunits interact with each other. In case of possible interaction, an additional peak is observed in the graph, which is expected to be on the left side of both peaks as the complex possesses a higher molecular weight and migrates faster in the gel. As indicated with the arrows, the Nmar_0272 and Nmar_0274 peaks were observed according to their 56 kDa and 18 kDa molecular weights, respectively as expected (***Fig. 3.3.1***). However, no additional peak indicating the Nmar_0272 and Nmar_0274 complex was detected. The result suggests that Nmar_0272 and Nmar_0274 subunits have not interacted with each other.

3.4 The Structure of Nmar_0272 Apo-form

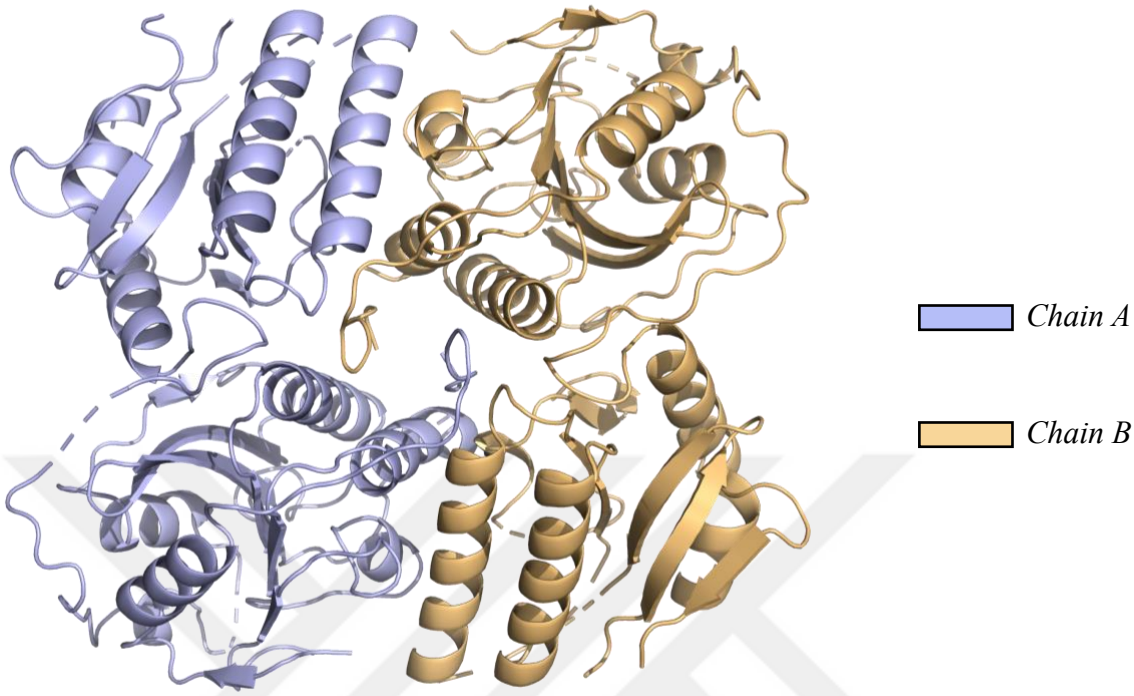


Figure 3.4.1: The cartoon representation of the apo-form Nmar_0272 dimer in space group P63*. The light blue and light orange bars represent chain A and chain B, respectively.*

The biotin-dependent bifunctional acetyl-CoA/propionyl-CoA carboxylase is the key carboxylation enzyme of the *N. maritimus* 3-HP/4-HB cycle. This enzyme catalyzes the carboxylation of acetyl-CoA and propionyl-CoA into malonyl-CoA and methylmalonyl-CoA, respectively. Acetyl-CoA/propionyl-CoA carboxylase possesses three subunits: carboxyltransferase Nmar_0272, biotin carboxylase Nmar_0273 and biotin carboxyl carrier protein Nmar_0274, in which each of these subunits forms multimeric protein complexes to form a functional enzyme (Könneke et al., 2014).

We determined an apo-form dimer structure of Nmar_0272 at 2.26 Å resolution in P63 space group symmetry. Each monomer contains 339 residues; however, due to areas with low electron density, a total of 176 residues are disordered, therefore missing in each chain. Each monomer in our model contains 10 α helices and 12 β sheets where the N-terminal α-helices of one monomer interacts with the C-terminal α-helices of the other monomer (**Fig. 3.4.1**). Due to the abundance of the areas with low electron density in the flexible regions, a portion of the structure is undetermined and contains multiple discrete

loop regions. The disordered regions from N-terminal to C-terminal include the residues: 1-33, 56-73, 93-101, 131-136, 168-172, 184-226, 256-258, 277-279, 367-374, 428-476 for chain A and 1-33, 54-72, 93-99, 129-136, 167-173, 185, 191-230, 276, 277, 367-377, 430-477 for chain B.

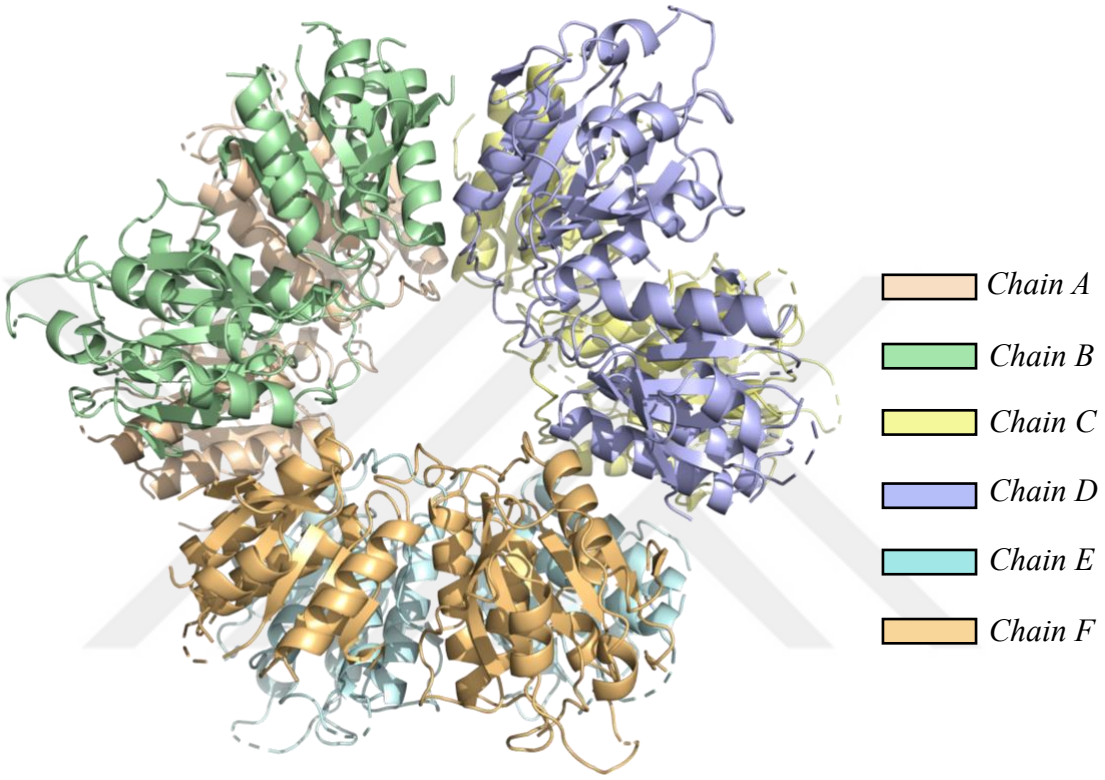


Figure 3.4.2: The cartoon representation of apo-form Nmar_0272 homohexamer structure*. The wheat, pale green, pale yellow, light blue, pale cyan, and light orange colors represent the chains A, B, C, D, E, and F, respectively. The structurally relevant chains C, D, E, and F, are the symmetry mates generated by the symmetry operator.*

Nmar_0272 is a homohexameric protein, arranged as a trimer of dimers and forming a 32-fold symmetry. Therefore the chains in the 3-fold axis interact perpendicular to the chains in the 2-fold axis, as illustrated in **Fig. 3.4.2**. In more detail, the chains B, D, and F interact with the chains A, C, and E, respectively, and form three interacting dimers. In order to show the homohexameric structure, symmetry mates were generated by applying a symmetry operation in PyMOL visualization software as indicated in **Fig 3.4.2**. The chains A and B belong to the original model, and chains C, D, E, and F are generated as symmetry mates.

The interfaces between the dimers play crucial roles for enzyme catalysis as the biotin and the acetyl/propionyl-CoA binding pockets are located between these interfaces (Diacovich et al., 2004; Lin et al., 2006). Since the majority of the residues involved in the catalysis are absent in our model, the missing regions were modeled based on homology. For this purpose, a homology modeling web server Swiss Model was utilized to obtain these missing regions. The model was generated based on the previously determined PDB model 1X0U, and the missing residues were merged with our original model, as demonstrated in **Fig. 3.4.3**. The modeled missing residues are shown with pale yellow. Among these missing regions, residues from 1 to 33 form an N-terminal alpha helix. The residues from 56 to 73 and 93 to 101 are loop regions. The residues from 131 to 136 and 168 to 172 are essential regions playing roles in substrate binding and catalysis in various carboxyltransferase homologs. The region covering the residues 184-226 and 367-374 is a large helix loop helix region and a small helical region, respectively carrying various highly conserved regions. These residues might be involved in interdimeric interactions. A large region covering the residues from 428 to 476 forms a C-terminal helix loop helix. This region also carries certain residues taking active roles in substrate binding and catalysis in many bacterial and archaeal homologs.

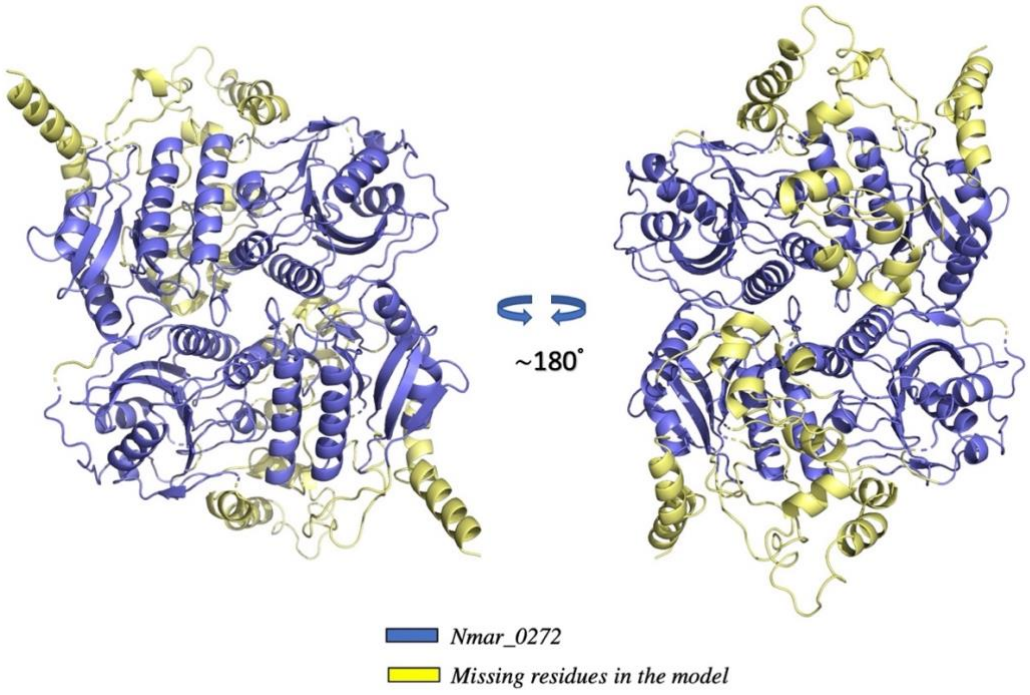
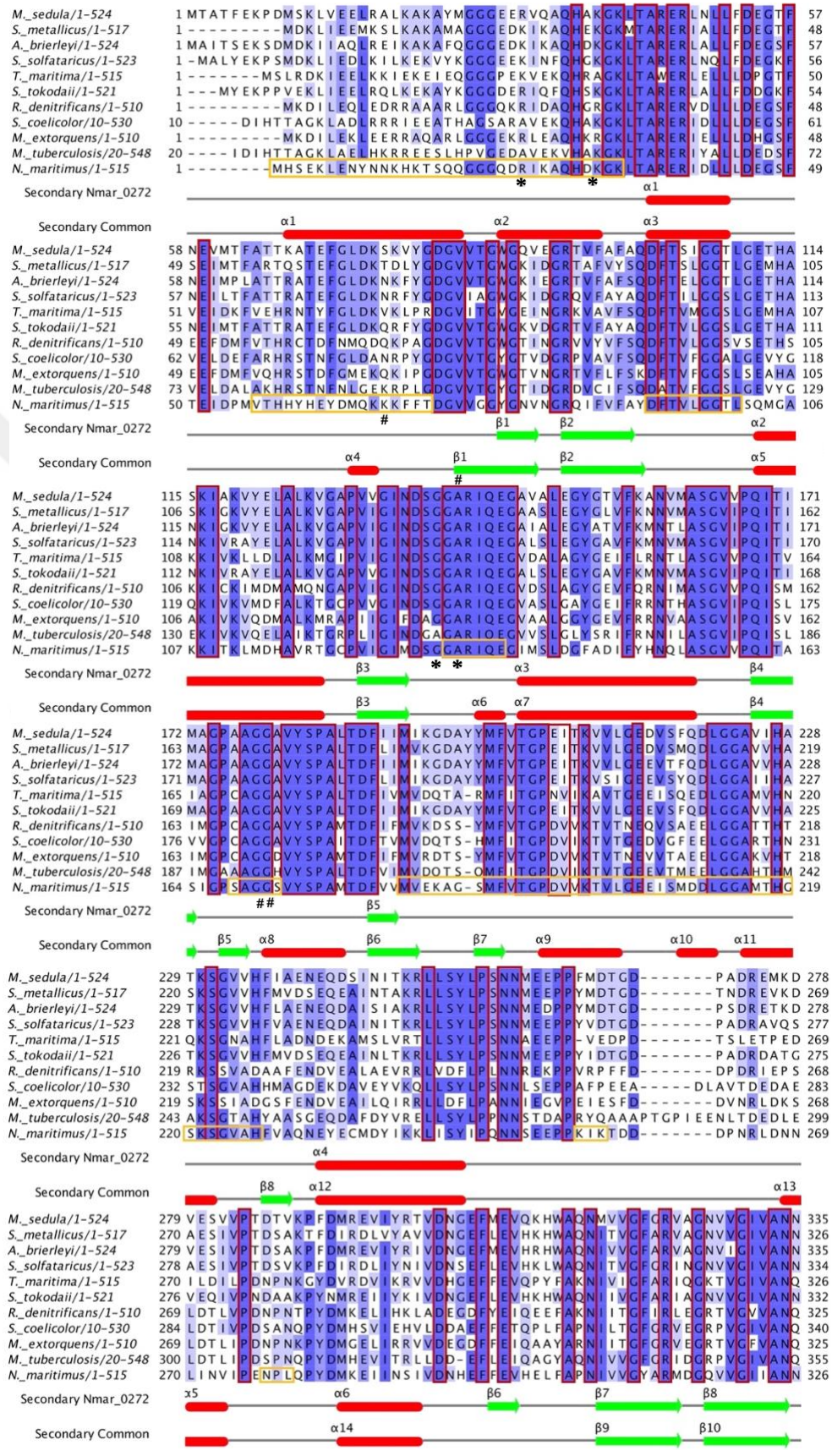


Figure 3.4.3: The cartoon representation of Nmar_0272-Swiss Model hybrid. *The slate blue regions in the model belong to the Nmar_0272 experimental model. The pale yellow regions represent the missing residues generated based on homology and are absent in the original model. The two images demonstrate the front and back view of the same protein model.*

