Structural Dynamics Studies of Wild-type and Mutant *Candida boidinii* NAD+-dependent Formate Dehydrogenase

by

Büşra YÜKSEL

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

Master of Science

in

Molecular Biology and Genetics

July 19, 2022

Structural Dynamics Studies of Wild-type and Mutant *Candida boidinii* NAD+-dependent Formate Dehydrogenase

Koç University

Graduate School of Sciences

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

Büşra Yüksel

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)

Prof. Dr. Özlem Keskin

Prof. Dr. Gizem Dinler Doğanay

Date: 28.07.2022

To my friends…To lost loved ones…

To daylight after the darkest hours, which help me to survive…

ABSTRACT

MSc Thesis

Structural Dynamics Studies of Wild-type and Mutant *Candida boidinii*

NAD+-dependent Formate Dehydrogenase

Master of Science in Molecular Biology and Genetics

July 28, 2022

Candida boidinii NAD+-dependent formate dehydrogenase (cbFDH) is a biotechnologically invaluable enzyme that recently gained more attention on biosynthetic applications for biofuel production. Structure-based mutants of this enzyme performed highly promising results of enhanced catalytic activity and thermal stability. In this study, we aimed to employ a holistic approach and combine X-ray crystallographic analysis of NAD+-dependent cboFDH with Gaussian Network Model analysis to determine the underpinnings of the cboFDH catalytic mechanism and allostery. Additionally, Alphafold2 predictions of a mutant form of cbFDH enzyme were discussed for mutation-based structure prediction efficiency. The 1.4 A resolution X-ray crystallography data of NAD+-dependent cbFDH revealed the structural dynamics of the enzyme with the assist of the computational approach. This understanding of enzymatic dynamics and the effect of mutations on structure and function may provide new insight for future biotechnology studies.

ÖZETÇE

Yüksek Lisans Tezi

Yabanıl-tip ve Mutant *Candida Boidinii* NAD+-bağımlı Formate Dehydrogenase Enziminin Yapısal Dinamik Çalışmaları

Moleküler Biyoloji ve Genetik, Yüksek Lisans

Temmuz 28, 2022

Candida boidinii NAD+-bağımlı format dehidrogenaz (CbFDH), son yıllarda biyoyakıt üretimi için biyosentetik uygulamalarda dikkat çeken, biyoteknolojik olarak paha biçilmez bir enzimdir. Bu enzimin yapı bazlı mutantları, geliştirilmiş katalitik aktivite ve termal stabilite açısından oldukça umut verici sonuçlar vermiştir. Bu çalışmada, bütünsel bir yaklaşım kullanarak, CbFDH katalitik mekanizmasının ve allosterisinin temellerini belirlemek için, NAD+-bağımlı CbFDH'nin X-ışını kristalografik analizini Gaussian Ağ Modeli (GNM) analizi ile birleştirmesi amaçlanmıştır. Ek olarak, mutasyona dayalı yapı tahmin verimliliği için CbFDH enziminin mutant formunun Alphafold2 tahminlerinin verimliliği tartışılmıştır. NAD+-bağımlı CbFDH'nin 1.4 A çözünürlüklü X-ışını kristalografi verileri, bilgisayarlı hesaplama yaklaşımının da yardımıyla enzimin yapısal dinamiklerini ortaya çıkarmıştır. Enzimatik dinamiklerin ve mutasyonların yapı ve fonksiyon üzerindeki etkisinin bu şekilde anlaşılması, gelecekteki biyoteknoloji çalışmaları için yeni bilgiler sağlayabilecektir.

ACKNOWLEDGMENTS

I would dedicate this thesis to our beloved lecturer Prof. Dr. Fatma Neşe Kök, who touched our lives by not only educating but also showing a bright path.

I would like to thank Dr. Huri Bulut for tutoring me for years and being the perfect sister that I never had and providing wild-type and mutant forms of CbFDH genes. I thank Dr. Çağdaş Dağ for leading me as a scientific brother and teaching all procedures and troubleshooting during my master's study.

I would like to thank my beloved friend Meryem Eren for teaching me on Gaussian Network Model (GNM) analysis with her all patience, providing the GNM script, and showing me troubleshooting when I stuck. I could not complete the GNM analysis without you and with my ignorance of programming on this quality.

I am also grateful to my brother like friends and mates Mehmet Gül, Can Özdoğar, Ömür Güven, and Barış Çakılkaya for supporting me in this crazy journey both personally and academically. Love you all guys. <3 I had lots of unforgettable memories in this jungle.

TABLE OF CONTENTS

List of Tables ix

List of Figures x

Abbreviations xii

Chapter 1: Introduction 1

1.1 Formate 1

1.1.1 Formate in Metabolism 1

1.1.2 Formate Level in Clinic 4

1.1.3 Formate in Biotechnological Applications 6

1.2 NAD+ Coenzyme 8

1.2.1 NAD+ Biosynthesis and Oxidoreductase Reactions 9

1.3 Formate Dehydrogenase 10

1.3.1 NAD+-dependent Formate Dehydrogenase 10

1.4 Candida boidinii NAD+-dependent Formate Dehydrogenase 12

1.5 Practical usage of FDH 14

1.6 Protein Engineering 16

1.6.1 Rational Design 17

1.6.2 Directed Evolution 18

1.6.3 Semi-Rational Design 19

1.7 Candida Boidinii NAD+-dependent FDH Protein Engineering Studies 19

1.7.1 Enzymatic activity enhancement of Candida Boidinii NAD+-dependent FDH 19

1.7.2 Thermal stability enhancement of Candida Boidinii NAD+-dependent FDH 21

1.8 Structural Biology 22

1.8.1 X-Ray Crystallography 23

1.8.2 Structural Biology Studies of Candida Boidinii NAD+-dependent FDH 26

Chapter 2: Methodology 27

2.1 Materials and equipment 27

2.2 Bacterial transformation 27

2.3 Small scale expression test 28

2.4 Large scale protein production and purification 29

2.5 Protein verification via SDS-Page gel electrophoresis 30

2.6 Protein crystallization 31

2.7 X-Ray data collection, reduction, and refinement 32

2.8 Gaussian Network Model (GNM) Analysis 34

2.9 Alphafold Analysis 35

Chapter 3: Results 36

Chapter 4: Dıscussıon 50

Chapter 5: Conclusion 53

Bibliography 54

LIST OF TABLES

Table 1.1: Comparison of protein engineering approaches……17

Table 1.2: Activity enhancement result of CbFDH mutants……...20

Table 1.3: Structures of CbFDH enzyme in PDB. Identity percentage and substrate are indicated with PDB codes……26

Table 2.1: His A, His B and Lysis buffer ingredient concentration as per liter….27

Table 2.2: Separating and stacking gel recipe for 15% SDS-Page Gel……..31

Table 2.3: The data collection statistic of 1.4 A resolution wt-CbFDH crystal….33

Table 2.4: The data refinement statistic of 1.4 A resolution wt-CbFDH crystal...34

LIST OF FIGURES

Figure 1.1: Formate major metabolic pathways and interorgan route…….2

Figure 1.2: Folate-mediated one-carbon pathway centered on formate……..4

Figure 1.3: Nicotinamide Adenine Dinucleotide redox reactions……...8

Figure 1.4: Catalytic mechanism of NAD+-dependent formate dehydrogenase…11

Figure 1.5: Seconder structure alignment of NAD+-dependent formate dehydrogenase and important residues for catalytic activity, formate, and NAD+ binding……13

Figure 1.6: Biotechnological of FDH-based enzymatic applications……..15

Figure 1.7: Circular dichroism result for melting temperature of wt- and His311Gln and Gln287Glu mutant-CbFDH enzyme……21

Figure 1.8: Relative activity of wt-CbFDH, His311Gln, and Gln287Glu mutant-CbFDH enzyme according to temperature…….22

Figure 1.9: Crystallization curve that represents the protein crystallization principles within supersaturated condition……..24

Figure 2.1: Protein crystallization procedure illustration……32

Figure 3.1: Small scale expression verification of wild-type and mutants CbFDH enzyme for 2 hours induction at 37°C……36

Figure 3.2: Large scale protein production and purification verification of wt-CbFDH at 37oC for 3h…….36

Figure 3.3: SDS-page gel electrophoresis for protein purification verifications of CbFDH-120* and CbFDH-285*……....37

Figure 3.4: CbFDH-311* and -287* mutants large scale protein purification verification by SDS-page gel electrophoresis……....37

Figure 3.5: Wt-CbFDH protein crystals with different conditions obtained by sitting-drop microbatch under oil method…….38

Figure 3.6: Wt-CbFDH protein crystals replicas gave results with Wizard III-47# condition that used for data collection……..38

Figure 3.7: Several crystals conditions produce protein crystals with mutant-CbFDH proteins……39

Figure 3.8: The diffraction pattern that is obtained while 1.0 A resolution data collection process……....40

Figure 3.9: 1.4 A apo-CbFDH structure and 5DNA apo-CbFDH structure comparison……..…41

Figure 3.10: Electron density map of new 1.4 A resolution apo-CbFDH structure with 1σ level 2Fo-Fc simulated annealing-omit map…….42