3.5 Multiple Sequence Alignment of Closely Related Homologs



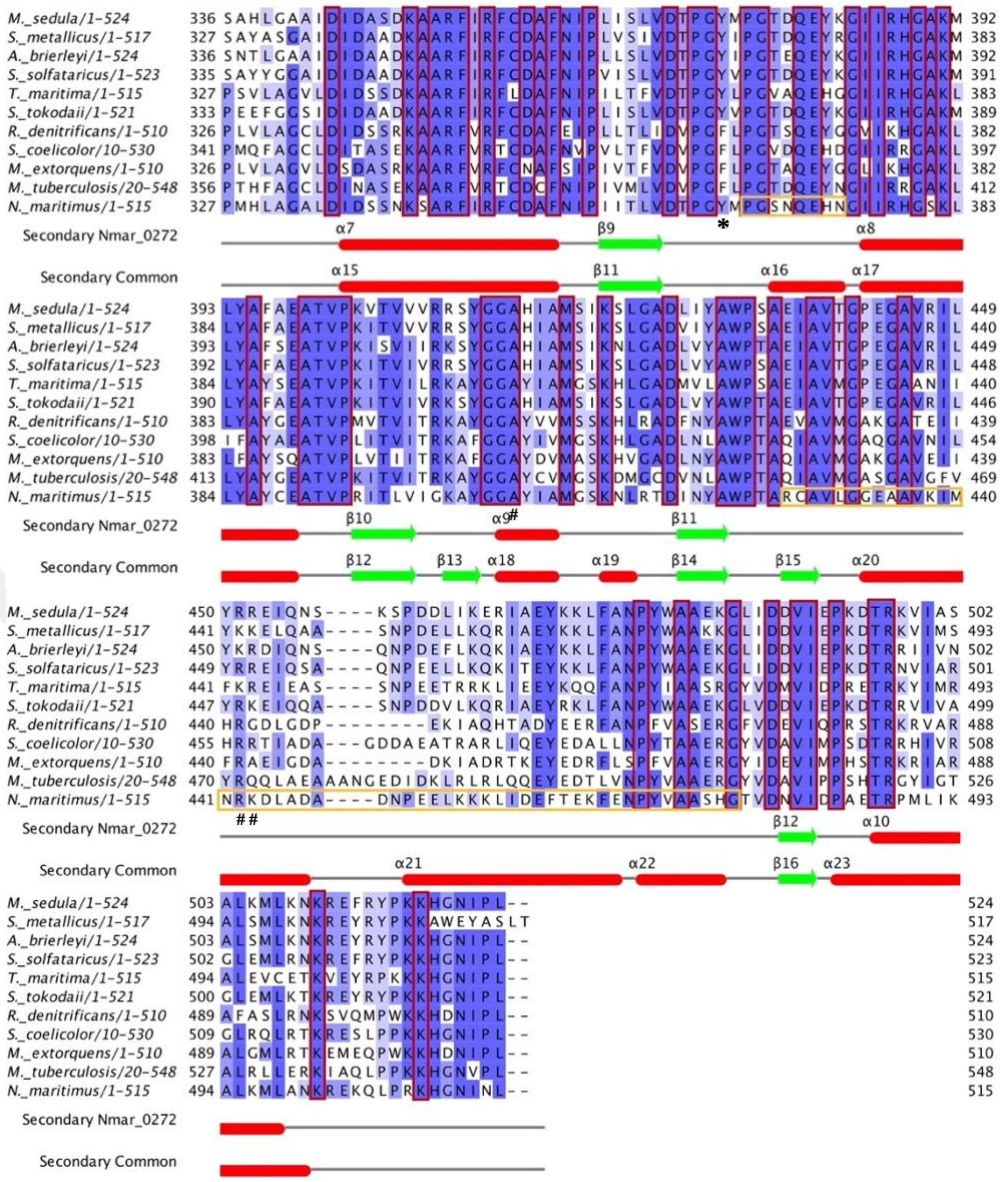


Figure 3.5.1: Multiple sequence alignment of Nmar_0272 and its closely related homologs. *The red frames indicate the highly conserved residues. The yellow frames correspond to the missing residues in the Nmar_0272 model. The red bars and the green arrows next to the Secondary Nmar_0272 title demonstrate the α-helices and the β-sheets that are present in the Nmar_0272 model, respectively. The red bars and the green arrows next to the Secondary Common title indicate the alpha helices and the beta-sheets present in the other homologs.*

The sequence similarity analysis among homologous enzymes is of great importance as they are crucial for protein structure and function. According to a homology search, the highest similarities were obtained from 10 bacterial and archaeal carboxyltransferases. To show the highly conserved and critical residues shared among these homologs, multiple sequence alignment was applied to these closely related homologs of Nmar_0272 by using the ClustalW algorithm of Jalview software (***Fig. 3.5.1***). Among

the aligned sequences, *M. sedula*, *S. metallicus*, *A.brierleyi,* and *S. solfataricus* do not possess determined structures. However, they share high sequence similarities with Nmar_0272 as 53.11%, 52.87%, 56.11%, and 53.51%, respectively. The sequences of determined protein structures are from *T. maritima* (PDB: 1VRG), *S. tokodaii* (PDB: 1X0U), *R. denitrificans* (PDB: 3N6R), *S. coelicolor* (PDB: 1XNY), *M. extorquens* (PDB: 6YBP), and *M. tuberculosis* (PDB: 2A7S) which share 55.47%, 53.89%, 53%, 51.95%, 50.59% and 49.23% sequence similarities respectively.

The sequence of *N. maritimus* carboxyltransferase subunit (Nmar_0272) contains 515 amino acids. The yellow frames indicate the residues absent in our Nmar_0272 model, which is a total of 177 residues. The highly conserved residues that are common in each sequence are indicated with red frames. The residues that are likely to involve polar interactions with the substrates based on the alignment between Nmar_0272, 6YBP, and 3N6R structures are shown with the symbol "*". Similarly, based on the alignment between Nmar_0272 and 1XNY structures, the potential residues involved in polar interactions are indicated with the symbol "#". The active site comparison of Nmar_0272 and 6YBP shows that Arg24 and Gly130 residues of Nmar_0272 are identical with the corresponding Arg23 and Gly129 in the 6YBP sequence. On the other hand, the Lys31 and Tyr365 residues of Nmar_0272 are replaced with the corresponding Arg30 and Phe364 residues in the 6YBP sequence. The active site comparison of Nmar_0272 and 1XNY sequences also demonstrates that the residues Ala132, Gly170, Gly171, Gly405, Ala406, and Arg442 are identical with the corresponding Ala144, Gly182, Gly183, Gly419, Ala420, and Arg456 residues of 1XNY. However, the residues Lys68 and Lys443 of the Nmar_0272 are replaced with Asn80 and Arg457 in the 1XNY sequence.

The α helices and the β sheets of our Nmar_0272 model are indicated with corresponding numbers next to the title "Secondary Nmar_0272". To demonstrate the potential missing secondary structures in our Nmar_0272 model, the possible α helices and the β sheets are also demonstrated based on the other homologs next to the title "Secondary Common". Due to the abundance of the missing residues in our Nmar_0272 model, these residues were homology modeled using the webserver Swiss Model as indicated in **Section 3.4**, **Fig. 3.4.3**.

3.6 Comparison of Nmar_0272 with Previously Identified Structures

3.6.1 Monomer Alignments of Nmar_0272 with its Previously Identified Homologs

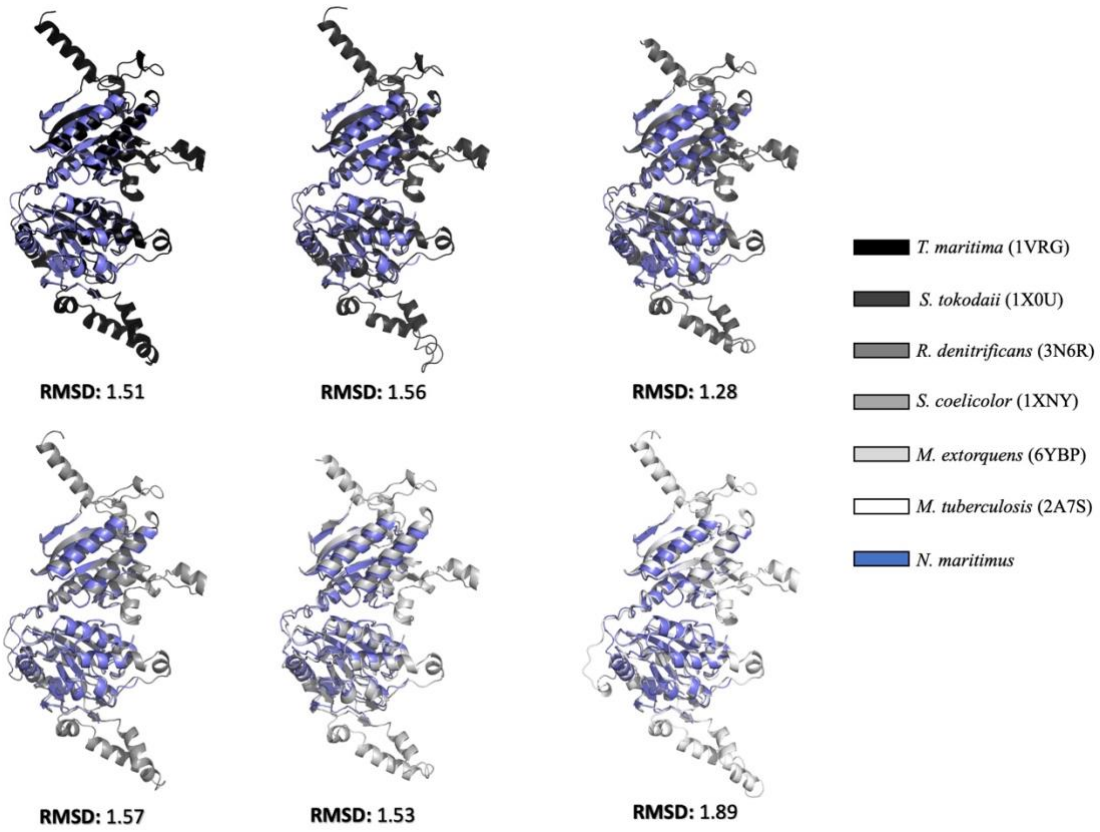


Figure 3.6.1: The structural alignments of the Nmar_0272 monomer with the previously determined PDB structures with the highest sequence and structure similarity. *Nmar_0272 is shown with the color slate. 1X0U, 1VRG, 1XNY, 3N6R, 6YBP, and 2A7S structures are indicated with black, gray20, gray40, gray60, gray80, and white colors, respectively. The sequence independent RMSD values are shown under each alignment.*

The structural similarities between Nmar_0272 and its closely related homologs were analyzed using the structural alignment utility of PyMOL. For this purpose, the PDB structures of 1VRG (black), 1X0U (gray20), 3N6R (gray40), 1XNY (gray60), 6YBP (gray80) and 2A7S (white) were superimposed with Nmar_0272 (slate) experimental model, where they share 55.47%, 53.89%, 53%, 51.95%, 50.59% and 49.23% sequence similarities respectively. The monomers of each of these homolog protein models were aligned with Nmar_0272 chain A (***Fig. 3.6.1***). The obtained RMSD values are 2.21 Å for 1VRG, 2.22 Å for 1X0U, 1.85 Å for 3N6R, 3.45 Å for 1XNY, 2.58 Å for 6YBP, and 2.99 Å for 2A7S. On the other hand, lower RMSD values were achieved from sequence-independent structural superpositions where the obtained RMSD values are 1.51 Å for

1VRG, 1.56 Å for 1X0U, 1.28 Å for 3N6R, 1.57 Å for 1XNY,1.53 Å for 6YBP, and 1.89 Å for 2A7S. The RMSD values of both alignments after the outlier rejection are 0.59 Å, 0.6 Å, 0.7 Å, 0.64 Å, 1.14 Å, and 0.63 Å, respectively.

Although our experimental Nmar_0272 model is a homodimer, the two chains displayed structural differences. The RMSD value of the structural alignment between chains A and B was observed to be 1.83 Å. This value was higher than expected for two identical chains. Nevertheless, the RMDS value decreased to 0.28 Å when the outlier rejection was applied (***Fig. 3.6.2***).

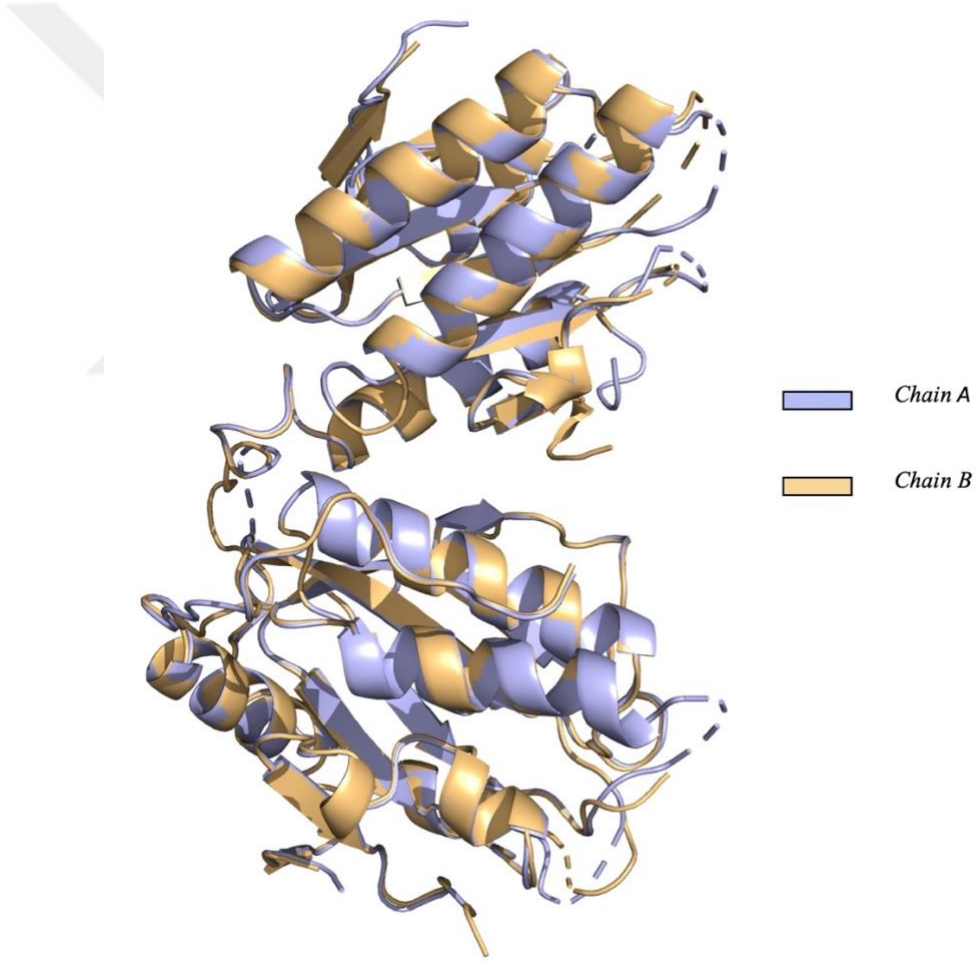


Figure 3.6.2: The structural alignment of Nmar_0272 chain A and chain B. *The chains A and B are colored with light blue and light orange, respectively. The RMSD values were calculated with PyMOL molecular visualization program.*