Figure 3.11: Alignment of 1.4 A resolution apo-CbFDH and 5DN9…….42

Figure 3.12: Morph analysis between apo- and holo- state of CbFDH reveal major conformational change…….…43

Figure 3.13: Cross-correlation maps of apo- and holo-CbFDH enzyme…....44

Figure 3.14: Cross-correlation differences apo-CbFDH over holo-CbFDH..45

Figure 3.15: Squared fluctuations of CbFDH for apo- and holo- forms……46

Figure 3.16: GNM mode variance represents that the first ten fastest modes are highest……46

Figure 3.17: Cross-correlation of intra- and inter-chain residues with NAD+, azide, and chloride molecules……....47

Figure 3.18: Alphafold2 prediction of CbFDH K47E mutant is compared with X-ray structure (PDB: 2FSS)…….48

ABBREVIATIONS

FDH	*Formate Dehydrogenase*
CbFDH	*Candida Boidinii* Formate Dehydrogenase
THF	Tetrahydrofolat
SAM	S-adenosylmethionine
CO_2	Carbon dioxide
NAD+	Nicotine amid Dinucleotide
NMR	Nuclear Magnetic Resonance
XRD	X-ray Diffraction
WC	Whole-cell lysate
SDS	Sodium dodecyl sulfate

Chapter 1:
INTRODUCTION

1.1 Formate

Formate is an important intermediate metabolite for both microorganisms and mammals. Formate, named after the Latin "Formica" meaning ant, because it was first isolated from ants, is the anion form of a formic acid (Pietzke, et al., 2020). Formic acid is characterized as a high-energy molecule due to its redox potential (-420 mV) (Sawers et al., 2004). With this potential, it is a source of energy in respiration in microorganisms and takes part in intermediate steps in one-carbon metabolism (Pinske and Sawers, 2016). (Gopalakrishnan et al., 2019). In general, formate is converted during energy production by various enzymatic pathways or removed by conversion to CO_2 and hydrogen (Gopalakrishnan et al., 2019). Formate, which is frequently encountered in biomass and CO_2 reduction studies due to its role in energy conversion, is synthesized and broken down by folate-dependent and folate-independent pathways as a co-factor in human metabolism. (Pinske and Sawers, 2016; Pietzke, et al., 2020). Since formate sources are serine, glycine, methionine, choline, formaldehyde, and methanol during synthesis, formate levels in mammalian body fluids can be evaluated as an indicator in cases such as vitamin B12 deficiency or methanol intoxication (Pietzke, et al., 2020).

1.1.1 Formate in Metabolism

Formate is the key intermediate of one-carbon metabolism and is present in blood in various mammals at a certain rate (10-100 mM) and has a half-life of 40-100 minutes, suggesting that formate is necessary for interorgan transport of one-carbon groups (Brosnan and Brosnan, 2016; Pietzke, et al., 2020). In addition, the cellular folate metabolism is closely related to formate and is regulated by mitochondrial and cytosolic pathways fed by various one-carbon donors.

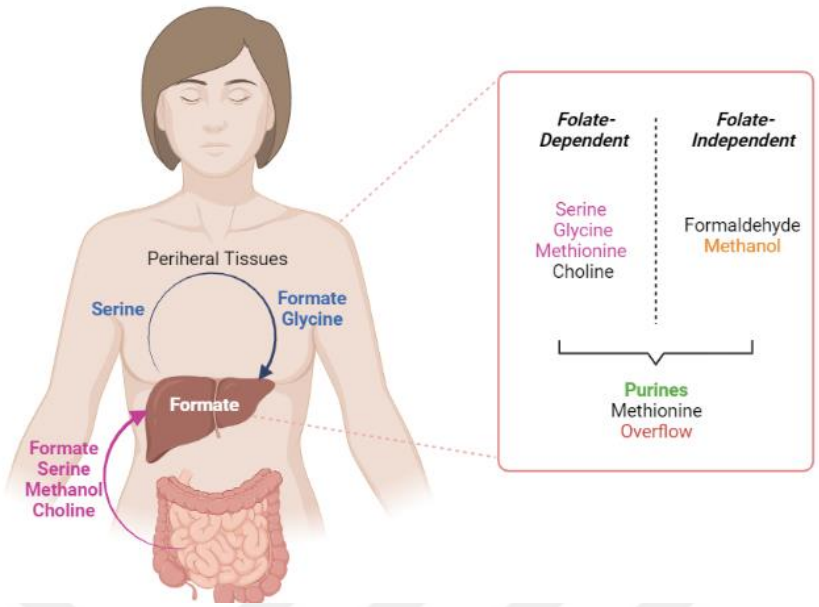

Figure 1.1 Formate major metabolic pathways and interorgan routes.

The production of folate-dependent formate is mainly produced in tetrahydrofolate (THF), one of its main sources being serine catabolism in the mitochondria. During the serine catabolism, glycine and 5,10-methylene are produced using THF and they are also used in the production of the one-carbon groups. Choline degradation, which also takes place in the mitochondria and uses THF as a cofactor, is also involved in the one-carbon synthesis. The glycine formed here is converted into ammonia and CO2 by the glycine cleavage system. While methionine is metabolized in the cytoplasm, it gives methyl groups to nucleophilic substrates through S-adenosylmethionine (SAM), which plays a key role in transmethylation, and helps the production of homocysteine, glycine, and other amino acids. Independent of folate, during the production of methionine, formate is produced during aminopropyl group transfer, which also takes place via SAM (Methionine salvage pathway). During the breakdown of histidine to formiminoglutamate, formate production takes place also in folate-dependent synthesis (Pietzke, et al., 2020).

In the production of folate-independent, one of the most important sources is methanol. Methanol is first degraded in the body by the catalase-peroxidase system and is converted to formaldehyde by alcohol dehydrogenase and then to formic acid by aldehyde dehydrogenase. It is thought that methanol in the body is formed by the fermentation of some fruits, such as pectin, in the gastro-intestinal tract (about half) or taken in low amounts from alcoholic beverages. In addition, it is thought that sweeteners

originating from aspartame are metabolized to methanol, and caffeine demethylation may be a source of formate due to formaldehyde production (Pietzke, et al., 2020). However, formate is used as an indicator of methanol toxicity since methanol poisoning directly increases formate production. In addition, the degradation of tryptophane and branched chain fatty acid by alpha-oxidation and the production of aromatase from lanosterol by demethylation of cholesterol and testosterone are also sources of folate-independent formate production. Because, during alpha-oxidation formyl-coA is broken down into formate and coenzyme A. (Lamarre et al., 2013; Brosnan and Brosnan, 2016; Pietzke, et al., 2020)

Formate metabolism is not limited to one-carbon and protein metabolism. Formate, which makes up 10-30% of the volate fatty acids in the blood of various mammals, is also vital for nucleic acid metabolism (Lamarre et al., 2013). Because formate is also required in the synthesis of purine nucleotide and thymidylate and the production of necessary metabolites (5,10-methylene-THF and 5-methyl-THF) during its remethylation from homocysteine to methionine. (Brosnan and Brosnan, 2016) Since formate is used as a precursor in the production of purine and deoxythymidine and adenine nucleotides, it is considered a growth-limiting factor directly related to nucleotide and energy metabolism. With an increase in formate, adenine nucleotide level and glycolysis rate increase, and AMPK activity decrease, in short, energy production increases. The increase in orotate and arginosuccinate levels also confirms the importance of formate in pyrimidine production (Oizel et al., 2020).

Formate elimination from the body is made by liver and urine disposal which are major and minor routes respectively (Lamarre et al., 2013). The formate route starts with serine catabolism in peripheral tissues and ends in liver and kidney by serine synthesis again (Pietzke, et al., 2020). Oxidation of formate to Co2 occurs through metabolic elimination with several mechanisms. One of these mechanisms is hydrogen peroxide (H2O2) included oxidation of formate by catalase enzyme. ($H_2O_2 + H^+ + HCOO^- \rightarrow 2H_2O + CO_2$) Another mechanism is the oxidation of formate by 10-formyltetrahydrofolate synthetase and 10-formyltetrahydrofolate dehydrogenase that is the folate-dependent mechanism. This mechanism catalyzes the reactions as:

Formate+ THF+ ATP→ 10-formyltetrahydrofolate+ADP+Pi,

10-formyltetrahydrofolate + NADP+ → THF + CO2 + NADPH) (Lamarre et al., 2013).

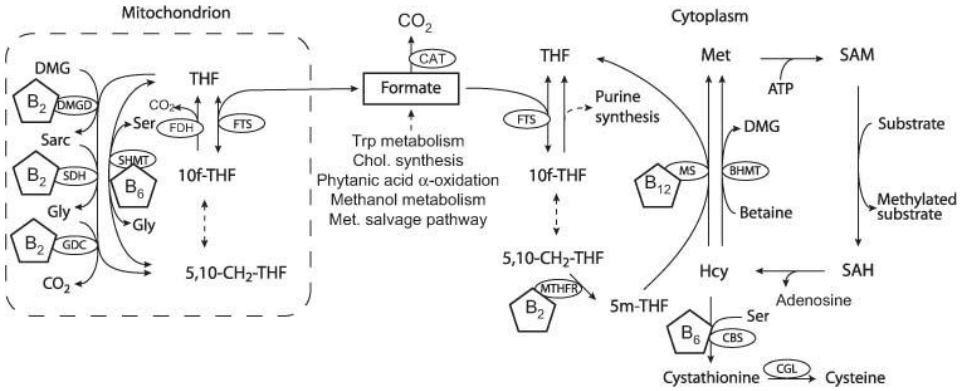

Figure 1.2 Folate-mediated one-carbon pathway centered on formate. (Lamarre et al., 2013)

1.1.2 Formate Level in Clinic

Vitamin B deficiency:

Since folate-dependent one-carbon metabolism of formate uses B-vitamin as a cofactor, the formate level in plasma and urinary excretion varies in B-vitamin deficiency, and it is thought that formate can be used as an indicator in different B-vitamin deficiency. Continuous synthesis of formate due to the impairment of one-carbon metabolism of folate-independent reactions is shown as the main reason for the increase in formate levels.

Vitamin B12, folate, B2 (riboflavin), and B6 deficiency commonly cause some imbalance in metabolism by causing accumulation in total homocysteine. However, in B12 deficiency, which indirectly causes folate deficiency through a mechanism called folate and methyl-folate, the plasma level of formate increases. Folate is also directly linked to formate production, as it is the active form of THF, the main intermediate in formate production (Pietzke et al., 2020). It is stated that vitamin B6 deficiency does not increase the amount of formate, since homocysteine accumulation is due to the homocysteine catabolism impairing caused by cystathionine-β-synthase (CBS). On the other hand, since B2 (riboflavin) is the precursor of FAD, which is a critical co-factor for folate-dependent one-carbon metabolism, its deficiency is thought to reduce the level of formate in the plasma, especially due to the decrease in the activity of the mitochondrial glycine cleavage system and methionine synthase (Lamarre et al., 2013).

Methanol toxicity:

In methanol toxicity, blindness (optic nerve injury) and metabolic acidosis occur, causing coma and even death. Also, during methanol toxicity, because of the central role of methanol in formate production, formate accumulates in the plasma and is one of the causes of acidosis (Lamarre et al., 2013). For this reason, various methods are being developed that can be used as a marker for methanol and ocular toxicity to determine the level of formate in the blood (Martin-Amat et al., 1978; Kerns et al., 2002). In one of the studies, the amount of formate in various body fluids was measured by converting the nonfluorescent dye resazurin to the fluorescent substance resorufin with the formate dehydrogenase enzyme produced from bacteria in combination with diaphorase-catalyzed reduction (Makar et al, 1975). In a similar study, the presence of formate with formate dehydrogenase and NAD+ were measured enzymatically by the ion-selective electrode analyzer (Cobas Mira), and the amount of formate was successfully detected in 14 of 15 patients who had methanol toxicity (Hovda et al., 2005). Alcohol dehydrogenase inhibitors such as ethanol and fomepizole are used in patients to prevent alcohol dehydrogenase from converting methanol to formate, which has a toxic effect during methanol toxicity (Pietzke, et al., 2020).

Cancer:

Cancer cells alter one-carbon metabolism, which is central to the production of amino acids, particularly serine and glycine, to provide amino acids necessary for high proliferation (Pietzke, et al., 2020). During this alteration, the amount of formate increases significantly, especially in folate metabolism. This increase has been demonstrated in both solid human tumors and various tumor cell lines. In addition, change in folate metabolism is not limited to mitochondrial formate production. At the same time, one-carbon metabolism is manipulated by cancer cells by changing the expressions of various enzymes. Since these enzymes are especially in THF production and glycine cleavage system, they increase the production of methyl groups required for amino acid and purine production and methylation (Brosnan and Brosnan, 2016). Increased levels of formate in cancer have also been demonstrated in serum samples in mouse models. This increase occurs with an increase in serine catabolism in mitochondria. Serine catabolism also increases tetrahydrofolate (THF) synthesis, which causes oxidative damage and toxic products, but the released THF is converted to the more stable 1-formyl-THF by formate.

(Oizel et al., 2020) The increase in formate is also associated with the production of formate by bacteria that cause chemoresistance in lung tumors (VanHook, 2022). For all these reasons, it has been recently stated that excess formate associated with cancer can be used in cancer treatment by stopping mitochondrial formate production through genetic and pharmaceutical trials (Pietzke, et al., 2020).

Other diseases:

Obesity causes an increase in purine synthesis. Since formate is a precursor in purine production, formate is consumed and its level decreases in the whole body. This decrease is evidenced by the fact that the circulating format level is lower in obese individuals than in healthy individuals. It is thought that the amount of formate may also increase in various neurodegenerative diseases. Knockdown of DJ-1 expression in DJ-1-dependent Parkinson's disease reduces the amount of formate by reducing mitochondrial formate-producing enzymes. Additionally, purine synthesis is known to affect brain functions. The decrease in the amount of formate, which is a purine precursor, will therefore also affect brain functions, and formate supplementation can be used therapeutically in diseases. Again, low plasma formate level is associated with cardiovascular disease, since it is thought that the decrease in purine synthesis is caused by mitral valve insufficiency and urate production (Pietzke, et al., 2020). Finally, due to the importance of formate in the production of various amino acids and purine, it is reported that its level is important in embryonic development (Brosnan and Brosnan, 2016).

1.1.3 Formate in Biotechnological Applications

Being a single-carbon organic compound, formate is an important mediator in both carbon metabolism and energy metabolism. But not just for organisms; It is also a good intermediate for biotechnological applications such as chemical production, carbon fixation/reduction, and biomass/biogas production. Because while providing CO2 sequestration, it also provides a cleaner and safer energy storage than hydrogen (Alissandratos et al., (2013); Schuchmann and Müller, 2013). Especially with synthetic biology applications, formate-fixing reactions are carried out with different enzymes and pathways in various species or conditions. The reduction by single electron pair transport from CO2 to the format can be achieved in various ways. These can be an enzymatic reduction, electroreduction, photoreduction, hydrogenation of CO2, biomass, and

oxidation of natural gas. In all these ways, formate is assimilated as an intermediate metabolite and takes part in physicochemical-biological processes in biotechnological applications such as microbial growth and chemical production. On the other hand, the use of the format in CO2 fixation/reduction pathways also makes an important contribution to the environment, especially by reducing greenhouse gas (Bar-Even, 2016).

During the enzymatic reduction of CO, the reverse reaction of the formate dehydrogenase enzyme is generally used. While low formate affinity is required for the reverse reaction for the enzyme, it can be used in different microorganisms by isolating it from the enzyme reaction or by altering its activity with different enzymes. The absence of limitations of enzymes such as high substrate purity or low concentration such as chemical reactions, and their high efficiency in different conditions such as anaerobic/aerobic or high temperature or methanogenic make the formate dehydrogenase enzyme an ideal candidate for formate assimilation (Alissandratos et al., 2013; Schuchmann and Müller, 2013; Schink et al., 2017).

Formate is also used in electrochemical energy-storage systems. Direct format fuel cells (DFFC), which have recently been the focus of attention, produce electricity from biochemical energy in the format. These cells, using palladium (Pd) as alkaline media, have a theoretical potential of 1.45V and can be easily used as biofuel by dissolving with water after being transported as a solid. For all these reasons, DFFC technology is an ideal system for clean energy production with its low cost and carbon-neutral fuel (An and Chen, 2016).

Formate serves not only in the production of fuel cells but also in biotechnology applications as an intermediate that is destroyed by biomethanation for the production of biofuels such as methanol, methane, and hydrogen. During biomethanation, there is the cooperation of fermentative, acetogenic, and methanogenic bacteria, during which formate is first broken down into CO2 and H2, and then CO2 is converted into methane. During the oxidation of formate, the FDH enzyme is frequently used. After formate is obtained by CO2 reduction of co-factor-dependent FDHs, formate is directed to methanol production via formaldehyde and alcohol dehydrogenase. (Crable et al., 2011).

1.2 NAD+ Coenzyme

Nicotinamide adenine dinucleotide (NAD+) is a metabolite that acts as an electron carrier in redox reactions and is a compound with phosphate groups and two nucleosides bonded to adenine and nicotinamide (Harden and Young, 1906; Smyth et al., 2004; Ziegler and Niere, 2004). While adenine and ribose ring with nicotinamide are attached to the nucleosides from the first carbon atom, two phosphate groups are attached from the fifth carbon atom (Pollak et al., 2007).

NAD+ functions as a coenzyme of DNA ligase, sirtuin, and ADP-ribosylation enzymes. While it is used as an electron carrier and in protein modifications, it acts as a substrate in enzymes and is reduced to form NADH (Figure 1.3). (Smyth et al., 2004; Billington et al., 2006). In redox reactions, while the proton is removed from the other substrate, the proton in the solution is oxidized and NADH is released by transferring it to the nicotinamide ring of NAD+. In the hydride electron pair transfer, one of the electrons is transferred to the positively charged nitrogen atom of the nicotinamide ring, and the other to the C4 carbon atom of the same ring. However, since this reaction is easily reversible, NAD+ is a good intracellular reducing-oxidizing intermediate metabolite (Unden and Bongaerts, 1997).