3.6.2 Active Site Comparison of Nmar_0272 with the Previously Identified Structures

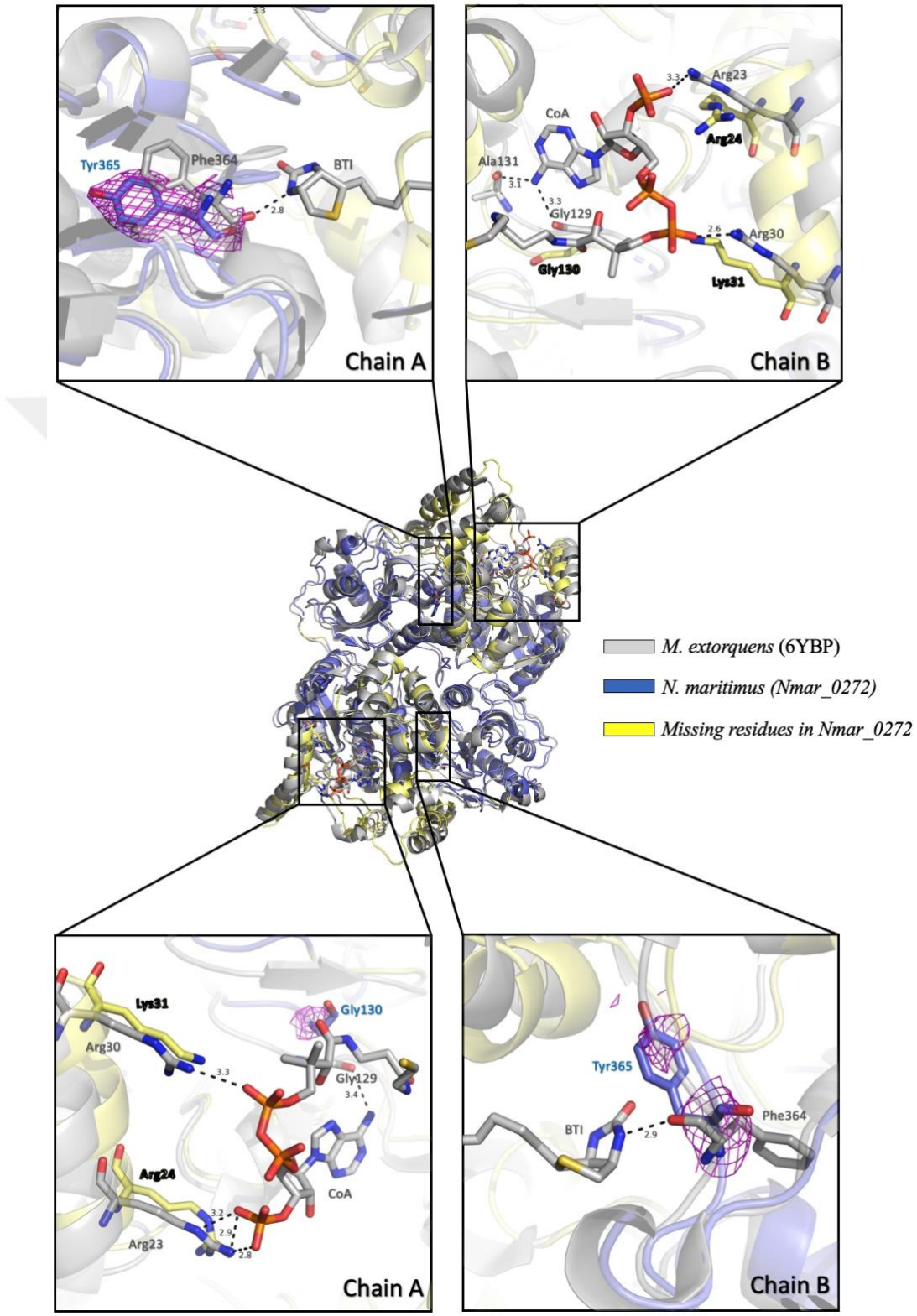


Figure 3.6.3: The comparison between Nmar_0272-Swiss Model hybrid and 6YBP with bound CoA and BTI. *Nmar_0272-Swiss Model hybrid is colored with slate and pale yellow to depict the experimental and missing residues, respectively. 6YBP structure is colored with gray80. The residues that form hydrogen bonds are shown with stick representation. The dashed lines indicate the hydrogen bonds and the distance between atoms in angstroms. Omit $2F_O$ - F_C electron density map of Nmar_0272 is shown with the color magenta.*

To have a broader idea about the catalytic properties of Nmar_0272, the substrate-binding regions of previously determined carboxyltransferase homologs were compared. In this respect, a sequence similarity based homology search was performed, and three substrate-bound structures: 6YBP, 3N6R, and 1XNY, were used for the comparison.

The comparison between Nmar_0272 and 6YBP substrate binding sites clearly demonstrates the residues that take part in BTI and CoA binding. As indicated in ***Fig. 3.6.3***, Phe364 is a critical residue for BTI binding in the active site, in which the N3 atom of BTI forms a hydrogen bond with the backbone O atom of Tyr365. The same interaction was also observed in 3N6R, as shown in ***Fig 3.6.4***. In Nmar_0272, Phe364 is replaced with Tyr365 and exhibits a well-defined electron density in Nmar_0272 chain A. Due to the low electron density in chain B, the same residue displays a different side-chain orientation.

A broader hydrogen bonding network was observed between CoA and the CoA binding site of the 6YBP structure. The Arg23 residue and CoA form a hydrogen bond with their NH2 and O8A atoms, respectively. Another hydrogen bond was also observed between the residues NH2 atom of Arg30 and O4A atom of CoA. Despite in Nmar_0272, the Arg30 residue is replaced with a corresponding Lys31. Moreover, the N6A atom of CoA is involved in forming two additional hydrogen bonds, where it interacts with O atoms of both Ala131 and Gly129 residues. These two residues are also highly conserved in many carboxyltransferase homologs. Due to the low electron density, only the Gly130 residue was present in our experimental model among the residues interacting with CoA. The electron density of Gly130 is demonstrated in ***Fig 3.6.3***.

The superimposition between Nmar_0272 and 1XNY revealed different binding pockets, as indicated in ***Fig 3.6.5***, where the interaction takes place at the dimeric interface. These binding pockets are present at the interface between the N-terminal of one chain and the C-terminal of the other chain. Unlike 3N6R and 6YBP structures, residues in both chains are involved in the formation of the binding pockets. In both of the binding pockets in the dimeric interface, the O3 atom of the biotin (BTN) and the N atom of Ala420 form a hydrogen bond.

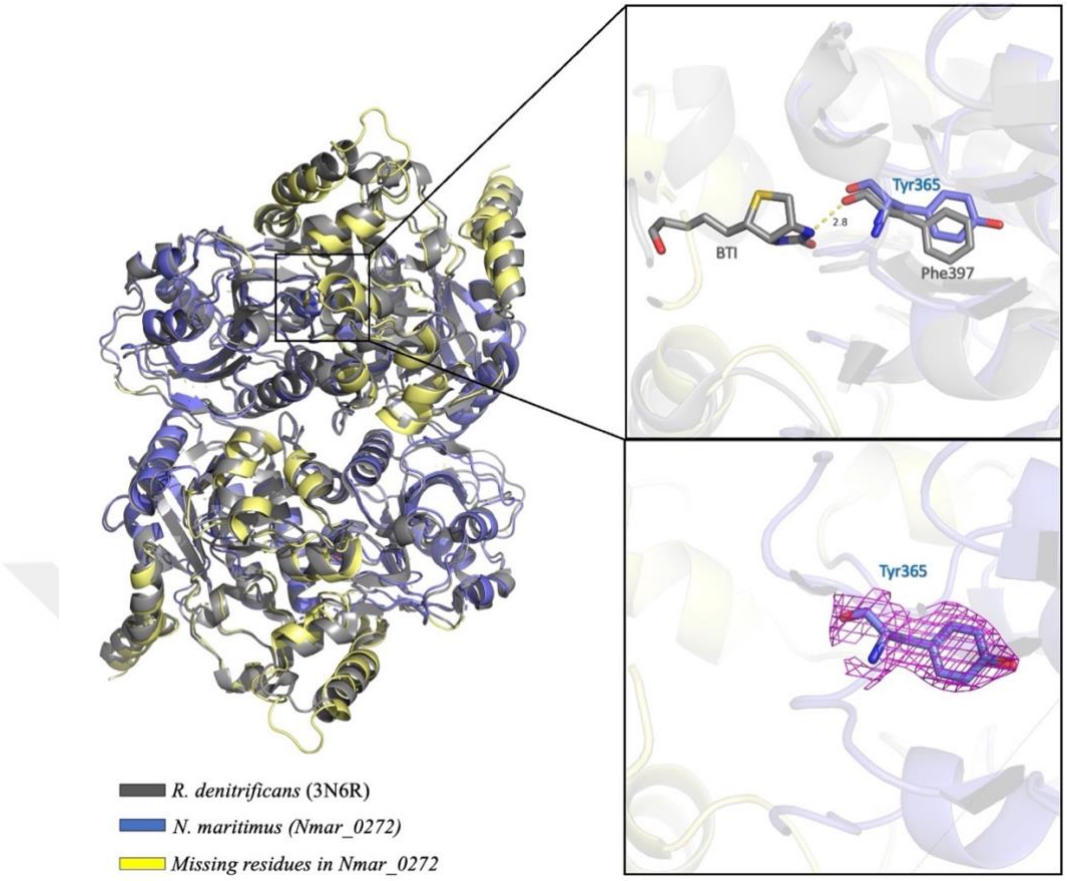


Figure 3.6.4: The comparison of substrate binding sites of Nmar_0272-Swiss Model hybrid and 3N6R with bound BTI. *Only the residues forming hydrogen bonds are shown with sticks. Nmar_0272- Swiss Model hybrid and 3N6R structures are shown with slate (experimental model)/pale yellow (missing residues) and gray40, respectively. $2F_O$ - F_C electron density map of Tyr365 is demonstrated with the color magenta. The dashed yellow line indicates the hydrogen bonds and the distance between the atoms in angstroms.*

In the propionyl-CoA binding site of the 1XNY structure, a hydrogen bond is present between the NH group (ND2) of Asn80 and the O14 atom of propionyl-CoA. In Nmar_0272, this residue is replaced with a Lys68 residue. Another polar contact is present between the O atom of Ala144 and the NH group (N4) of propionyl-CoA. The study conducted by Diacovich et al. suggests that Gly182 and Gly183 are involved in forming an oxyanion hole, and the NH groups form a hydrogen bond with the carbonyl group of propionyl-CoA. The same carbonyl group forms another hydrogen bond with the N1 atom of BTN. The second oxyanion hole is formed by the residues Gly419 and Ala420, where the NH group forms a hydrogen bond with the carbonyl group of the BTN. These residues also exhibited well electron densities in Nmar_0272. Propionyl-CoA is further stabilized with the hydrogen bonds formed between the guanidine groups (NH1) of Arg456 and Arg457 residues and O4, O5 and O7 and O9 atoms of propionyl-CoA (Diacovich et al., 2004). In Nmar_0272, Arg457 is replaced with a Lys443 residue (***Fig 3.6.5***).

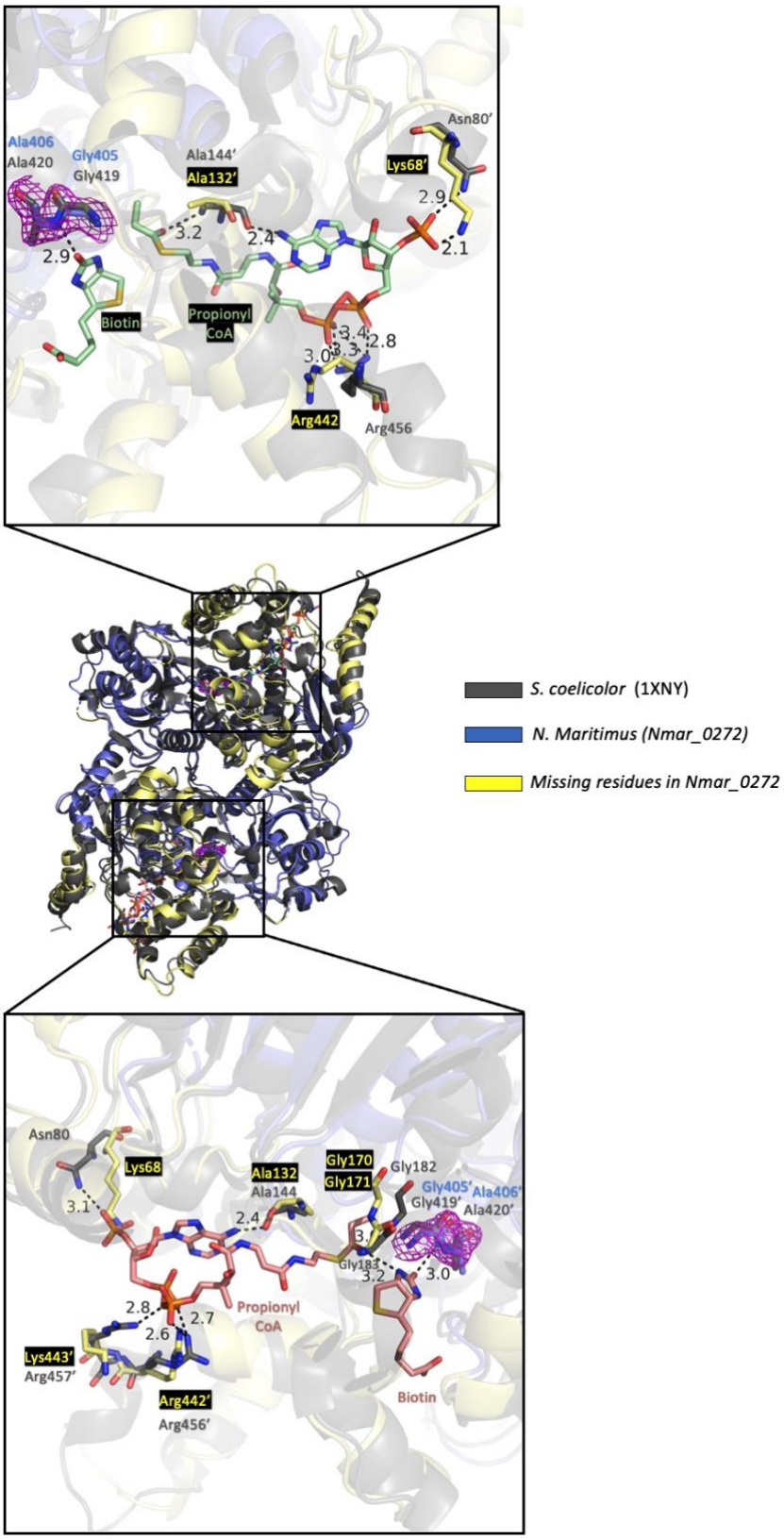


Figure 3.6.5: Comparison between the propionyl-CoA and biotin-binding sites of Nmar_0272-Swiss Model hybrid and 1XNY structures. *Only the residues forming polar contacts are shown. Nmar_0272 and its missing residues are shown with the colors slate and pale yellow, respectively. 1XNY structure and its substrates are shown with gray60, pale green (top), and salmon (bottom), respectively. Omit $2F_O$ - F_C electron densities are demonstrated with the color magenta. The residues names without and with apostrophes stand for chain A and chain B, respectively. The dashed lines indicate the hydrogen bonds and the distance between atoms in angstroms.*

3.7 Prediction of Nmar_0272 Structure with Alpha Fold and RoseTTA Fold

3.7.1 Alpha Fold Prediction of Nmar_0272

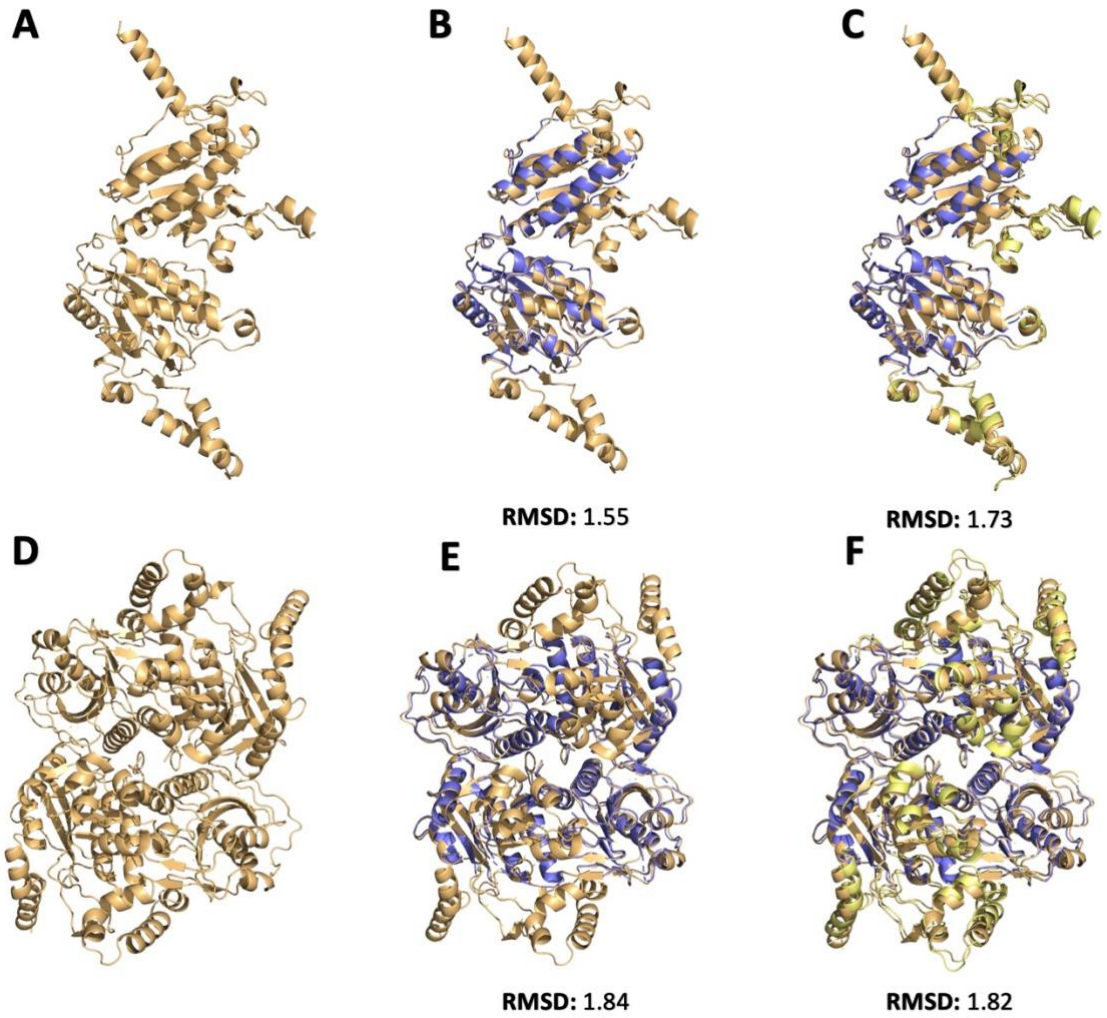


Figure 3.7.1: Alpha Fold predictions of Nmar_0272 (A-D) *Monomer and dimer structures. (**B-E**) The monomer and dimer alignment of Nmar_0272 experimental and the Alpha Fold models. (**C-F**). The monomer and dimer alignment of Nmar_0272 experimental-Swiss Model hybrid and Alpha Fold models. The Alpha Fold and the experimental models are colored with light orange and slate, respectively. The missing residues in the Nmar_0272 experimental-Swiss Model hybrid are shown with pale yellow. The RMSD values are depicted under each structure.*