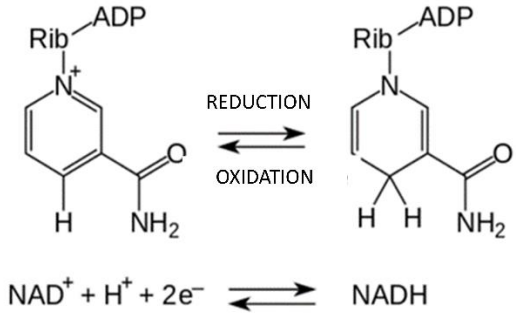

Figure 1.3: Nicotinamide Adenine Dinucleotide redox reactions (Pollak et al., 2007).

NAD+ not only functions as an oxidizer in intracellular metabolic events but also plays a role as a signal molecule by being secreted between tissues from organs such as neurons, large intestine, brain, and urinary bladder (Unden and Bongaerts, 1997; Koch-Nolte et al., 2011; Breen et al., 2006; Mutafova-Yambolieva et al., 2007; Durnin et al., 2012). Finally, the difference in UV-absorption peaks is utilized for the spectroscopic detection of NAD+ and NADH and enzyme kinetic studies (295 nm / 339 nm respectively) (Dawson, 1985; Lakowicz et al., 1992; Jameson et al., 1989).

1.2.1 NAD+ Biosynthesis and Oxidoreductase Reactions

There are two main pathways involved in NAD+ biosynthesis. While new NAD+ is produced from different amino acids in de novo pathway, NAD+ is recycled from preformed compounds such as nicotinic acid, nicotinamide, and nicotinamide riboside in the salvage pathway (Lin and Guarente, 2003).

In the de novo pathway, NAD+ is produced by using tryptophan or aspartic acid, known as quinolinic acid, as a precursor (Katoh et al., 2006; Foster and Moat, 1980). Nicotinic acid mononucleotide is formed by adding phosphoribosyl to these amino acids, and then an adenylate group is added to produce nicotinic acid adenine dinucleotide. Finally, nicotinamide adenine dinucleotide is formed by amination (Belenky et al., 2007). In the salvage pathway, simple amino acids and pyrimidine contents are recycled to obtain NAD+. Vitamins called niacin are used as the main precursors (Lin and Guarente, 2003; Anderson et al., 2002). NAD+ level is usually stable in metabolism after its use in reduction and oxidation reactions by enzymes (Lin and Guarente, 2003). However, it is important to maintain the current balance during synthesis and catabolism. For this reason, the intake of especially essential amino acids and vitamins is important (Henderson, 1983).

NAD+ is used as a coenzyme in metabolism by the oxidoreductase enzyme group, which includes dehydrogenase and reductase enzymes. Here, the task of NAD+ is electron transfer in reducing reactions, as mentioned before. The NAD+/NADH binding enzyme family is very large and has a Rossman fold in common (Hanukoglu, 2015; Rao and Rossmann, 1973). In NADP+ binding, which is very similar in structure but very different in metabolic functions, the active site of enzymes has different amino acids, and they also show different and high specificity (Carugo and Argos, 1997). While there is a basic amino acid that will create an ionic interaction with the acidic phosphate group in NADP+ binding, specificity is provided for NAD+ with a more acidic binding pocket. While the main task of enzymes is to make the reaction efficient by removing water from the reaction medium in their active sites, on the other hand, it is to bring the coenzyme and the substrate to the appropriate position for the reaction. As mentioned earlier, since electron transfer occurs over the C4 of nicotinamide, this ring is positioned by oxidoreductase enzymes to perform electron transfer to the other substrate. However, this

positioning may differ as a planar conformation according to the enzymes (Rao and Rossmann, 1973).

1.3 Formate Dehydrogenase

Formate dehydrogenase (FDH) enzyme is an oxidoreductase enzyme found in both prokaryotes and eukaryotes, operating in both aerobic and anaerobic conditions, and its size can vary between 79 and 590 kDa. The main function of the enzyme is to oxidize the formate and convert it to CO2, and for this reaction, natural electron carriers such as NAD+, NADP+, ferredoxin, or cofactors such as tungsten, molybdenum, flavin are used (Ljungdahl, 1980; Fauque and Barton, 2012). In some anaerobic bacteria, this reaction reversibly takes place, performing CO2 fixation/reduction. In these anaerobic species, the activity of the FDH enzyme is generally dependent on the presence of cofactors such as tungsten, molybdenum, and heme. (Fauque and Barton, 2012) This class of FDH is called metal-dependent formate dehydrogenase, and hydride transfer is accomplished directly with the oxidized metal-contained active site. Formate reducing is achieved here by the exchange of electrons between metal cofactors and the substrate. Although this prokaryotic protein can be produced in high amounts, especially with recombinant protein technology, strictly anaerobic conditions are needed in studies because it is a metal-containing protein. Another class of formate dehydrogenase is NAD+-dependent formate dehydrogenase and NAD+, an organic oxidizing molecule, acts as a coenzyme for formate reduction (Niks and Hille, 2018).

Various structural-functional studies have been carried out so far to elucidate the working mechanism of different types of enzymes, especially due to CO2 reduction and its economic potential (Thomas et al., 2016). The inhibitors of the FDH enzyme are azide and cyanide, which are covalent inhibitors, but they are not preferred in studies due to their high toxicity (Ljungdahl, 1980).

1.3.1 NAD+-dependent Formate Dehydrogenase

NAD+-dependent formate dehydrogenase is a fundamental enzyme for metabolic regulation of one-carbon metabolism for organisms, especially for yeast and aerobic bacteria. NAD+-dependent FDH is classified within-specific 2-hydroxy acid dehydrogenase family (Nielsen et al., 2019). Conversion of formate to CO2 and NADH

by coenzyme NAD+ is performed within the redox-active catalytic site of the enzyme which mostly does not include cofactors (Torres et al., 1999). NAD+-dependent FDH is used easily in aerobic conditions, does not require costly cofactors, and due to its high efficiency, it is especially used for the pure production of invaluable chemicals and CO2 fixation, as well as for the determination of formate from body fluids biomedically. Because of this value, this enzyme has been widely studied both in enzymatic and structural studies and in improvement studies with mutations.

1.3.1.1 Molecular Mechanism of NAD+-dependent FDH

The redox-active site of the enzyme has mostly 8 conserved amino acids such as Arg, His as basic amino acids to create a binding site for formic acid or CO2; Asp, Thr, Gln, Ser as polar amino acids for NAD+/NADH binding and Ile as a hydrophobic amino acid. The polar Thr, Asp, and His create NAD+ binding site with a hydrogen bond network that traps NAD+ from the amide group's two hydrogens and carbonyl oxygen. NAD+ trap is believed as the first step of the reaction mechanism of NAD+-dependent FDH. Then, formic acid binds to positively charged amino acid Arg and His from its oxygen, and formic acid oxidation occurs during NAD+ reduction (Amao, 2018). The site-directed mutagenesis studies by Popov et al. and structural studies also proved the importance of the His332-Gln313 pair and Arg284 for substrate binding and catalytic activity, respectively. The reaction is the result of the electrostatic effects of those residues and the reversible process depends on substrate concentration because of the potential affinity of both substrate and product in NAD+-dependent FDH (Castillo et al., 2008; Sato and Amao, 2020).

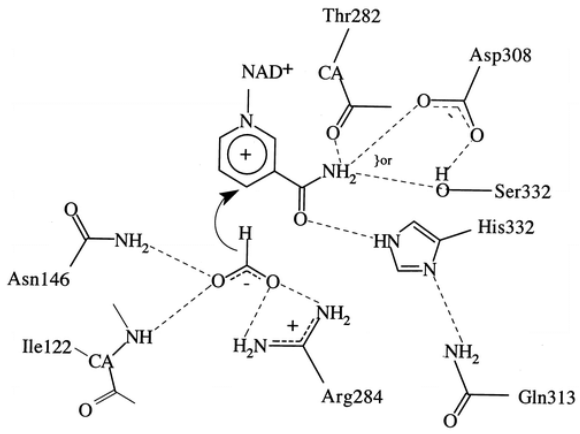

Figure 1.4: Catalytic mechanism of NAD+-dependent formate dehydrogenase. NAD+ and formate binding is provided by conserved residues that bring this molecule in an approximate position for hydride transfer (Torres et al., 1999)

In molecular dynamics (MD) simulations, it has been reported that these preserved residues (Arg284, Asn146, Ile122) provide appropriate conformation for hydride transfer by bringing the formate closer to the C4N group of the nicotinamide ring of NAD+ with electrostatic interactions. At the same time, it has been proven that Thr, Asp, His, and Ser residues and NAD+ maintain this conformation by binding to the active site with a trans conformation of NAD+. (Figure 1.4) (Torres et al., 1999).