Recently, Google's DeepMind company has developed an algorithm named Alpha Fold which predicts protein structures using artificial intelligence. This program promises to predict protein structures with high consistencies, making it a valuable tool for the structural biology field (Jumper et al., 2021). Since the program claims to predict high accuracy structures, Alpha Fold was used to predict the structure of Nmar_0272 (***Fig. 3.7.1***). The model generated by Alpha Fold predicted a structure with high accuracy with a predicted local Distance Difference Test (plDDT) point of more than 90%. This test is used to show how much the experimental model fits with the computational model based

on local distance differences of the atoms instead of structural superposition (Mariani et al., 2013). The RMSD values of the monomer and dimer alignment between Nmar_0272 experimental and Alpha Fold structures were calculated as 1.55 Å and 1.84 Å, respectively (***Fig 3.7.1 B-E***). However, following the outlier rejection, these values dropped to 0.47 Å and 0.54 Å, respectively. The monomer and dimer superposition between the Nmar_0272-Swiss Model hybrid gave RMSD values of 1.73 Å and 1.82 Å, respectively (***Fig 3.7.1 C-F)***. The outlier rejection resulted in the RMSD values of 0.75 Å and 0.78 Å, respectively. By comparing the models, one may interpret that the Alpha Fold predicted Nmar_0272 structure has high consistency with the experimental and hybrid models.

3.7.2 Monomer Alignments of Nmar_0272 Alpha Fold Model and its Homologs

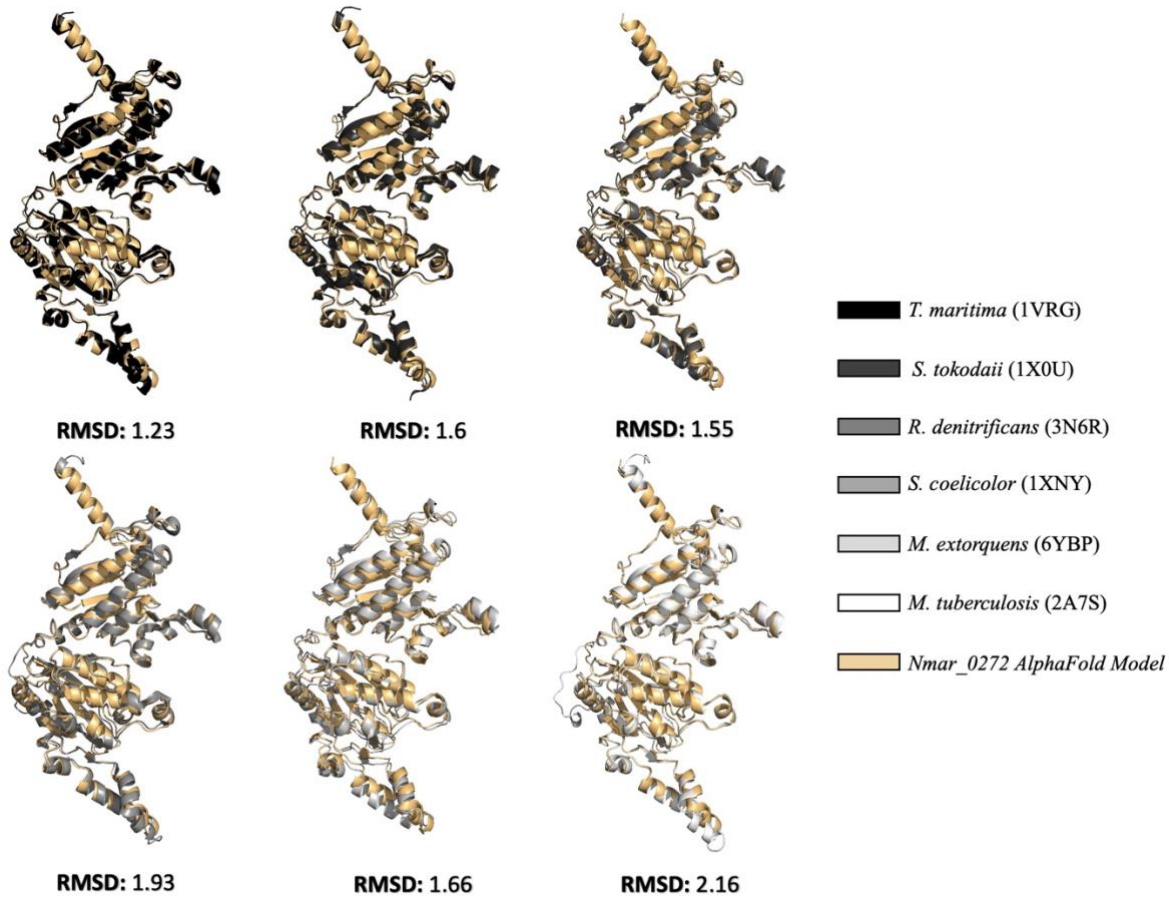


Figure 3.7.2: The monomer alignment of Nmar_0272 Alpha Fold model with the close related previously determined PDB structures. *The Alpha Fold is illustrated with the color light orange. The structures of 1VRG, 1X0U, 3N6R, 1XNY, 6YBP, and 2A7S are demonstrated with black, gray20, gray40, gray60, gray80, and white, respectively. All the superpositions and RMSD calculations were performed by PyMOL. The RMSD values are indicated under each figure. The order of the structures is arranged from the highest to the lowest sequence similarities.*

The superpositions between the Nmar_0272 Alpha Fold model were compared with closely related homologs to analyze their structural similarities. These homologs were chosen based on sequence and structural similarities. In this respect, 1VRG (black), 1X0U (gray20), 3N6R (gray40), 1XNY (gray60), 6YBP (gray80) and 2A7S (white) structures were used (***Fig. 3.7.2***). Following the superimpositions, the RMSD values of each superposition were calculated. Based on the alignments between the Nmar_0272 Alpha Fold model with the homologs, the following RMSD values were obtained: 1.23 Å for 1VRG, 1.6 Å for 1X0U, 1.55 Å for 3N6R, 1.93 Å for 1XNY, 1.66 Å for 6YBP, and 2.16 Å for 2A7S. After the application of outlier rejection, the values dropped to 0.66 Å for 1VRG, 0.72 Å for 1X0U, 0.77 Å for 3N6R, 0.71 Å for 1XNY, 1.24 Å for 6YBP and 0.73 Å for 2A7S structures. Compared with the superpositions between the Nmar_0272 experimental model and the identical homologs mentioned in **Chapter 3.6.1**, similar RMSD values were obtained. Therefore, one may interpret that the Alpha Fold model might reflect a consistent prediction compared with the Nmar_0272 experimental model.

3.7.3 RoseTTA Fold Prediction of Nmar_0272 and Monomer Alignments

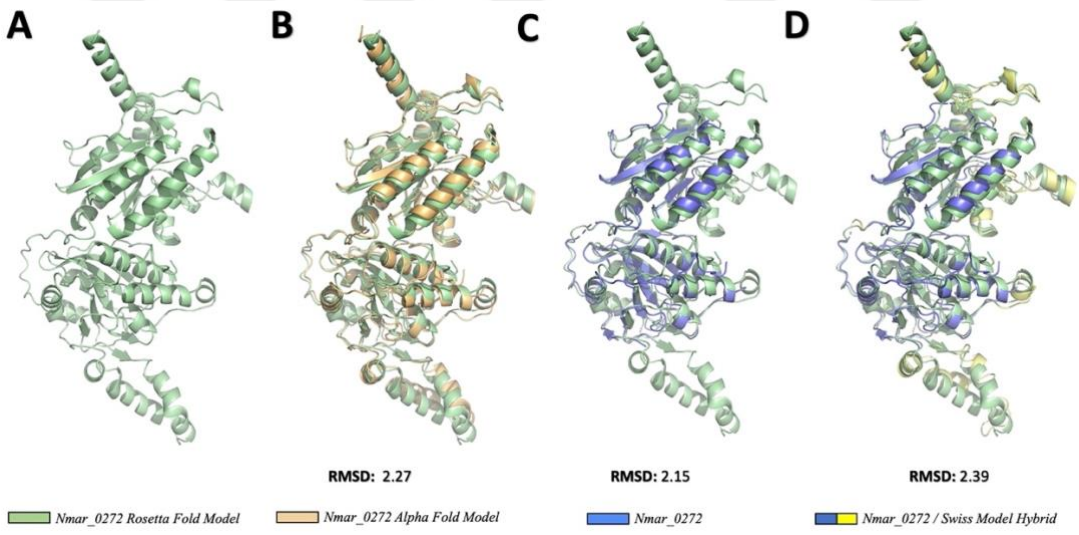


Figure 3.7.3: Rosetta Fold prediction of Nmar_0272 monomer structure (A) and its alignments with Alpha Fold prediction (B) Nmar_0272 experimental model (C), Nmar_0272 / Swiss Model hybrid (D). *The Rosetta Fold and Alpha Fold models are colored with pale green and light orange, respectively. Nmar_0272 experimental model and the Swiss Model prediction of its missing residues are depicted with slate and pale yellow, respectively. The RMSD values are shown under each superposition. All the alignments and RMSD calculations were executed with PyMOL.*

RoseTTA Fold is an alternative method for protein structure prediction. Similar to Alpha Fold, this method promises to solve the modeling issues encountered in X-ray crystallography and Cryo-EM (Baek et al., 2021). To analyze our experimental, homology-based hybrid and Alpha Fold models, the Nmar_0272 structure was also

predicted using RoseTTA Fold. Based on the prediction, a monomer model was obtained with a 0.86 confidence score based on IDDT (***Fig 3.7.3 A***). The alignment of the RoseTTA Fold and Alpha Fold model predictions exhibited an RMSD value of 2.27 Å (***Fig 3.7.3 B***). Moreover, the alignment of the RoseTTA Fold model between the experimental model and the hybrid model also resulted in the RMSD values of 2.15 Å and 2.39 Å, respectively (***Fig 3.7.3 C-D***). The outlier rejections resulted in a noticeable decrease in the RMDS values where the mentioned values dropped to 1.21 Å, 1.07 Å, and 1.35 Å, respectively. Therefore the results suggest that a relatively lower consistency is obtained from the RoseTTA Fold prediction compared to the Alpha Fold monomer prediction.

3.7.4 Monomer Alignments of Nmar_0272 Alpha Fold Model and its Homologs

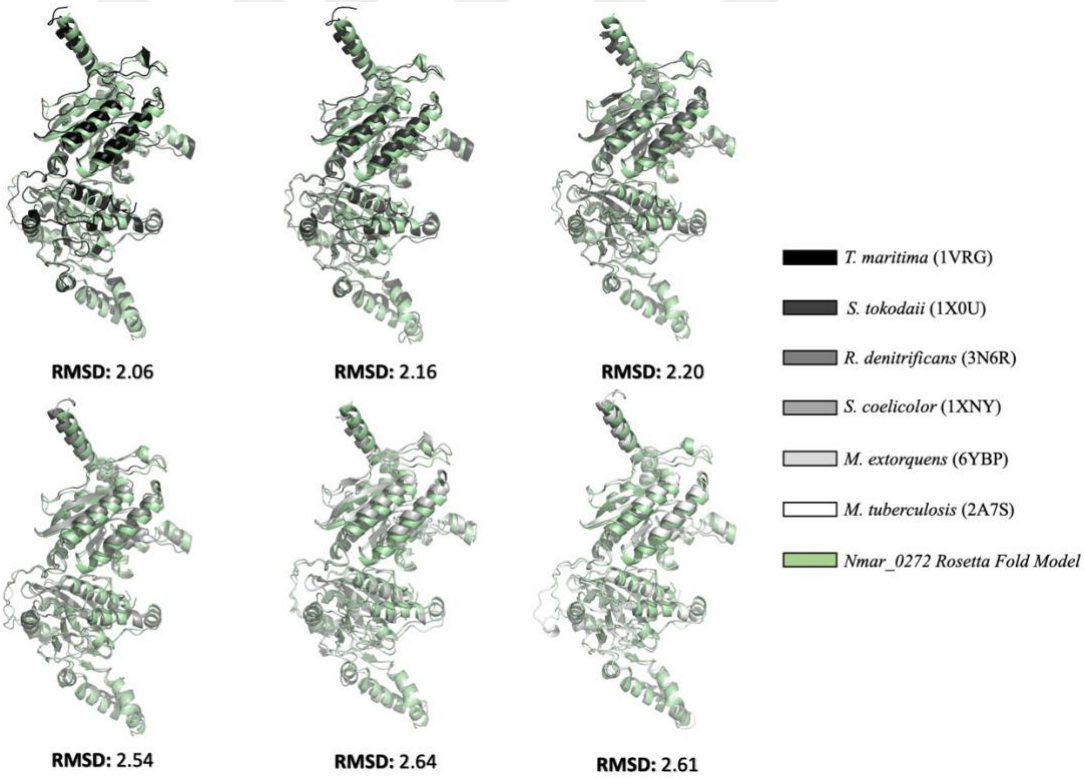


Figure 3.7.4: The alignment of Nmar_0272 Rosetta Fold model with the closely related carboxyltransferase homologs. *The Rosetta Fold model is colored pale green. The homologs 1VRG, 1X07, 3N6R, 1XNY, 6YBP, 2A7S, are depicted with black, gray20, gray40, gray60, gray80, and white, respectively. The RMDS values are shown under each superposition. The alignments and RMSD calculations were performed by using PyMOL.*

The RoseTTA Fold model was also compared with the previously determined carboxyltransferase homologs with high sequence similarities. The 1VRG, 1X0U, 3N6R, 1XNY, 6YBP, 2A7S structures were used for superposition and RMSD calculations. Contrary to the results obtained from Alpha Fold prediction, the RoseTTA Fold

prediction demonstrated slightly lower structural similarities. The obtained RMSD values are 2.06 Å for 1XNY, 2.16 Å for 1X0U, 2.20 Å for 3N6R, 2.54 Å for 1XNY, 2.64 Å for 6YBP, and 2.61 Å for 2A7S. (***Fig. 3.7.4***). When the outliers were denied, these values decreased further to 1.29 Å, 1.17 Å, 1.33 Å, 1.41 Å, 1.86 Å, and 1.44 Å, respectively. Therefore, the structure alignment results suggest that, compared with the RoseTTA Fold, the Alpha Fold prediction of Nmar_0272 is more consistent with our experimental data, homology model, and the previously determined carboxyltransferase homologs.

3.7.5 Comparison of Biotin and Propionyl-CoA Binding Pockets with Predicted Models

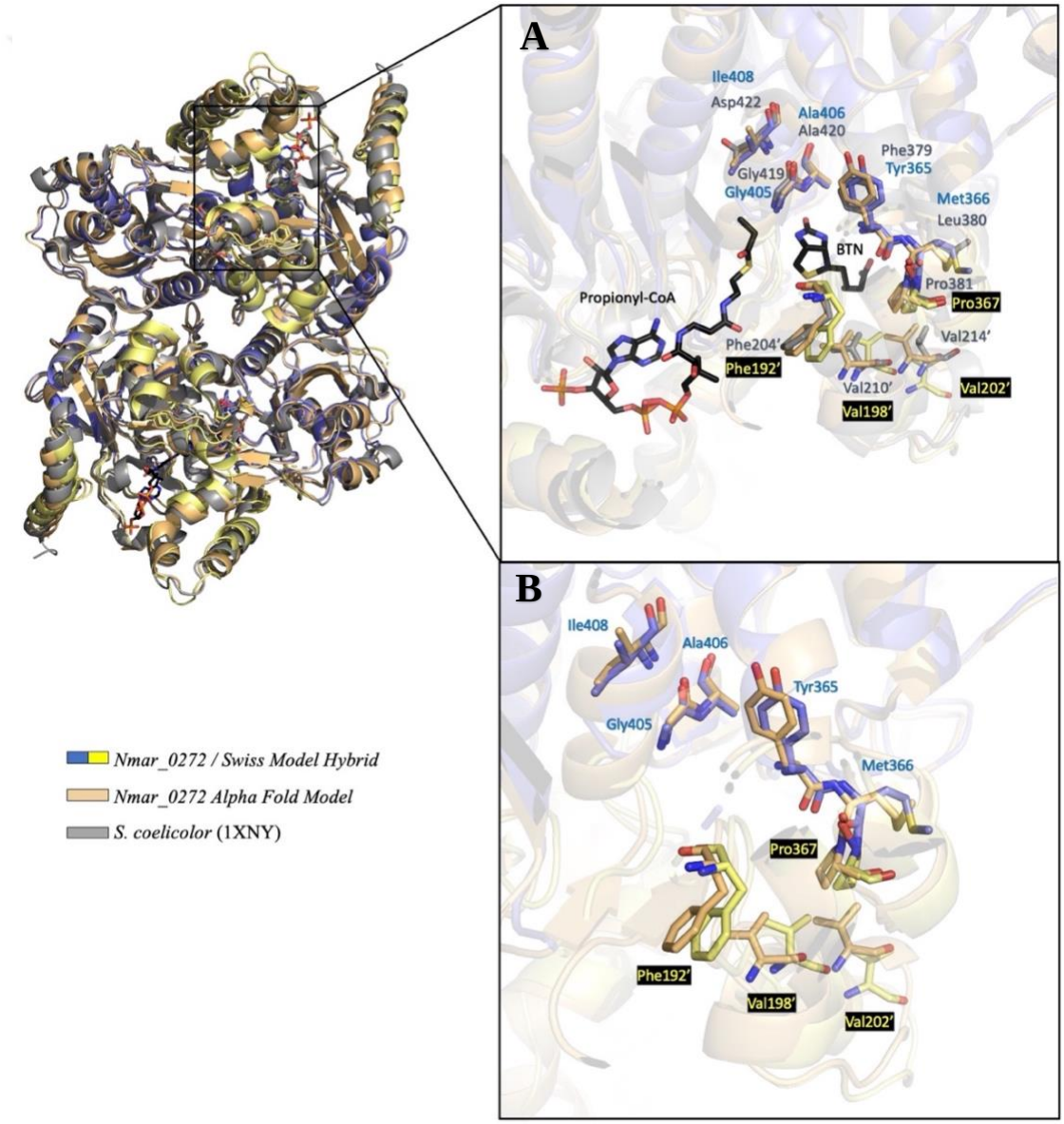


Figure 3.7.5: A) The comparison of the biotin (BTN) binding pocket of Nmar_0272 / Swiss Model hybrid, Alpha Fold, and 1XNY models. *The residues belonging to the experimental model and missing residues of the*

Nmar_0272/Swiss Model hybrid are shown with slate and pale yellow, respectively. The residues of Alpha Fold and 1XNY models are depicted with the colors light orange and gray60, respectively. All the residues are labeled with their corresponding colors (except the Alpha Fold model). The residue names without and with apostrophes correspond to the chains A and B, respectively. Propionyl-CoA and biotin were indicated with black sticks. ***B)*** *The comparison of Nmar_0272/Swiss Model hybrid and Alpha Fold models only.*

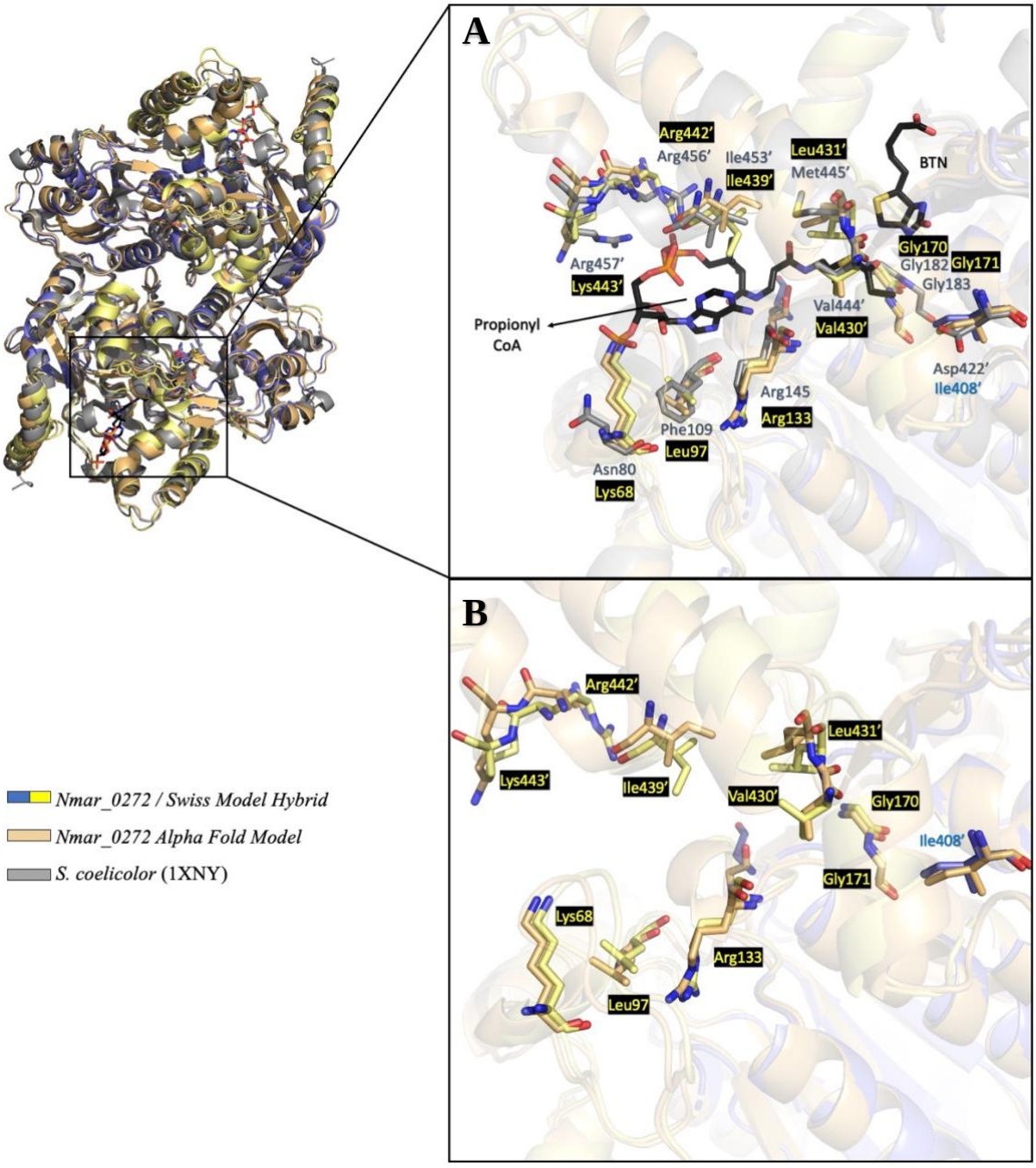


Figure 3.7.6: A) Comparison of the propionyl-CoA binding pockets of Nmar_0272/ Swiss Model hybrid, Alpha Fold, and 1XNY structures. *In the Nmar_0272/Swiss Model hybrid model, the experimental and missing residues are depicted with slate and pale yellow, respectively. Nmar_0272 Alpha Fold model and 1XNY structure are colored with light orange and gray6o, respectively. The residues of the Nmar_0272 experimental model, the missing residues, and 1XNY structures are labeled with their corresponding colors. Propionyl-CoA and biotin were shown with black sticks. The residue labels without and with apostrophes are used to indicate the chains A and B, respectively.* ***B)*** *The comparison between Alpha Fold prediction and Nmar_0272/ Swiss Model hybrid only.*

Since a relatively more accurate model was obtained from the Alpha Fold prediction, this model was used as a reference to analyze the biotin and propionyl-CoA binding pockets of our model. In this respect, the critical binding pocket residues of Nmar_0272/ Swiss Model hybrid, Nmar_0272 Alpha Fold model, and 1XNY structures were superimposed. The residues Phe204, Val210, Val214, Phe379, Leu380, Pro381, Gly419, Ala420, and Asp422 are critical for biotin binding in the 1XNY structure. These hydrophobic and aromatic residues form a highly hydrophobic biotin binding pocket (Diacovich et al., 2004). Interestingly, this pocket also reveals several non-conserved residues when compared to Nmar_0272. For instance, the residues Asp422, Phe379, Leu380 in 1XNY are replaced with Ile408, Tyr365, and Met366 in Nmar_0272, respectively (***Fig. 3.7.5 A***).

In the biotin-binding pocket of Nmar_0272/ Swiss Model hybrid, the residues Tyr365, Met366, Gly405, Ala406, and Ile408 belong to the experimental model. The comparison of the Nmar_0272 Alpha Fold prediction, experimental model, and 1XNY structure exhibited high consistency in residue conformations and positions. Moreover, both the side chains and backbones aligned well with each other. On the other hand, the residues Phe192, Val198, Val20,2, and Pro367, which belong to the Swiss Model, displayed slight differences. Compared with the experimental data, a relatively lower fit was observed in the side chain orientations and backbone positions between the Alpha Fold and Swiss Model predictions. Nevertheless, the accuracy between the 1XNY and Alpha Fold side-chain positions and orientations was relatively higher (***Fig. 3.7.5 B***).

The propionyl-CoA binding pocket exhibits a larger binding pocket in the 1XNY structure. The main residues forming this binding pocket are Asn80, Phe109, Arg145, Gly182, Gly183, Asp422, Val444, Met445, Ile453, Arg456, and Arg457. Interestingly, many critical residues in the 1XNY structure are replaced with alternative amino acids in Nmar_0272. For instance, the residues Asn80, Phe109, Asp422, Me,t445, and Arg457 in the 1XNY structure are replaced with Lys68, Leu97, Ile408, L,eu431, and Lys443, respectively (***Fig. 3.7.6 A***).

Among the residues in this binding pocket, only Ile 422 displayed a well-defined electron density. Furthermore, the side-chain orientation and backbone position of Ile422 are significantly consistent with the corresponding residues in both the 1XNY and

Nmar_0272 Alpha Fold models. All the other residues were predicted based on homology, using the webserver Swiss Model. The comparison of both the Swiss Model and Alpha Fold predictions shows nearly identical side chain and backbone orientations for the residues Lys68, Leu97, Arg133, Gly170, and Gly171. The residues Val430, Leu431, and Ile439 also displayed decent, however, relatively fewer consistencies. The only inconsistencies were observed for the residues Arg442 and Lys443, where the side chain and backbone orientations differed considerably (***Fig 3.7.6 B***). The surface representation of the 1XNY, Nmar_0272/Swiss Model hybrid, and Nmar_0272 Alpha Fold models also displayed nearly identical binding pockets. On the other hand, slight differences were observed in all the models, especially in the acyl-CoA (acetyl-CoA/propionyl-CoA) binding pocket (***Fig 3.7.7 A-B-C***).

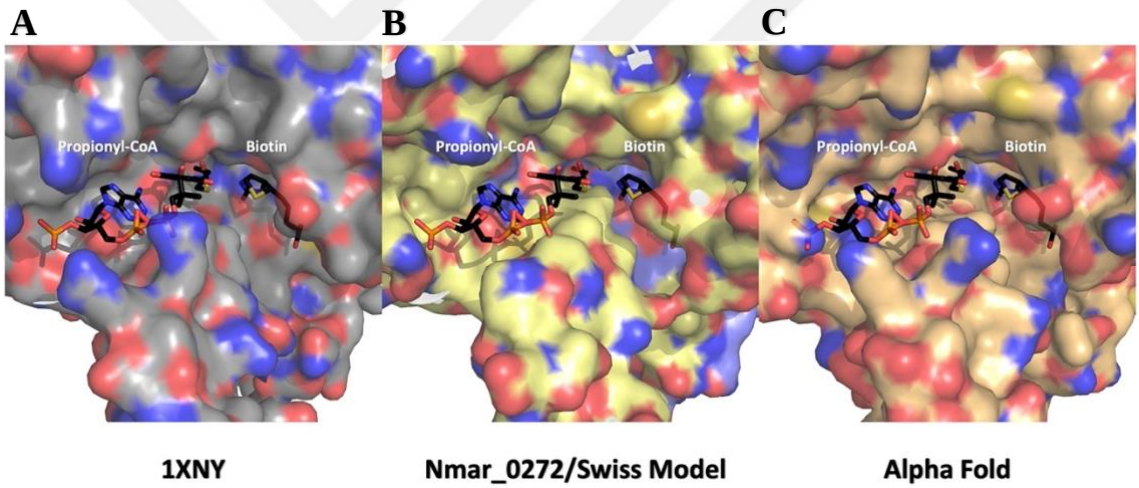


Figure 3.7.7: The surface representation of acyl-CoA and biotin-binding pockets of A) 1XNY, B) Nmar_0272/ Swiss Model hybrid, and C) Nmar_0272 Alpha Fold models*. The propionyl-CoA and biotin are shown with black sticks and belong to the 1XNY structure. The identical propionyl-CoA and biotin are also illustrated in **B** and **C** to simplify the comparison.*

Chapter 4:
DISCUSSION AND CONCLUSION

Autotrophic carbon fixation mechanisms are widely distributed among prokaryotes and eukaryotes to maintain the balance of inorganic and organic carbon sources. 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle is one of the carbon fixation mechanisms utilized by the archaeal phyla Crenarchaeota and Thaumarchaeota. In this cycle, two bicarbonate molecules are fixed into organic carbon precursors. The first half-reactions of the cycle convert the acetyl-CoA into succinyl-CoA by using a 3-hydroxypropionate intermediate. The other half set of the reactions convert succinyl-CoA into two acetyl-CoA molecules using the intermediate 4-hydroxybutyrate. The precursors produced in these reactions are utilized to synthesize more complex molecules and the building blocks of the cells with a process called biosynthesis (Berg et al., 2007).

Both the Crenarchaeal and Thaumarchaeal HP/HB cycles use oxygen tolerant enzymes and produce identical precursor molecules. Despite that, recent studies discovered that the species *Nitrosopumilus maritimus* from the phylum Thaumarchaea uses an altered version of the HP/HB cycle, making this pathway the most energy-efficient aerobic carbon fixation cycle that has ever been identified. Compared with the Crenarchaeal HP/HB cycle, bifunctional and more energy-efficient enzymes were identified in *N. maritimus*. For instance, rather than using separate hydratase and dehydratase enzymes, N. maritimus uses a promiscuous enzyme 3-hydroxypropionyl-CoA dehydratase/crotonyl-CoA hydratase (Nmar_1308). Additionally, instead of using AMP-producing ligases, the ADP-forming alternatives (Nmar_0206 and Nmar_1309) are present in *N. maritimus*. These features grant *N. maritimus* to consume nearly half of the ATP spent in the Crenarchaeal HP/HB cycle. (Könneke et al., 2014)

In both the Crenarchaeal and Thaumarchaeal HP/HB cycles, the carboxylation reactions are catalyzed by a single biotin-dependent bifunctional acetyl-CoA/propionyl-CoA carboxylase. *N. maritimus* carboxylase contains three subunits: carboxyltransferase (CT) Nmar_0272, biotin carboxylase (BC) Nmar_0273, and biotin carboxyl carrier protein (BCCP) Nmar_0274. These subunits are involved in a two-step reaction. In the first step of the reaction, a biotin molecule attached to Nmar_0274 (BCCP) is carboxylated by the

Nmar_0273 (BC) subunit. This reaction follows the translocation of the resulting carboxybiotin intermediate to the Nmar_0272 (CT) subunit with the help of Nmar_0274 (BCCP). The second reaction takes place by transferring the carboxyl group of carboxybiotin to acetyl-CoA or propionyl-CoA substrates by Nmar_0272 (CT). As a result of these carboxylation reactions, two bicarbonate molecules are fixed to convert the acetyl-CoA and propionyl-CoA into malonyl-CoA and methylmalonyl-CoA, respectively (Tong, 2013; Könneke et al., 2014).