In general metal-independent FDHs such as NAD+-dependent FDH have specificity for NAD+/NADH binding and do not constitutively permit binding of NADP. In addition, NAD+-dependent enzymes favor formate oxidation more than CO2 reduction. The enzyme favors hydride transfer to NADH from formate with the help of creating an orientation of NAD+ and formate within the transition state (Nielsen et al., 2019). Quantum mechanics/molecular dynamics (QM/MM) studies have confirmed that the enzyme positions the substrate and coenzyme close to the transition state, thus providing the appropriate conformation for the hydride transfer of the enzyme and favoring the enzyme in the forward direction. In the same study, it was observed that the presence of water in the active site disrupted this transition state and made the reaction difficult. In this way, it proves the necessity of the presence of hydrophobic amino acids like Ile in the active site and one of the functions of the enzyme is to lower the energy barrier by removing water (Castillo et al., 2008).

1.4 Candida boidinii NAD+-dependent Formate Dehydrogenase

Candida boidinii is an industrial-important yeast strain that is mostly used for chemical synthesis such as methanol or biofuel and recombinant protein production. *Candida boidinii* NAD+-dependent enzyme (CbFDH) is one of those invaluable enzymes that are suitable for many industrial applications. CbFDH is a homodimeric enzyme that has 84 kDa total molecular weight (Castillo et al., 2008). Like other NAD+-dependent FDHs, CbFDH has not also had metal or cofactor requirements of catalytic center. The reaction mechanism is also conserved as oxidation of formate by the reduction of NAD+. According to the study of Labrou and Rigden, Val-123 and Ile-175 provide a suitable anhydrous environment for the reaction by forming hydrophobic pockets, as was the case with NAD+-dependent FDHs before. At the same time, Arg-258 and Asn-119 are important for hydride transfer while providing formate binding. While Arg-258 stabilizes

the negative charges of the formate's oxygens, Asn-119 ensures that the formate stays in proper conformation. In addition, His311-Gln287 residues (especially the unprotonated state of His311), which replace the conserved His-Glu relay system in other NAD+-dependent FDHs, are thought to contribute to the binding of formate with their electrostatic properties and are important for the enzymatic reaction (Labrou and Rigden, 2001).

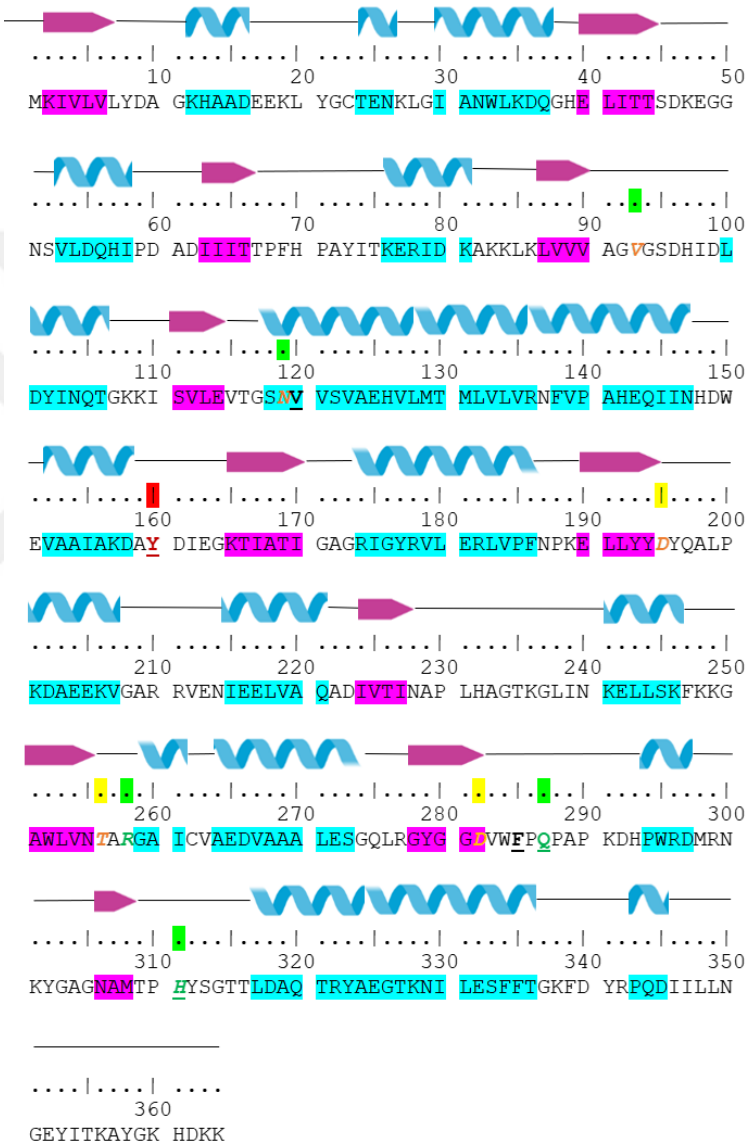

Figure 1.5: Seconder structure alignment of NAD+-dependent formate dehydrogenase and important residues for catalytic activity, formate, and NAD+ binding. Alpha-helices and beta-sheets are indicated with blue string and pink arrows, respectively. Formate-binding residues are shown by orange letters and green boxes. Catalytic activity residues are represented by green letters and boxes. NAD+ binding residues are indicated in orange letters and yellow boxes.

Like other NAD+-dependent FDHs, CbFDH generally does not favor forwards. However, in a study, the reverse reaction of CbFDH from CO2 to formate production was tested and it was found that the amount of formate produced with increasing CO2 concentration was directly proportional. The fact that the amino acids (Asn119, Ile175, and Arg258) to which formate binds are also residues to which CO2 is attached, make the reverse version of the reaction possible (Sato and Amao, 2020).

Not only application in pharmaceutical products but also the carbon reduction potential of the CbFDH with less toxic and more stable environmentally-friendly application makes this enzyme more valuable if the enzymatic turnover rate becomes higher relative to chemical applications (Nielsen et al., 2019). Because of the industrial and environmental importance of CbFDH, site-directed mutagenesis for the enzymatic activity or thermal stability enhancement and understanding of the molecular mechanism of the enzyme is fundamental (Castillo et al, 2008).

1.5 Practical usage of FDH

NAD+-dependent formate dehydrogenase can be used in various metabolic engineering applications to change the cofactor availability, which is one of the rate-limiting factors, and to increase system efficiency. In this way, not only the rate-limiting steps of the reactions are supported, but it also supports fermentation as an oxidizing agent, as NADH is required for anaerobic cell growth (Berríos -Rivera et al., 2002).

For example, in the study of Balzer et.al, Candida boidinii originated NAD+-dependent FDH was added during fermentation to prevent formate formation, which is the main problem in the production of succinic acid, as industrially important. As a result of this application, the formate by-product formed during succinate production was eliminated and NADH availability was increased (Balzer et al., 2013). In addition, the fact that overexpressed CbFDH in E. coli increases NADH and increases cell density as expected, again proving the effect of NADH availability on cell growth (Berríos -Rivera et al., 2002). Again, it has been observed that FDH overexpressing S. cerevisiae strains obtained by adaptive laboratory evolution have increased tolerance to formate and acetate, which are important for bioethanol production. In this study, FDH not only removed the excessive amount of formate from the environment by converting it to CO2 as is known

but also increased cell viability by converting it to acetate formate and ATP as a result of CO2 reduction in the absence of formate (Du et al., 2022).

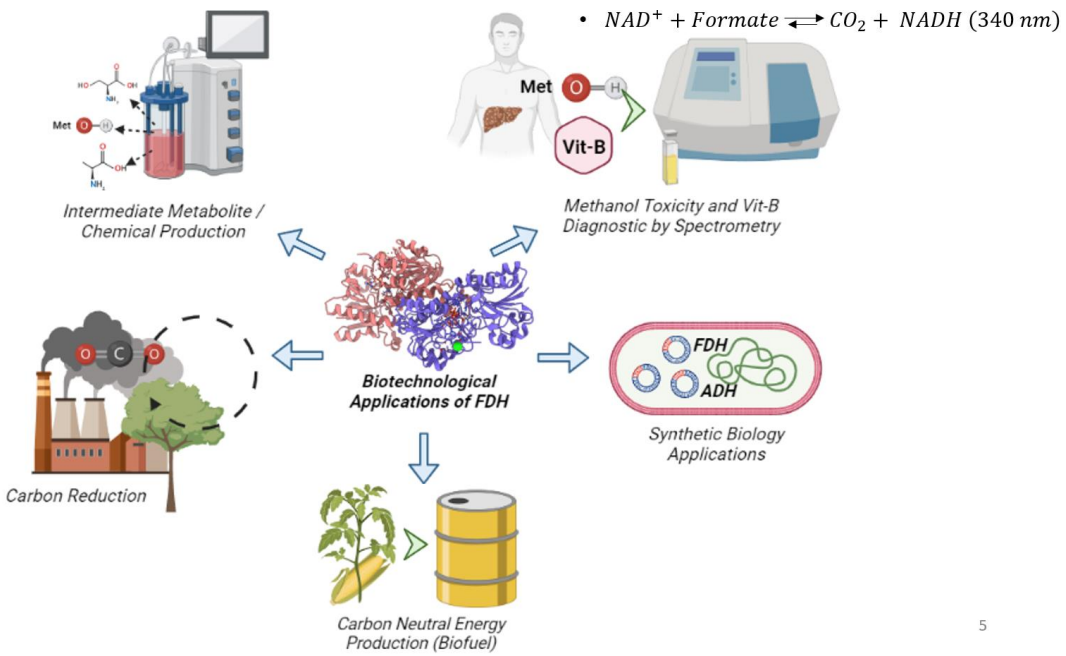

Figure 1.6: Biotechnological of FDH-based enzymatic applications

Apart from carbon neutral energy production such as biomass production, the use of FDH is also environmentally friendly with CO2 reduction. In a recent study, they operated the reverse reaction mechanism of CbFDH for CO2 reduction using an artificial cofactor, methyl viologen radical cation (Jayathilake et al., 2019). Another cathodic cell application similar to this study is the use of NAD+-dependent FDH in electrochemical NADH regeneration for CO2 reduction. In the study, NAD+-dependent FDH immobilized to electrospun polystyrene nanofibers increased both long-term stability and reusability (Barin et al., 2018).