In this study, we identified the structure of apo-form Nmar_0272 of *N. maritimus* acetyl-CoA/propionyl-CoA carboxylase at 2.26 Å resolution and in the P63 space group. The structure displayed a dimer structure where the N-terminal α helices of a chain directly interact with the C-terminal alpha-helices of the other chain, forming a di-domain dimeric interface. The application of the symmetry operator generates a trimer of dimers homo-hexamer structure in which three chains interact perpendicular with their dimeric partners, as demonstrated in ***Fig 3.4.2***. Importantly, the structure of the Nmar_0272 experimental model displayed high structural and sequence-related similarities with its previously identified carboxyltransferase homologs 1VRG, 1X0U, 3N6R, 1XNY, 6YBP, and 2A7S, where all these structures share about 50% sequence similarities and displayed low RMSD values when aligned.

The dimer structure of Nmar_0272 exhibited some regions with low electron density. Therefore 176 residues in these low electron density regions could not be modeled for each chain. Unfortunately, some of these residues were critical for the formation of acetyl-CoA/propionyl-CoA and biotin-binding pockets. Therefore, only the missing residues were constructed based on homology by the webserver Swiss Model, using the 1X0U structure as a reference. Similar to the Nmar_0272 experimental model, the obtained Swiss Model prediction also showed high structural similarities with the previously determined homologs and the exhibited low RMSD values.

The comparison between Nmar_0272/Swiss Model hybrid and substrate-bound homolog structures displayed possible binding pocket residues. The superposition between our hybrid model and the cryo-EM structures 6YBP and 3N6R demonstrated that the backbone of the Phe364 and Phe397 residues are critical for BTI binding, respectively.

In Nmar_0272, this residue is replaced with the backbone of a Tyr365 residue. Multiple sequence alignment results also indicate that the same tyrosine replacements are observed in many homologs. Since only the backbones of tyrosine and phenylalanine participate in these polar contacts, it can be interpreted that the side chains might only have structural roles rather than their roles in catalysis. For the CoA binding, the residues Arg23, Arg30, Gly129, and Ala131 were involved in polar contacts in the 6YBP structure. However, in Nmar_0272, Arg30 is replaced with an alternative positively charged Lys31 residue. Despite that, all the other residues of 6YBP involved in CoA binding were identical with the corresponding residues of Nmar_0272. Among these residues, Gly129 and especially Ala131 are highly conserved in almost all the homologs, including Nmar_0272, suggesting that these residues are very likely to take part in protein function and catalysis.

Compared to the 3N6R and 6YBP structures, the co-crystal structure of 1XNY with propionyl-CoA and biotin displays different and more realistic binding pockets. This difference was not unexpected as propionyl-CoA is a much larger molecule. Therefore it cannot bind to the CoA binding pocket as observed in 6YBP due to steric hindrance. Similar to the other homologs, these binding pockets are located at the dimeric interfaces. The 1XNY structure reveals that Asn80, Ala144, Gly182, Gly183 Arg456, and Arg457 are critical residues mediating the binding of propionyl-CoA to the binding pocket. In Nmar_0272, the uncharged Asn80 residue is substituted with a positively charged Lys68. Interestingly this residue is not conserved and highly variable in many different homologs. The Arg457 of 1XNY is also replaced with a similarly charged residue Lys443 in Nmar_0272. Importantly, in the 1XNY structure, Diacovich et al. showed that the residues Gly182 and Gly183 form an oxyanion hole and stabilizes the propionyl-CoA during catalysis. The same oxyanion hole is formed by the residues Gly170 and Gly171 in Nmar_0272. More importantly, this glycine doublet is very well conserved in most carboxyltransferase homologs, therefore playing vital roles during catalysis. On the other hand, the binding and stability of the biotin are provided by another oxyanion hole formed by the residues Gly419 and Ala420 in 1XNY, which corresponds to Gly405 and Ala406 in Nmar_0272. Similar to the Gly-Gly, the Gly-Ala doublet also forms an oxyanion hole. Notably, this doublet is also highly conserved in many carboxyltransferase homologs and is crucial for enzymatic activity. Moreover, rather than the residues around, the biotin is further stabilized with a carbonyl group that belongs to propionyl-CoA.

Recently, two artificial intelligence-based protein prediction methods Alpha Fold and RoseTTA Fold, have been released. As these programs claim to predict highly consistent structures with the experimental models, both Alpha Fold and RoseTTA Fold were used to predict the structure of Nmar_0272. Based on the predictions, two monomers and one dimer structures were obtained. The Alpha Fold and RoseTTA Fold prediction models revealed high similarities with our Nmar_0272 experimental model. These models were also used to determine how consistent our homology-based hybrid model was with the AI-based predictions. The structural alignments between Alpha Fold and RoseTTA Fold models with Nmar_0272/Swiss Model hybrid displayed decent structural similarities with low RMSD values. Similar results were obtained when the close-related carboxyltransferase homologs were compared with the Alpha Fold and RoseTTA Fold models. On the other hand, relatively better results and higher consistencies were obtained from the Alpha Fold predictions.

The biotin and acetyl-CoA/propionyl-CoA binding pockets of Nmar_0272 were analyzed using the 1XNY structure as a reference. For a more detailed comparison, the Alpha Fold dimer model was also used as it provided more confident results. In both Nmar_0272 and 1XNY, the biotin-binding pocket reveals highly hydrophobic residues. The comparison of these two structures displays certain residue substitutions. In 1XNY, the negatively charged Asp422 is replaced with a hydrophobic Ile408 residue in Nmar_0272. However, all the other substituted residues are replaced with alternative hydrophobic amino acids in Nmar_0272. This hydrophobicity is conserved in all the carboxyltransferase homologs compared, suggesting functional importance. Furthermore, the experimental Nmar_0272 and Alpha Fold residues displayed nearly identical side-chain positions and orientations in this pocket. However, slight differences between the residues of Alpha Fold and Swiss Model predictions were observed.

The structure of 1XNY displayed a larger binding pocket for propionyl-CoA. In our Nmar_0272/Swiss Model hybrid, the residues of the propionyl-CoA binding pocket were modeled based on homology due to low electron density. Similar to that seen in the biotin-binding pocket, some residues were also replaced by amino acids with similar properties in this binding pocket. Interestingly, certain residues in the propionyl-CoA binding pocket of 1XNY are replaced with less bulky amino acids in Nmar_0272. For instance, Phe109,

Met445, and Arg457 of 1XNY are replaced with Lys443, Leu431, Leu97 in Nmar_0272, respectively. Furthermore, the uncharged Asn80 residue of 1XNY is substituted with a positively charged Lys68 residue in Nmar_0272. These substitutions might grant flexibility to the binding pocket of Nmar_0272 and may facilitate the binding of both acetyl-CoA and propionyl-CoA. The superimposition of the Alpha Fold and homology model predictions displayed decent but relatively less consistent side-chain conformations compared to the consistency observed in the biotin-binding pocket. On the other hand, Lys68 and Gly-Gly doublet exhibited highly accurate side-chain conformations.

In conclusion, our results present a novel carboxyltransferase structure Nmar_0272, which operates in the most energy-efficient aerobic carbon fixation pathway. Our model exhibited significant similarities with the previously determined carboxyltransferase structures. On the other hand, our structure displayed certain disordered regions with low electron densities. The homology and AI-based predictions were performed to reveal these undetermined regions. Nevertheless, an experimental repeat is required to determine the flexible and disordered parts for a more reliable Nmar_0272 model. The stability of the disordered regions of Nmar_0272 can be provided with the co-expression of its missing subunits Nmar_0273 and Nmar_0274. Furthermore, co-crystallizations of Nmar_0272 with acetyl-CoA, propionyl-CoA, and biotin are also necessary to identify the binding pockets and protein substrate interactions more accurately. This study also presents the purification and crystallization procedures of Nmar_0274, however structural determination of this subunit has not been performed yet. Therefore, future studies should focus on the undetermined Nmar_0273 and Nmar_0274 subunits, together with the substrate-bound structure of the Nmar_0272, to shed light on the catalytic mechanism of acetyl-CoA/propionyl-CoA carboxylase.

BIBLIOGRAPHY

Adams, P. D., Afonine, P. v., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). *PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*, *66*(2). https://doi.org/10.1107/S0907444909052925

Alber, B. E., & Fuchs, G. (2002). Propionyl-Coenzyme A Synthase from Chloroflexus aurantiacus, a Key Enzyme of the 3-Hydroxypropionate Cycle for Autotrophic CO2 Fixation. *Journal of Biological Chemistry*, *277*(14). https://doi.org/10.1074/jbc.M110802200

Alber, B. E., Kung, J. W., & Fuchs, G. (2008). 3-Hydroxypropionyl-Coenzyme A Synthetase from Metallosphaera sedula, an Enzyme Involved in Autotrophic CO2 Fixation. *Journal of Bacteriology*, *190*(4). https://doi.org/10.1128/JB.01593-07

Alber, B., Olinger, M., Rieder, A., Kockelkorn, D., Jobst, B., Hügler, M., & Fuchs, G. (2006). Malonyl-Coenzyme A Reductase in the Modified 3-Hydroxypropionate Cycle for Autotrophic Carbon Fixation in Archaeal Metallosphaera and Sulfolobus spp. *Journal of Bacteriology*, *188*(24). https://doi.org/10.1128/JB.00987-06

Attwood PV, Wallace JC. Chemical and catalytic mechanisms of carboxyl transfer reactions in biotin-dependent enzymes. Acc Chem Res. 2002 Feb;35(2):113-20. doi: 10.1021/ar000049+. PMID: 11851389.

Audi, M., Ali, A., & Kassem, M. (2020). GREENHOUSE GASES: A REVIEW OF LOSSES AND BENEFITS. *International Journal of Energy Economics and Policy*, *10*(1). https://doi.org/10.32479/ijeep.8416

Bachofen, R., Buchanan, B. B., & Arnon, D. I. (1964). FERREDOXIN AS A REDUCTANT IN PYRUVATE SYNTHESIS BY A BACTERIAL EXTRACT. *Proceedings of the National Academy of Sciences*, *51*(4). https://doi.org/10.1073/pnas.51.4.690

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. v., van Dijk, A. A., Ebrecht, A. C., … Baker, D. (2021b). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, *373*(6557). https://doi.org/10.1126/science.abj8754

Bassham, J. A., Benson, A. A., Kay, L. D., Harris, A. Z., Wilson, A. T., & Calvin, M. (1954). The Path of Carbon in Photosynthesis. XXI. The Cyclic Regeneration of Carbon Dioxide Acceptor [1]. *Journal of the American Chemical Society*, *76*(7). https://doi.org/10.1021/ja01636a012

Bassham, J. A., & Calvin, M. (1960). The path of carbon in photosynthesis. In *Die CO2-Assimilation / The Assimilation of Carbon Dioxide*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-94798-8_30

Bassham, J. A., & Calvin, M. (1962). The way of CO2 in plant photosynthesis. *Comparative Biochemistry and Physiology*, *4*(2–4). https://doi.org/10.1016/0010-406X(62)90004-X

Beh, M., Strauss, G., Huber, R., Stetter, K.-O., & Fuchs, G. (1993). Enzymes of the reductive citric acid cycle in the autotrophic eubacterium Aquifex pyrophilus and in the archaebacterium Thermoproteus neutrophilus. *Archives of Microbiology*, *160*(4). https://doi.org/10.1007/BF00292082

Berg, I. A., Kockelkorn, D., Buckel, W., & Fuchs, G. (2007). A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea. *Science*, *318*(5857). https://doi.org/10.1126/science.1149976

Berg, I. A., Kockelkorn, D., Ramos-Vera, W. H., Say, R. F., Zarzycki, J., Hügler, M., Alber, B. E., & Fuchs, G. (2010). Autotrophic carbon fixation in archaea. *Nature Reviews Microbiology*, *8*(6). https://doi.org/10.1038/nrmicro2365

Berg, I. A., Ramos-Vera, W. H., Petri, A., Huber, H., & Fuchs, G. (2010). Study of the distribution of autotrophic CO2 fixation cycles in Crenarchaeota. *Microbiology*, *156*(1). https://doi.org/10.1099/mic.0.034298-0

Brochier-Armanet, C., Boussau, B., Gribaldo, S., & Forterre, P. (2008). Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Reviews Microbiology*, *6*(3). https://doi.org/10.1038/nrmicro1852

Buchanan, B. B., & Arnon, D. I. (1990). A reverse KREBS cycle in photosynthesis: consensus at last. *Photosynthesis Research*, *24*(1). https://doi.org/10.1007/BF00032643

Buchanan, B. B., & Evans, M. C. (1965). The synthesis of alpha-ketoglutarate from succinate and carbon dioxide by a subcellular preparation of a photosynthetic bacterium. *Proceedings of the National Academy of Sciences*, *54*(4). https://doi.org/10.1073/pnas.54.4.1212

Buckel, W., & Golding, B. T. (2006). Radical Enzymes in Anaerobes. *Annual Review of Microbiology*, *60*(1). https://doi.org/10.1146/annurev.micro.60.080805.142216

Cane, D. E. (1998). Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations. *Science*, *282*(5386). https://doi.org/10.1126/science.282.5386.63

Chaban, B., Ng, S. Y. M., & Jarrell, K. F. (2006). Archaeal habitats — from the extreme to the ordinary. *Canadian Journal of Microbiology*, *52*(2). https://doi.org/10.1139/w05-147

Chayen, N. E., & Saridakis, E. (2008). Protein crystallization: from purified protein to diffraction-quality crystal. *Nature Methods*, *5*(2). https://doi.org/10.1038/nmeth.f.203

Chuakrut, S., Arai, H., Ishii, M., & Igarashi, Y. (2003). Characterization of a Bifunctional Archaeal Acyl Coenzyme A Carboxylase. *Journal of Bacteriology*, *185*(3). https://doi.org/10.1128/JB.185.3.938-947.2003

Clamp, M., Cuff, J., Searle, S. M., & Barton, G. J. (2004). The Jalview Java alignment editor. *Bioinformatics*, *20*(3). https://doi.org/10.1093/bioinformatics/btg430

Clark JE, Ljungdahl LG. Purification and properties of 5,10-methylenetetrahydrofolate reductase, an iron-sulfur flavoprotein from Clostridium formicoaceticum. J Biol Chem. 1984 Sep 10;259(17):10845-9. PMID: 6381490.

Clark JE, Ljungdahl LG. Purification and properties of 5,10-methenyltetrahydrofolate cyclohydrolase from Clostridium formicoaceticum. J Biol Chem. 1982 Apr 10;257(7):3833-6. PMID: 7061514.

Cowtan, K. (2003). Phase Problem in X-ray Crystallography, and Its Solution. In *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd. https://doi.org/10.1038/npg.els.0002722

Cronan, Jr. , J. E., & Rock, C. O. (2008). Biosynthesis of Membrane Lipids. *EcoSal Plus*, *3*(1). https://doi.org/10.1128/ecosalplus.3.6.4

Diacovich, L., Mitchell, D. L., Pham, H., Gago, G., Melgar, M. M., Khosla, C., Gramajo, H., & Tsai, S.-C. (2004c). Crystal Structure of the *β* -Subunit of Acyl-CoA Carboxylase: Structure-Based Engineering of Substrate Specificity † , ‡. *Biochemistry*, *43*(44). https://doi.org/10.1021/bi049065v

Drake, H. L., Gößner, A. S., & Daniel, S. L. (2008). Old Acetogens, New Light. *Annals of the New York Academy of Sciences*, *1125*(1). https://doi.org/10.1196/annals.1419.016

Eisenreich, W., Strauss, G., Werz, U., Fuchs, G., & Bacher, A. (1993). Retrobiosynthetic analysis of carbon fixation in the phototrophic eubacterium Chloroflexus aurantiacus. *European Journal of Biochemistry*, *215*(3). https://doi.org/10.1111/j.1432-1033.1993.tb18073.x

Emsley, P., & Cowtan, K. (2004). *Coot* : model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography*, *60*(12). https://doi.org/10.1107/S0907444904019158

Erb, T. J., & Zarzycki, J. (2018). A short history of RubisCO: the rise and fall (?) of Nature's predominant CO2 fixing enzyme. *Current Opinion in Biotechnology*, *49*. https://doi.org/10.1016/j.copbio.2017.07.017

Estelmann, S., Hu gler, M., Eisenreich, W., Werner, K., Berg, I. A., Ramos-Vera, W. H., Say, R. F., Kockelkorn, D., Gad'on, N., & Fuchs, G. (2011). Labeling and Enzyme Studies of the Central Carbon Metabolism in Metallosphaera sedula. *Journal of Bacteriology*, *193*(5). https://doi.org/10.1128/JB.01155-10

Ettema, T. J. G., Makarova, K. S., Jellema, G. L., Gierman, H. J., Koonin, E. v., Huynen, M. A., de Vos, W. M., & van der Oost, J. (2004). Identification and Functional Verification of Archaeal-Type Phosphoenolpyruvate Carboxylase, a Missing Link in Archaeal Central Carbohydrate Metabolism. *Journal of Bacteriology*, *186*(22). https://doi.org/10.1128/JB.186.22.7754-7762.2004

Evans, M. C., Buchanan, B. B., & Arnon, D. I. (1966). A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proceedings of the National Academy of Sciences*, *55*(4). https://doi.org/10.1073/pnas.55.4.928

Field, C. B. , B. M. J. , R. J. T. , F. P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, *281*(5374). https://doi.org/10.1126/science.281.5374.237

Fuchs, G. (2011). Alternative Pathways of Carbon Dioxide Fixation: Insights into the Early Evolution of Life? *Annual Review of Microbiology*, *65*(1). https://doi.org/10.1146/annurev-micro-090110-102801

Fukuda, W., Fukui, T., Atomi, H., & Imanaka, T. (2004). First Characterization of an Archaeal GTP-Dependent Phosphoenolpyruvate Carboxykinase from the Hyperthermophilic Archaeon Thermococcus kodakaraensis KOD1. *Journal of Bacteriology*, *186*(14). https://doi.org/10.1128/JB.186.14.4620-4627.2004

Gerhardt, A., Çinkaya, I., Linder, D., Huisman, G., & Buckel, W. (2000). Fermentation of 4-aminobutyrate by Clostridium aminobutyricum : cloning of two genes involved in the formation and dehydration of 4-hydroxybutyryl-CoA. *Archives of Microbiology*, *174*(3). https://doi.org/10.1007/s002030000195

Grace, J. (2004). Understanding and managing the global carbon cycle. *Journal of Ecology*, *92*(2). https://doi.org/10.1111/j.0022-0477.2004.00874.x

Hallam, S. J., Mincer, T. J., Schleper, C., Preston, C. M., Roberts, K., Richardson, P. M., & DeLong, E. F. (2006). Pathways of Carbon Assimilation and Ammonia Oxidation Suggested by Environmental Genomic Analyses of Marine Crenarchaeota. *PLoS Biology*, *4*(4). https://doi.org/10.1371/journal.pbio.0040095

Hawkins, A. B., Adams, M. W. W., & Kelly, R. M. (2014). Conversion of 4-Hydroxybutyrate to Acetyl Coenzyme A and Its Anapleurosis in the Metallosphaera sedula 3-Hydroxypropionate/4-Hydroxybutyrate Carbon Fixation Pathway. *Applied and Environmental Microbiology*, *80*(8). https://doi.org/10.1128/AEM.04146-13

Herter, S., Busch, A., & Fuchs, G. (2002). l-Malyl-Coenzyme A Lyase/β-Methylmalyl-Coenzyme A Lyase from Chloroflexus aurantiacus, a Bifunctional Enzyme Involved in Autotrophic CO2 Fixation. *Journal of Bacteriology*, *184*(21). https://doi.org/10.1128/JB.184.21.5999-6006.2002

Herter, S., Farfsing, J., Gad'On, N., Rieder, C., Eisenreich, W., Bacher, A., & Fuchs, G. (2001). Autotrophic CO2 Fixation by Chloroflexus aurantiacus: Study of Glyoxylate Formation and Assimilation via the 3-Hydroxypropionate Cycle. *Journal of Bacteriology*, *183*(14). https://doi.org/10.1128/JB.183.14.4305-4316.2001

Herter, S., Fuchs, G., Bacher, A., & Eisenreich, W. (2002). A Bicyclic Autotrophic CO2 Fixation Pathway in Chloroflexus aurantiacus. *Journal of Biological Chemistry*, *277*(23). https://doi.org/10.1074/jbc.M201030200

Himes, R. H., & Cohn, M. (1967). Formyltetrahydrofolate Synthetase. *Journal of Biological Chemistry*, *242*(16). https://doi.org/10.1016/S0021-9258(18)95855-2

Himes, R. H., Harmony, J. A. K., & Rabinowitz, J. C. (1973). Formyltetrahydrofolate Synthetas. *CRC Critical Reviews in Biochemistry*, *1*(4). https://doi.org/10.3109/10409237309105441

Holo, H. (1989). Chloroflexus aurantiacus secretes 3-hydroxypropionate, a possible intermediate in the assimilation of CO2 and acetate. *Archives of Microbiology*, *151*(3). https://doi.org/10.1007/BF00413138

Holo, H., & Sirevg, R. (1986). Autotrophic growth and CO2 fixation of Chloroflexus aurantiacus. *Archives of Microbiology*, *145*(2). https://doi.org/10.1007/BF00446776

Huber, G., Spinnler, C., Gambacorta, A., & Stetter, K. O. (1989). Metallosphaera sedula gen, and sp. nov. Represents a New Genus of Aerobic, Metal-Mobilizing, Thermoacidophilic

Archaebacteria. *Systematic and Applied Microbiology*, *12*(1). https://doi.org/10.1016/S0723-2020(89)80038-4

Huber, H., Gallenberger, M., Jahn, U., Eylert, E., Berg, I. A., Kockelkorn, D., Eisenreich, W., & Fuchs, G. (2008). A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum Ignicoccus hospitalis. *Proceedings of the National Academy of Sciences*, *105*(22). https://doi.org/10.1073/pnas.0801043105

Hügler, M., & Fuchs, G. (2005). *Assaying for the 3-Hydroxypropionate Cycle of Carbon Fixation*. https://doi.org/10.1016/S0076-6879(05)97012-2

Hügler, M., Krieger, R. S., Jahn, M., & Fuchs, G. (2003). Characterization of acetyl-CoA/propionyl-CoA carboxylase in *Metallosphaera sedula*. *European Journal of Biochemistry*, *270*(4). https://doi.org/10.1046/j.1432-1033.2003.03434.x

Hügler, M., Menendez, C., Schägger, H., & Fuchs, G. (2002). Malonyl-Coenzyme A Reductase from Chloroflexus aurantiacus, a Key Enzyme of the 3-Hydroxypropionate Cycle for Autotrophic CO2 Fixation. *Journal of Bacteriology*, *184*(9). https://doi.org/10.1128/JB.184.9.2404-2410.2002

Hügler, M., & Sievert, S. M. (2011). Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. *Annual Review of Marine Science*, *3*(1). https://doi.org/10.1146/annurev-marine-120709-142712

Hügler, M., Wirsen, C. O., Fuchs, G., Taylor, C. D., & Sievert, S. M. (2005). Evidence for Autotrophic CO2 Fixation via the Reductive Tricarboxylic Acid Cycle by Members of the ε Subdivision of Proteobacteria. *Journal of Bacteriology*, *187*(9). https://doi.org/10.1128/JB.187.9.3020-3027.2005

Ivanovsky, R. N., Krasilnikova, E. N., & Fal, Y. I. (1993). A pathway of the autotrophic CO2 fixation in Chloroflexus aurantiacus. *Archives of Microbiology*, *159*(3). https://doi.org/10.1007/BF00248481

Ivanovsky, R. N., Sintsov, N. v., & Kondratieva, E. N. (1980). ATP-linked citrate lyase activity in the green sulfur bacterium Chlorobium limicola forma thiosulfatophilum. *Archives of Microbiology*, *128*(2). https://doi.org/10.1007/BF00406165

Johnson, C. E., & Derwent, R. G. (1996). Relative radiative forcing consequences of global emissions of hydrocarbons, carbon monoxide and NOx from human activities estimated with a zonally-averaged two-dimensional model. *Climatic Change*, *34*(3–4). https://doi.org/10.1007/BF00139301

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021a). Highly accurate protein structure prediction with AlphaFold. *Nature*. https://doi.org/10.1038/s41586-021-03819-2

Kanao, T., Kawamura, M., Fukui, T., Atomi, H., & Imanaka, T. (2002). Characterization of isocitrate dehydrogenase from the green sulfur bacterium *Chlorobium limicola*. *European Journal of Biochemistry*, *269*(7). https://doi.org/10.1046/j.1432-1033.2002.02849.x

Khalil, M. A. K. (1999). Non-CO2 Greenhouse Gases In The Atmosphere. *Annual Review of Energy and the Environment*, *24*(1). https://doi.org/10.1146/annurev.energy.24.1.645

Knowles, J. R. (1989). The Mechanism of Biotin-Dependent Enzymes. *Annual Review of Biochemistry*, *58*(1). https://doi.org/10.1146/annurev.bi.58.070189.001211

Kockelkorn, D., & Fuchs, G. (2009). Malonic Semialdehyde Reductase, Succinic Semialdehyde Reductase, and Succinyl-Coenzyme A Reductase from Metallosphaera sedula: Enzymes of the Autotrophic 3-Hydroxypropionate/4-Hydroxybutyrate Cycle in Sulfolobales. *Journal of Bacteriology*, *191*(20). https://doi.org/10.1128/JB.00794-09

Könneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., & Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*, *437*(7058). https://doi.org/10.1038/nature03911

Konneke, M., Schubert, D. M., Brown, P. C., Hugler, M., Standfest, S., Schwander, T., Schada von Borzyskowski, L., Erb, T. J., Stahl, D. A., & Berg, I. A. (2014). Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation. *Proceedings of the National Academy of Sciences*, *111*(22). https://doi.org/10.1073/pnas.1402028111

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., Prosser, J. I., Schuster, S. C., & Schleper, C. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*, *442*(7104). https://doi.org/10.1038/nature04983

Li, L.-F., Ljungdahl, L., & Wood, H. G. (1966). Properties of Nicotinamide Adenine Dinucleotide Phosphate-Dependent Formate Dehydrogenase from Clostridium thermoaceticum. *Journal of Bacteriology*, *92*(2). https://doi.org/10.1128/JB.92.2.405-412.1966

Lin, T.-W., Melgar, M. M., Kurth, D., Swamidass, S. J., Purdon, J., Tseng, T., Gago, G., Baldi, P., Gramajo, H., & Tsai, S.-C. (2006a). Structure-based inhibitor design of AccD5, an essential acyl-CoA carboxylase carboxyltransferase domain of Mycobacterium

tuberculosis. *Proceedings of the National Academy of Sciences*, *103*(9). https://doi.org/10.1073/pnas.0510580103

Liu, D., Guo, X., & Xiao, B. (2019). What causes growth of global greenhouse gas emissions? Evidence from 40 countries. *Science of The Total Environment*, *661*. https://doi.org/10.1016/j.scitotenv.2019.01.197

Liu, L., Brown, P. C., Könneke, M., Huber, H., König, S., & Berg, I. A. (2021). Convergent Evolution of a Promiscuous 3-Hydroxypropionyl-CoA Dehydratase/Crotonyl-CoA Hydratase in *Crenarchaeota* and *Thaumarchaeota*. *MSphere*, *6*(1). https://doi.org/10.1128/mSphere.01079-20

Ljungdahl, L. G. (2009). A Life with Acetogens, Thermophiles, and Cellulolytic Anaerobes. *Annual Review of Microbiology*, *63*(1). https://doi.org/10.1146/annurev.micro.091208.073617

Ljungdahl, L. G., & Wood, H. G. (1969). Total Synthesis of Acetate From CO2 by Heterotrophic Bacteria. *Annual Review of Microbiology*, *23*(1). https://doi.org/10.1146/annurev.mi.23.100169.002503

Ljungdhal, L. G. (1986). The Autotrophic Pathway of Acetate Synthesis in Acetogenic Bacteria. *Annual Review of Microbiology*, *40*(1). https://doi.org/10.1146/annurev.mi.40.100186.002215

Mariani, V., Biasini, M., Barbato, A., & Schwede, T. (2013a). lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, *29*(21). https://doi.org/10.1093/bioinformatics/btt473

Martens-Habbena, W., Berube, P. M., Urakawa, H., de la Torre, J. R., & Stahl, D. A. (2009). Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature*, *461*(7266). https://doi.org/10.1038/nature08465

Martins, B. M., Dobbek, H., Cinkaya, I., Buckel, W., & Messerschmidt, A. (2004). Crystal structure of 4-hydroxybutyryl-CoA dehydratase: Radical catalysis involving a [4Fe-4S] cluster and flavin. *Proceedings of the National Academy of Sciences*, *101*(44). https://doi.org/10.1073/pnas.0403952101

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). *Phaser* crystallographic software. *Journal of Applied Crystallography*, *40*(4). https://doi.org/10.1107/S0021889807021206

McPherson, A. (2009). *Introduction to Macromolecular Crystallography*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470391518

Michelet, L., Zaffagnini, M., Morisse, S., Sparla, F., Pérez-Pérez, M. E., Francia, F., Danon, A., Marchand, C. H., Fermani, S., Trost, P., & Lemaire, S. D. (2013). Redox regulation of the Calvin–Benson cycle: something old, something new. *Frontiers in Plant Science*, *4*. https://doi.org/10.3389/fpls.2013.00470

Moore MR, O'Brien WE, Ljungdahl LG. Purification and characterization of nicotinamide adenine dinucleotide-dependent methylenetetrahydrofolate dehydrogenase from Clostridium formicoaceticum. J Biol Chem. 1974 Aug 25;249(16):5250-3. PMID: 4153026.

Moss, J., & Lane, M. D. (1971). *The Biotin-Dependent Enzymes*. Adv Enzymol Relat Areas Mol Biol. 1971;35:321-442. doi: 10.1002/9780470122808.ch7. PMID: 4150153.

Patel, H. M., Kraszewski, J. L., & Mukhopadhyay, B. (2004). The Phosphoenolpyruvate Carboxylase from Methanothermobacter thermautotrophicus Has a Novel Structure. *Journal of Bacteriology*, *186*(15). https://doi.org/10.1128/JB.186.15.5129-5137.2004

Quayle, J. R., Fuller, R. C., Benson, A. A., & Calvin, M. (1954). ENZYMATIC CARBOXYLATION OF RIBULOSE DIPHOSPHATE [1]. *Journal of the American Chemical Society*, *76*(13). https://doi.org/10.1021/ja01642a089

Ragsdale, S. W., & Pierce, E. (2008). Acetogenesis and the Wood–Ljungdahl pathway of CO2 fixation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1784*(12). https://doi.org/10.1016/j.bbapap.2008.08.012

Ramos-Vera, W. H., Berg, I. A., & Fuchs, G. (2009). Autotrophic Carbon Dioxide Assimilation in Thermoproteales Revisited. *Journal of Bacteriology*, *191*(13). https://doi.org/10.1128/JB.00145-09

Schauder, R., Widdel, F., & Fuchs, G. (1987). Carbon assimilation pathways in sulfate-reducing bacteria II. Enzymes of a reductive citric acid cycle in the autotrophic Desulfobacter hydrogenophilus. *Archives of Microbiology*, *148*(3). https://doi.org/10.1007/BF00414815

Scherf, U., S&#xF;hling, B., Gottschalk, G., Linder, D., & Buckel, W. (1994). Succinate-ethanol fermentation in Clostridium kluyveri: purification and characterisation of 4-hydroxybutyryl-CoA dehydratase/vinylacetyl-CoA ?3-?2-isomerase. *Archives of Microbiology*, *161*(3). https://doi.org/10.1007/BF00248699

Sellers, P. J., Schimel, D. S., Moore, B., Liu, J., & Eldering, A. (2018). Observing carbon cycle–climate feedbacks from space. *Proceedings of the National Academy of Sciences*, *115*(31). https://doi.org/10.1073/pnas.1716613115

Sharkey, T. D. (2019). Discovery of the canonical Calvin–Benson cycle. *Photosynthesis Research*, *140*(2). https://doi.org/10.1007/s11120-018-0600-2

Shiba, H., Kawasumi, T., Igarashi, Y., Kodama, T., & Minoda, Y. (1985). The CO2 assimilation via the reductive tricarboxylic acid cycle in an obligately autotrophic, aerobic hydrogen-oxidizing bacterium, Hydrogenobacter thermophilus. *Archives of Microbiology*, *141*(3). https://doi.org/10.1007/BF00408058

Smyth, M. S. (2000). x Ray crystallography. *Molecular Pathology*, *53*(1). https://doi.org/10.1136/mp.53.1.8

Song, L., & Gouaux, J. E. (1997). *[4] Membrane protein crystallization: Application of sparse matrices to the α-hemolysin heptamer*. https://doi.org/10.1016/S0076-6879(97)76050-6

Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit, W., Stahl, D. A., Wagner, M., & Schleper, C. (2010). Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends in Microbiology*, *18*(8). https://doi.org/10.1016/j.tim.2010.06.003