For the production of amino acids such as methanol, succinic, L-alanine, and L-serine with in vivo systems, formate is also produced from CO2 by using FDH as a liquid one-carbon intermediate metabolite. It is thought that the formation of formate assimilation pathways by FDHs overexpression in different systems such as E. coli will be seen frequently in the bioproduction of different valuable chemicals in the future (Calzadiaz-Ramirez and Meyer, 2022).

As stated in the previous section, CbFDH is also a preferred enzyme during CO2 reduction by binding both formate and CO2 to critical amino acid residues. However, one of the new FDH strains tried in a more recent study, Thiobacillus sp. (TsFDH) showed

84.2-fold higher CO2-reducing activity, showing that there are other promising enzyme strains in this area. In the same study, it was also shown that low pH significantly increased the activity for CO2 reduction (Choe et al., 2014).

In terms of synthetic biology applications, the use of carbonic anhydrase and FDH for carbon reduction by using microbial transglutaminase as a cross-linking medium can be shown as a good example. This hybrid enzyme both strengthened in terms of thermal stability and preserved enzymatic activity. Overall catalytic efficiency increased 5.8 times compared to free FDH for formate production (Zhang et al, 2021).

Finally, formate change in body fluids in various physiological conditions in mammals has paved the way for the use of formate dehydrogenase enzyme in clinical applications. Various studies are carried out to measure NADH, which is formed as a result of the reaction of the FDH enzyme, especially in the early diagnosis of vitamin B deficiency and methanol toxicity, by spectrophotometric methods (Pietzke et al., 2020; Lamarre et al., 2013; Martin-Amat et al., 1978).

1.6 Protein Engineering

In recent years, as proteins provide more environmentally friendly processes, the expectations that these biological macromolecules should meet significantly changed with the increase in applications in different pharmaceutical and industrial fields. For example, while industrial applications require enzymes that can withstand extreme conditions such as high temperatures and different pH, enzymes with high substrate and cofactor specificity are also being developed for increased efficiency. In pharmaceutical applications, protein engineering applications are used to meet important parameters such as more effective drug-target molecule interaction and binding, fewer side effects, and longer half-life (Singh, et al., 2018). Protein engineering is to provide the desired properties such as high yield and durability, as in the examples, by causing changes in the peptide sequences of proteins with DNA manipulations (Bansal and Kundu, 2022).

There are three main types of protein engineering applications: rational design, directed evolution, and semi-rational design. Rational design requires knowledge of the protein's three-dimensional structure, sequence, and function. This structure-function information is used to directly provide the desired properties of the protein by changing amino acids. Although the variants obtained as a result of the design provide cost-

effectiveness, the complex structure of the proteins makes these applications difficult as they cannot predict all the allostery-folding changes up to the product (Pitela et al., 2007; Chen, 2001; Bansal and Kundu, 2022). Directed evolution applications, on the other hand, include efficiency testing with a high-throughput screen for structure and function after random mutagenesis and do not require prior knowledge. Although this method seems efficient in terms of diversity, it is not cost-effective (Williams et al., 2004; Singh, et al., 2018). A semi-rational design has been developed to overcome the limitations of these two methods. With the semi-rational design, a smaller pool is created by random mutagenesis in a certain sequence range using the three-dimensional structure of the protein, in silico predictions, and certain functional features and this pool is scanned to provide the desired feature more efficiently (Singh, et al., 2018). (Bansal and Kundu, 2022). For this method, de novo protein design is also used, which provides the computational design with protein folding algorithms, and provides in silico prediction of proteins with unknown structures through steps such as sequence optimization and validation starting from backbone building. (Gulati and Poluri, 2016; Singh, et al., 2018).

Table 1.1: Comparison of protein engineering approaches.

	Rational Design	**Directed Evolution**	**Semi-Rational Desing**
3-D Structure	Required	Not required	Required
Mechanism	Required	Not required	Required
Screeing Methods	Not required	Required	Required

1.6.1 Rational Design

Rational design is a method of protein engineering in which amino acid sequences are expressed using the protein's three-dimensional structure and function relationship (Singh et al., 2013). Site-directed mutagenesis is the most widely studied application. In this method, which is used today with optimized commercial reaction kits, residual manipulations related to function/stability on the sequence can be designed with the structure of the protein or homology models if it is not present. Amino acid residues can be inserted, deleted, or substituted made by this method, and the designed mRNAs create the difference between the method (Cedrone et al., 2000). With this method, the cofactor/substrate affinity and kinetic performance of various enzymes can be increased-decreased (Getzoff et al., 1992). Or target protein-drug binding and function changes can

be examined (Que´me´ neur et al., 1998). In this way, the direct structure-function relationship can be examined retrospectively. Finally, purification and testing of each mutant are less cost-effective compared to high-throughput methods (Chen, 2001; Bansal and Kundu, 2022).

1.6.2 Directed Evolution

Directed evolution, as the name suggests, is the modification of a protein's properties such as affinity, stability, and specificity by creating a large DNA library as evolutionary. In this method, it is necessary to have a known and tested enzymatic/binding method for high-throughput monitoring of genotype or phenotypic change. In addition, the change must be phenotypically observable and available for screening (Zeymer and Hilvert, 2018; Yuan et al., 2005; Farinas et al., 2001).

In this method, in which three-dimensional structure information is not required, diversification is achieved by creating a DNA library with random mutagenesis, and then the desired variants are selected by screening and selection processes. Finally, the screening cycles are repeated to optimize the selected variants (Singh et al., 2013). In the directed evolution method, gene diversification can be performed non-recombinatively with random mutagenesis or recombinative by altering gene blocks with homologous/non-homologous methods (Pitela et al., 2007; Yuan et al., 2005). The screening part can be performed in vitro with automated systems or on a suitable host such as bacteria, or viruses (Zeymer and Hilvert, 2018).

In non-recombinative design, random mutations are created by point mutations such as insertion, deletion, and substitution, as in directed evolution, and can be generated by damaging DNA with mutagens or UV, or by error-prone PCR (Yuan et al., 2005). Error-prone PCR uses error-generating primers and the probability of error in DNA replication (Wong et al., 2004). In the recombinative design, on the other hand, imitates the homologous or non-homologous gene changes seen in sexual reproduction and thus creates diversity (Stemmer, 1994). In this method, like the DNA shuffling method, different genes are fragmented and then combined with PCR with the help of designed primers (Crameri et al., 1998).

1.6.3 Semi-Rational Design

Semi-rational design, the inability of directed evolution to provide effectiveness due to randomness, and the low cost-effectiveness of rational design, combined these two methods to remove these restrictions (Leisola and Turunen, 2007; Bansal and Kundu, 2022). Using the known structure and in silico predictions in the method, functionally important residues are detected and more specific libraries are created by site-saturation mutagenesis (Yoshikuni et al., 2006). After selecting efficient variants by screening as in rational design, desired properties can be developed further with site-directed mutagenesis on them (Singh et al., 2013). The effectiveness of this method, which is performed by PCR with primer sets containing codon variants for selected residues, has been proven by various studies (Parikh and Matsumura, 2005; Spadiut et al., 2008).

1.7 Candida Boidinii NAD+-dependent FDH Protein Engineering Studies

Although the first protein engineering studies for FDH were for metal-dependent FDHs, the most frequently studied group was metal-independent FDHs. Complex metal-dependent FDHs containing heterooligomers and multiple cofactors are difficult to study with methods such as cryo-EM. In addition, the reason for this is that they do not have high homology such as metal-independent FDHs. Protein engineering studies for FDH, like other enzymes with industrial potential, have focused on increasing the catalytic activity and substrate affinity and enhancing thermal stability (Galkin et al., 2002; Calzadiaz-Ramirez and Meyer, 2022). In addition, CO2 reduction enhancement studies, such as other metal-independent FDH, such as *C. thermophilum*, are also carried out due to its potential in carbon-reduction pathways, which have been intensified in recent years (Calzadiaz-Ramirez and Meyer, 2022). *Candida boidinii* NAD+-dependent FDH is the focal point for industrial FDH improvements because it has the highest thermal stability and catalytic activity among other NAD+-dependent FDHs (Bulut et al., 2020).