Stahl, D. A., & de la Torre, J. R. (2012). Physiology and Diversity of Ammonia-Oxidizing Archaea. *Annual Review of Microbiology*, *66*(1). https://doi.org/10.1146/annurev-micro-092611-150128

Strauss, G., Eisenreich, W., Bacher, A., & Fuchs, G. (1992). 13C-NMR study of autotrophic CO2 fixation pathways in the sulfur-reducing Archaebacterium Thermoproteus neutrophilus and in the phototrophic Eubacterium Chloroflexus aurantiacus. *European Journal of Biochemistry*, *205*(2). https://doi.org/10.1111/j.1432-1033.1992.tb16850.x

Strauss, G., & Fuchs, G. (1993). Enzymes of a novel autotrophic CO2 fixation pathway in the phototrophic bacterium Chloroflexus aurantiacus, the 3-hydroxypropionate cycle. *European Journal of Biochemistry*, *215*(3). https://doi.org/10.1111/j.1432-1033.1993.tb18074.x

Teufel, R., Kung, J. W., Kockelkorn, D., Alber, B. E., & Fuchs, G. (2009). 3-Hydroxypropionyl-Coenzyme A Dehydratase and Acryloyl-Coenzyme A Reductase, Enzymes of the Autotrophic 3-Hydroxypropionate/4-Hydroxybutyrate Cycle in the Sulfolobales. *Journal of Bacteriology*, *191*(14). https://doi.org/10.1128/JB.00068-09

Thauer, R. K. (1972). CO_2 -reduction to formate by NADPH. The initial step in the total synthesis of acetate from CO_2 in *Clostridium thermoaceticum*. *FEBS Letters*, *27*(1). https://doi.org/10.1016/0014-5793(72)80421-6

Thompson, J. D., Gibson, Toby. J., & Higgins, D. G. (2003). Multiple Sequence Alignment Using ClustalW and ClustalX. *Current Protocols in Bioinformatics*, *00*(1). https://doi.org/10.1002/0471250953.bi0203s00

Tong, L. (2013). Structure and function of biotin-dependent carboxylases. *Cellular and Molecular Life Sciences*, *70*(5). https://doi.org/10.1007/s00018-012-1096-0

Udara Willhelm Abeydeera, L. H., Wadu Mesthrige, J., & Samarasinghalage, T. I. (2019). Global Research on Carbon Emissions: A Scientometric Review. *Sustainability*, *11*(14). https://doi.org/10.3390/su11143972

Vorholt, J. A., Hafenbradl, D., Stetter, K. O., & Thauer, R. K. (1997). Pathways of autotrophic CO 2 fixation and of dissimilatory nitrate reduction to N 2 O in Ferroglobus placidus. *Archives of Microbiology*, *167*(1). https://doi.org/10.1007/s002030050411

Vornolt, J., Kunow, J., Stetter, K. O., & Thauer, R. K. (1995). Enzymes and coenzymes of the carbon monoxide dehydrogenase pathway for autotrophic CO2 fixation in Archaeoglobus lithotrophicus and the lack of carbon monoxide dehydrogenase in the heterotrophic A. profundus. *Archives of Microbiology*, *163*(2). https://doi.org/10.1007/BF00381784

Walker, C. B., de la Torre, J. R., Klotz, M. G., Urakawa, H., Pinel, N., Arp, D. J., Brochier-Armanet, C., Chain, P. S. G., Chan, P. P., Gollabgir, A., Hemp, J., Hugler, M., Karr, E. A., Konneke, M., Shin, M., Lawton, T. J., Lowe, T., Martens-Habbena, W., Sayavedra-Soto, L. A., … Stahl, D. A. (2010). Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proceedings of the National Academy of Sciences*, *107*(19). https://doi.org/10.1073/pnas.0913533107

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, *74*(11). https://doi.org/10.1073/pnas.74.11.5088

Wolosiuk, R. A., Ballicora, M. A., & Hagelin, K. (1993). The reductive pentose phosphate cycle for photosynthetic CO $_2$ assimilation: enzyme modulation. *The FASEB Journal*, *7*(8). https://doi.org/10.1096/fasebj.7.8.8500687

Wood, H. G. (1991). Life with CO or CO $_2$ and H $_2$ as a source of carbon and energy. *The FASEB Journal*, *5*(2). https://doi.org/10.1096/fasebj.5.2.1900793

Wuchter, C., Abbas, B., Coolen, M. J. L., Herfort, L., van Bleijswijk, J., Timmers, P., Strous, M., Teira, E., Herndl, G. J., Middelburg, J. J., Schouten, S., & Sinninghe Damste, J. S.

(2006). Archaeal nitrification in the ocean. *Proceedings of the National Academy of Sciences*, *103*(33). https://doi.org/10.1073/pnas.0600756103

Yakimov, M. M., Cono, V. la, Smedile, F., DeLuca, T. H., Juárez, S., Ciordia, S., Fernández, M., Albar, J. P., Ferrer, M., Golyshin, P. N., & Giuliano, L. (2011). Contribution of crenarchaeal autotrophic ammonia oxidizers to the dark primary production in Tyrrhenian deep waters (Central Mediterranean Sea). *The ISME Journal*, *5*(6). https://doi.org/10.1038/ismej.2010.197

Zarzycki, J., Brecht, V., Müller, M., & Fuchs, G. (2009). Identifying the missing steps of the autotrophic 3-hydroxypropionate CO $_2$ fixation cycle in *Chloroflexus aurantiacus*. *Proceedings of the National Academy of Sciences*, *106*(50), 21317–21322. https://doi.org/10.1073/pnas.0908356106

APPENDIX A: Chemicals

Chemicals and Media Components	Supplier Company
Acetic Acid	ISOLAB
Acrylamide/Bis-acrylamide	Bio-Rad
APS	Sigma-Aldrich
BME	Merck
Chloramphenicol	GOLDBIO
Coomassie Brilliant Blue	Bio-Rad
DTT	Roche
Ethanol	ISOLAB
Glycerol	ISOLAB
Glycine	ISOLAB
Imidazole	Biofroxx
IPTG	GOLDBIO
Kanamycin	GOLDBIO
LB	Thermo Fisher
LB Agar	Caisson
NaCl	ISOLAB
NaOH	ISOLAB
HCl	ISOLAB
Ni-NTA Resin	QIAGEN
Paraffin Oil	TEKKİM
SDS	Sigma-Aldrich
TEMED	Sigma-Aldrich
Thrombin Protease	Sigma-Aldrich
Tris-HCl	Sigma-Aldrich
Triton X-100	Sigma-Aldrich
Tryptone	BD Biosciences
Yeast Extract	BD Biosciences

APPENDIX B: Crystal Screening Kits

Screening Kit	Supplier Company
3D Structure Screen	Molecular Dimensions
Additive Screen	Hampton Research
Clear Strategy Screen	Molecular Dimensions
Crystal Screen Cryo	Hampton Research
Crystal Screen	Hampton Research
Grid Screen Ammonium Sulfate	Hampton Research
Grid Screen MPD	Hampton Research
Grid Screen Sodium Chloride	Hampton Research
Grid Screen PEG/LiCl	Hampton Research
Grid Screen PEG6000	Hamptron Research
Grid Screen Sodium Malonate	Hampton Research
Helix	Molecular Dimensions
Index	Hampton Research
Ionic Liquid Screen	Hampton Research
JBScreen Nuc-Pro	Jena Bioscience
JCSG	Molecular Dimensions
MacroSol	Molecular Dimensions
MembFac	Hampton Research
MIDAS	Molecular Dimensions
Morpheus	Molecular Dimensions
MultiXtal	Molecular Dimensions
Natrix	Hampton Research
NeXtal Protein Complex Suite	NeXtal Biotechnologies
NR-LBD	Molecular Dimensions
PACT Premier	Molecular Dimensions
PEG/Ion	Hampton Research
PEGRx	Hampton Research
PGA Screen	Molecular Dimensions

ProPlex Screen	Molecular Dimensions
Quik Screen	Hampton Research
SaltRx	Hampton Research
Structure Screen	Molecular Dimensions
Stura FootPrint Screen	Molecular Dimensions
Wizard	Rigaku
Wizard Cryo	Rigaku
Wizard Precipitant Synergy	Rigaku

APPENDIX C: Equipment

Equipment	**Supplier Company**
AKTA GO	Cytiva
AKTA Prime	Cytiva
Beckman Allegra 15R Centrifuge	Beckman Coulter
Beckman Avanti J-26S	Beckman Coulter
Beckman OptimaTM L-80 XP Ultracentrifuge	Beckman Coulter
Beckman Ti45 Rotor	Beckman Coulter
Branson W250 Sonifier	Branson
Cellulose Dialysis Tubing	Thermo Scientific
Cryovials	Wuxi NEST Biotechnology
Falcon Tubes	FIRATMED
Innova 44R	Eppendorf New Brunswick
Innova 4430R	Eppendorf New Brunswick
Kimwipes	Kimberly-Clark
Nanodrop2000c	Thermo Scientific
Superdex 200 Increase 10/300 GL	Cytiva
Terasaki Plate	Greiner Bio
TGX-Mini Protean Gradient SDS-PAGE System	Bio-Rad
Ultra-Pure Water System	Merck Millipore

Appendix D: Protein Ladders

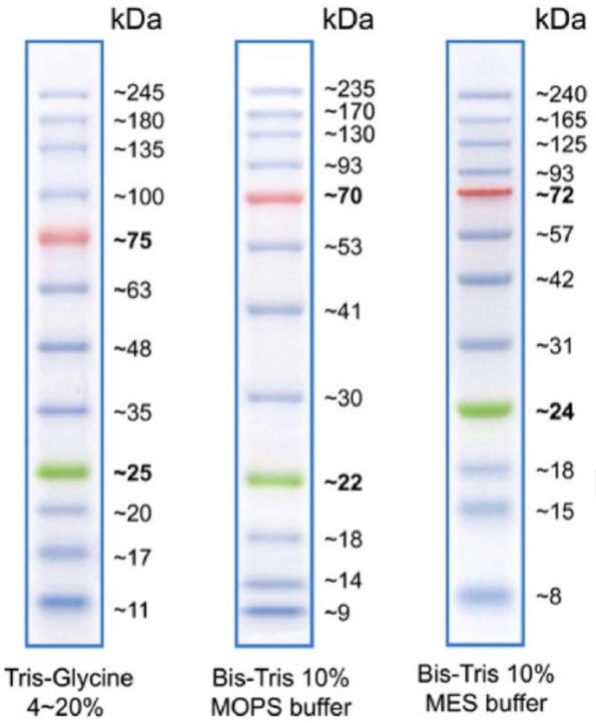


Figure D1: BLUEstain™ Protein Ladder, 11-245 kDa, GOLDBIO, USA

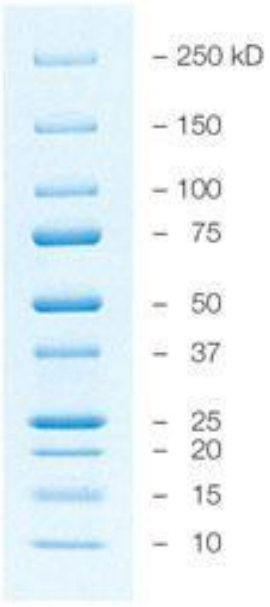
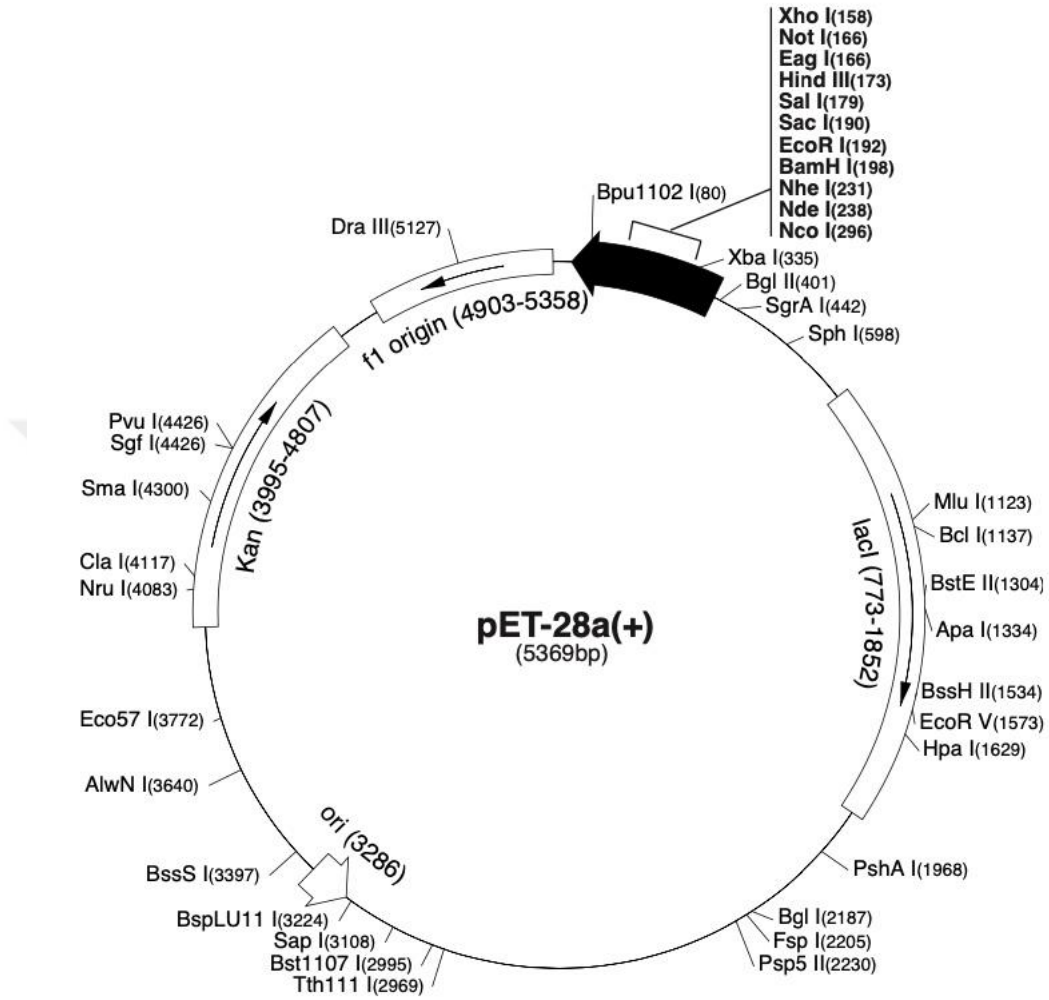


Figure D2: Precision Plus™ All Blue Prestained Protein Standards, BIO-RAD, USA

Appendix E: Plasmid Map



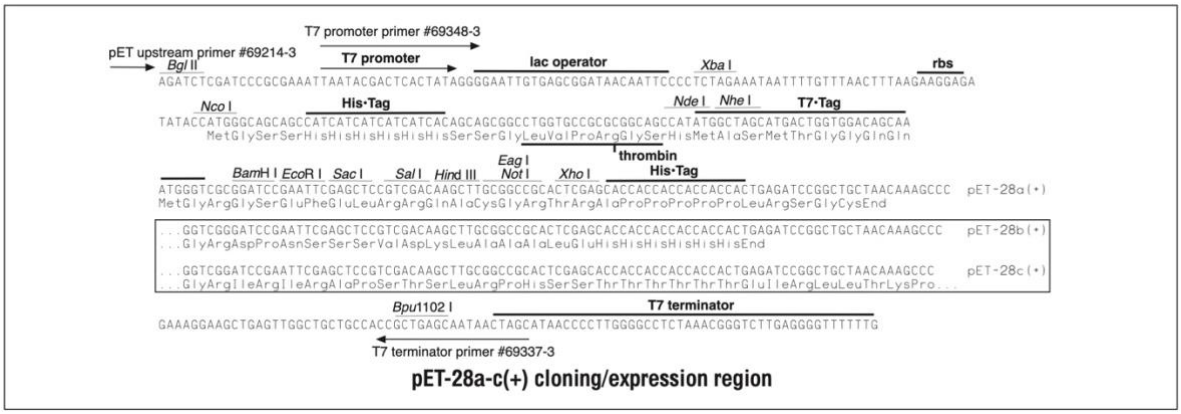


Figure E1: The plasmid map of pET-28a-c(+), NOVAGEN, USA

Retrieved from https://www.merckmillipore.com/TR/tr/product/pET-28a+-DNA-Novagen,EMD_BIO-69864?ReferrerURL=https%3A%2F%2Fwww.google.com%2F