1.7.1 Enzymatic activity enhancement of Candida Boidinii NAD+-dependent FDH

The results of the CbFDH activity mutation studies also shed light on the importance of residues in the substrate/cofactor binding or activity of the enzyme. For example, in the studies of Labrou et.al, Arg258Ala silencing mutation caused loss of enzyme activity. This finding provides evidence for the necessity of Arg258 both in binding to formate

and in its transfer to hydride. Similarly, His311Gln mutation, which is known to be important for substrate binding, significantly reduced formate binding (10-fold), but NAD affinity did not change. Asn119His and 119-311 combined mutants, which are also known for their importance in binding the NAD+ co-enzyme, significantly reduced the affinity of both NAD+ and formate (between 70-200-fold) (Labrou and Rigden, 2001). Ile175Ala and Phe69Ala mutations, which are known to be crucial in hydrophobic active site formation, were observed to decrease format and NAD+ affinity similarly. Mutation studies have shown that Pro-68 and Phe-97 residues, which are important in the formation of the hydrophobic wall, are also important for the binding of formate (Labrou and Rigden, 2001).

In different CbFDH activity mutation studies, Cys23Ser and Phe285Ser mutations increased the activity (Felber, 2001), while Val120Ser and Asn187Asp mutations decreased it (Jiang et al., 2016). The basic principle here is to convert non-polar residues to polar or charged residues, while the effect of active site positions on electrostatic interactions is due to this difference. It has been observed that His311Phe mutation prevents formate binding and activity in mutations made for His311 and Gln287, which are thought to contribute to the activity by forming a proton-relay system, while His311Gln and Gln287Glu mutations reduce the activity (Labrou and Rigden, 2001).

Table 1.2: Activity enhancement result of CbFDH mutants: Val120Thr, Phe285Thr, Gl287Glu, and His311Gln.

	Formate			NAD^+		
	K_M(mM)	k_{cat}(s-1)	k_{cat}/K_M($M^{-1}s^{-1}$)	K_M(mM)	k_{cat}(s-1)	k_{cat}/K_M ($M^{-1}s^{-1}$)
Wt-CbFDH	2.00±0.12	9.48±0.19 x10^2	4.74 x10^2	1.50±0.34	2.62±0.47 x10^4	1.74 x 10^4
Val120Thr-CbFDH	1.30±0.28	6.23±0.17 x10^3	4.79 x10^3	7.90±0.23	1.04±0.52 x10^4	1.32 x10^3
Phe285Thr-CbFDH	0.97±0.23	6.23±0.38 x10^3	6.40 x10^3	2.30±0.41	4.59±0.19 x10^3	1.99 x10^3
Gln287Glu-CbFDH	1.65±0.31	4.42±0.26 x10^3	2.67 x10^3	3.30±0.18	9.51±0.43 x10^3	2.90 x10^3
His311Gln-CbFDH	1.50±0.16	3.28±0.22 x10^3	2.18 x10^3	1.60±0.34	8.20±0.19 x10^3	4.93 x10^3

On the other hand, our preliminary data on the His311Gln and Gln287Glu provided a better turnover rate and higher affinity for formate while the result was similar to Labrou's study for NAD+ co-enzyme. Finally, novel mutants of the CbFDH enzyme in our preliminary study had a promising result for activity enhancement in the biotechnological application. Val120Thr and Phe285Thr mutants had higher enzymatic turnover rates (almost double) and higher formate binding affinity which can be seen by lower K_m value and k_{cat} value respectively.

1.7.2 Thermal stability enhancement of Candida Boidinii NAD+-dependent FDH

Thermal stability studies for NAD+-dependent FDH have generally been done by manipulation of the cysteine residue. For example, mutations Cys255Met, Cys255Ser, and Cys255Ala for PseFDH increased NAD+ affinity while decreasing thermal stability (Tishkov et al., 1993). For CbFDH, the Cys23Ser and Cys262Ala mutations similarly decreased the thermal stability but did not cause a change in the activity, but the stability of these mutants against metal-induced oxidative stress increased. The reason for this is that the hydrophilic cysteine residue makes the enzyme more water exposed and vulnerable to metal effects (Slusarczyk et al., 2000). Finally, it has been observed that thermal stabilization is increased by making the helix structures more hydrophobic with Ser131Ala, Ser160Ala, Ser184Ala, and Ser228Ala mutations in PseFDH, which is also a homolog of CbFDH, and a similar strategy can be applied for CbFDH (Rojkova et al., 1999).

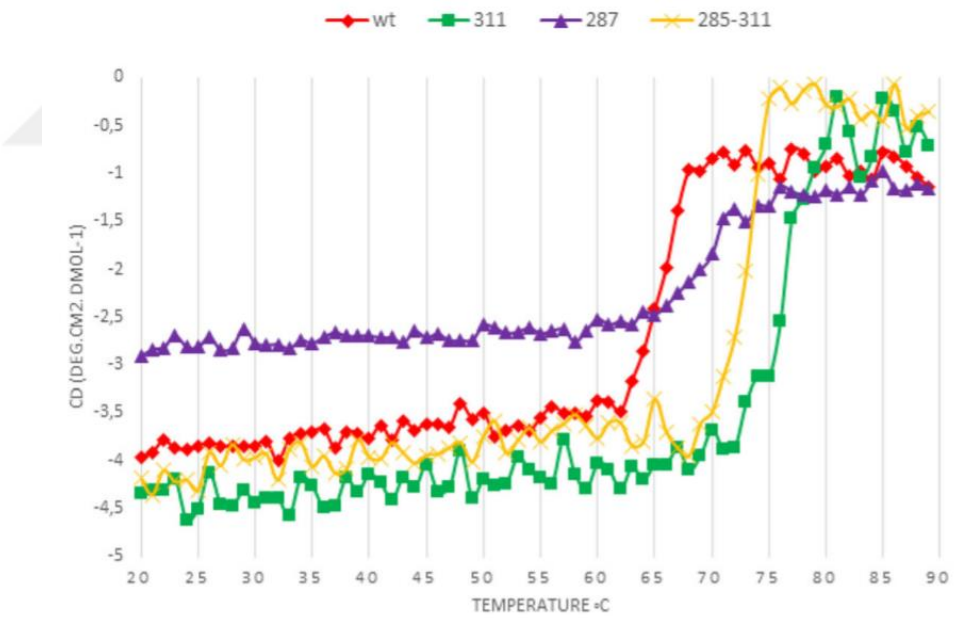

Figure 1.7: Circular dichroism result for melting temperature of wt- and His311Gln and Gln287Glu mutant-CbFDH enzyme. The baseline of each curve represents the folding state while the topline indicates the unfolding state of the protein. The melting point of each curve is the middle point of the slope between folding and unfolding states (Bulut et al., 2020).

Moreover, our recent study revealed that His311Gln and Gln287Glu mutations which are on conserved active site residues observed a significant increase in thermal stability.

According to the circular dichroism result, the His311Gln and Gln287Glu mutations increased melting temperature points from wt-CbFDH as 65℃ to 78℃ and 73℃ respectively (Figure 1.7). On the other hand, the optimum temperature of enzymatic activity was detected with a 25-30℃ increase by those point mutations, which may provide an invaluable opportunity for the usage of CbFDH biotechnological application in extreme conditions (Figure 1.8).

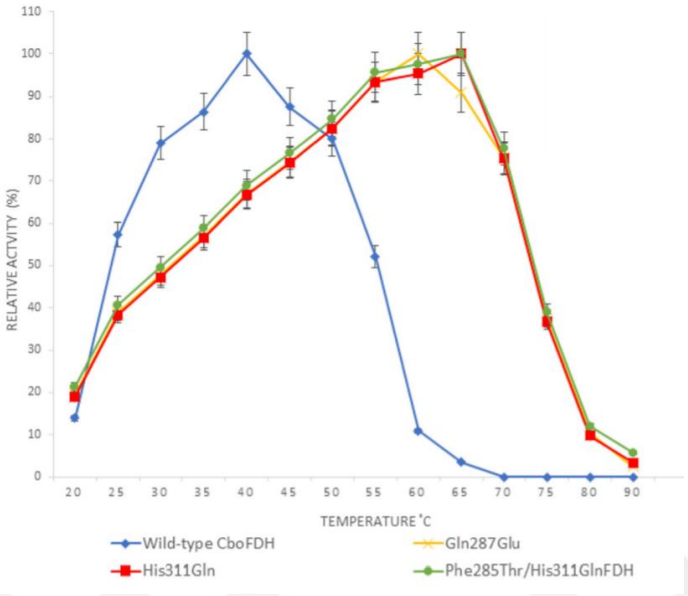

Figure 1.8: Relative activity of wt-CbFDH, His311Gln, and Gln287Glu mutant-CbFDH enzyme according to temperature. The peak points of each curve represent the optimum temperature for enzymatic activity (Bulut et al., 2020).

1.8 Structural Biology

Although X-ray technology is much older and the first protein crystals were grown in the 1890s, the relationship of X-ray crystallography technique with biological sciences dates back to the 1950s. Thanks to the structural biology techniques that started with X-ray crystallography, it has become possible to reach the structure of proteins, DNA, and RNAs at the molecular level and thus to understand the molecular mechanisms of life (Nitta et al., 2018). Structural biology techniques used today are X-ray crystallography, neutron crystallography, electron microscopy, nuclear magnetic resonance (NMR), and solution scattering (Curry, 2015). Each of these techniques has its advantages and limitations. For example, the first NMR and X-ray crystallography methods are the preferred methods to study molecules with high resolution and detail. Due to the

examination of the samples in the solution in the NMR method, the conformational differences of various environments (such as pH, temperature), binding, or drug interactions can be observed in motion. However, since the operable size range is a maximum of 25 kDa, this allows only small proteins to be studied. On the other hand, although the workable protein size range is wider in the X-ray crystallization technique, crystallization of mobile or large proteins is very difficult. Cryo-electron microscopy (Cryo-EM) technique and X-ray crystallography-like data collection and reduction process created by Henderson et al. have been very effective in eliminating the limitations of these two methods (Nitta et al., 2018). Today, high-resolution structures of proteins of 50 kDa and above can be solved by the cryo-EM technique. With the newly created techniques, it has become possible to examine smaller proteins or single-particle molecules by the cryo-EM method. In addition, it has been possible to examine various membranes or intrinsically disordered proteins in NMR spectroscopy with the understanding of solid-state NMR (Dobson, 2019). Neutron crystallography, on the other hand, is a method that can be used as a complementary method to X-ray in finding the positions of H atoms by the technology providing diffraction of neutrons from the cell nucleus (Curry, 2015).

The existence of all these techniques, however, did not allow to reach the structure of some challenging proteins. For this reason, prediction tools such as AlphaFold (DeepMind2) and RoseTTAFold (University of Washington) have come to the fore in recent years. Machine learning algorithms on protein folding applied in these methods can give high predictions about at least the secondary structure of proteins. Although these predictions are weak in data such as molecular interaction and side chain conformation, they are promising for the future (Subramaniam and Kleywegt, 2022).

1.8.1 X-Ray Crystallography

X-ray crystallography is the first and most frequently used method to extract the structures of biological molecules at high resolution and atomic levels. Thanks to the detailed structures of the method and the opportunity to examine low-weight molecules, interaction information and conformational changes caused by ligand binding can be examined. Therefore, X-ray crystallography is a unique method for the study of various pharmaceutical and target proteins (Maveyraud and Mourey, 2020). However, this

method is not limited to drug studies, it is also very valuable for understanding the bindings such as enzyme-substrate/cofactor with similar advantages and examining the conformational changes created by small mutations.

The first step of the X-ray crystallography technique is crystallization, which is the most important part. In this technique, the aim is to make the protein thermodynamically stable, that is, energy minimized. For this, the protein is expected to crystallize in a supersaturated (precipitant) solution, staying between the solid and solution phase. Precipitant solutions usually contain organic solvents, PEG, or salts in different concentrations and combinations (Ilari and Savino, 2008). Ammonium sulfate is the most commonly known salt and is harmless to proteins. Salts aid crystallization by reducing surface charges on proteins and stabilizing waters. Similarly, PEG immobilizes water, but on the other hand, PEG molecules help energy minimization by wrapping around the protein. Organic solvents such as MPD, on the other hand, lower the dielectric constant, allowing proteins to converge and crystallize. Briefly, the task of precipitants is to regulate surface charge disturbance and solvent access for protein crystallization and to bring proteins closer together for supersaturation (Drenth, 2007).

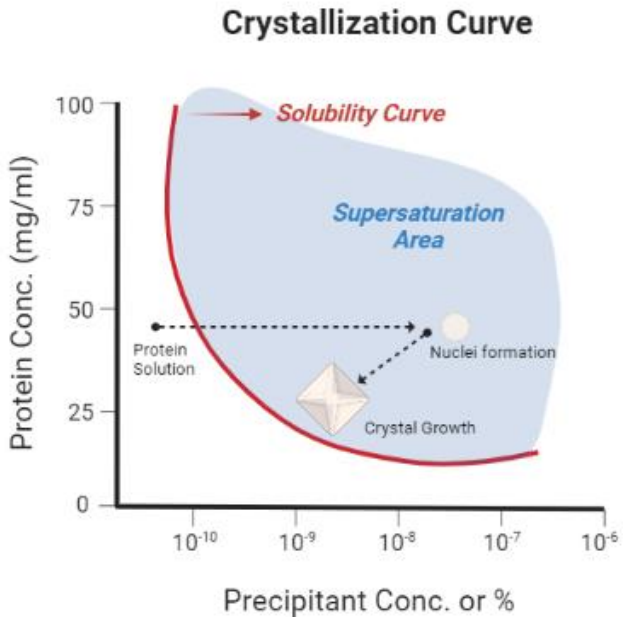

Figure 1.9: Crystallization curve that represents the protein crystallization principles within the supersaturated condition.

The most important factor in the crystallization of proteins is purity and 90-95% purity is required for a crystal with good diffraction. The second most important point for the crystallization of the protein is the presence of a solvent in which the protein can be

precipitated for the crystal. Finally, the protein-solvent mixture must reach supersaturation. The protein first creates nuclei in the presence of the solvent, then these nuclei gradually allow crystal growth with the addition of new proteins (Ilari and Savino, 2008; Drenth, 2007).

Crystallization techniques can be batch, liquid-liquid diffusion, vapor diffusion, or dialysis. Batch is one of the oldest crystallization methods, and it takes place by keeping the protein under oil with precipitation reagents in the presence of 1-2 ul of protein, and the protein gradually grows over time. Preventing salt crystal growth by rapid evaporation is also the reason for keeping precipitant condition and the protein under oil (Drenth, 2007).

After obtaining protein crystals the proteins must be diffracted with an X-ray source. In general, this technique is based on solving the three-dimensional structure of biological samples by diffraction, collecting two-dimensional images by detectors after the electron of the proteins hit by the X-ray rays, and processing these images in three-dimensional form. The data of a protein sample examined by X-ray crystallography first demonstrates residues' atomic positions as an "electron density" (Curry, 2015). During data processing and measurement, exposure time length, rotation angle per image, signal-to-noise ratio and absence of salt are important. During data analysis, indexing and scaling are performed and structure factors are determined from diffraction intensities and reflections. Finally, the obtained data or “electron density map” is integrated into the protein sequence with molecular replacement technique. During molecular replacement, a known or predicted structure is used as the starting point, then this structure is refined to be placed according to its differences in the electron density map. Structure refinement performed by CCP4i suite for rigid body refinement, coordinates, and B-factor refinements. After getting the initial model, which includes mostly secondary structure but not side chain conformations, model building is performed by COOT program. Via COOT the structure of the protein can be altered according to electron density map. This help to decrease R values and outliers for better statistic of the structure (Ilari and Savino, 2008).

1.8.2 Structural Biology Studies of Candida Boidinii NAD+-dependent FDH

There are not many structures or construct differences of *Candida boidinii* NAD+-dependent formate dehydrogenase. Among the 6 constructs in the Protein Data Bank (PDB), mutant and holo-forms 6D4C and 6D4B-encoded CbFDH constructs were collected at 1.45 A resolution, while apo- and holo-CbFDH constructs belonging to the wild-type form were collected at 1.75 and 1.5 A resolution. The other two mutant constructs reported with the codes 2J6I and 2FSS are in the apo form. The study from 2016 also revealed physiological relevant CbFDH structures with PDB code 5DNA for apo-CbFDH and 5DN9 for holo-FDH with azide covalent inhibitor and NAD+ co-enzyme. Due to the high enzymatic turn-over rate of the CbFDH enzyme, the structures containing the NAD+ coenzyme did not contain formate and could be resolved with a covalent azide inhibitor. Therefore, there is no data on the exact binding site of the formate.

Table 1.3: Structures of CbFDH enzyme in PDB. Identity percentage and substrate are indicated with PDB codes. % identity represents sequence similarity with CbFDH.

PDB	***% identity***		***Substrate***
6D4B	99.73	Crystal structure of Candida boidinii formate dehydrogenase V123A mutant complexed with NAD+ and azide	***azide***
6D4C	99.73	Crystal structure of Candida boidinii formate dehydrogenase V123G mutant complexed with NAD+ and azide	***azide***
2J6I	98.62	Candida boidinii formate dehydrogenase (FDH) C-terminal mutant	***apo***
2FSS	98.36	Candida boidinii formate dehydrogenase (FDH) K47E mutant	***apo***

Chapter 2:
METHODOLOGY

2.1 Materials and equipment

In this study, the gene was contained in the pET23a (+) expression vector. The competent cell strain was BL21 Rosetta-2 *E. coli* cells treated with the calcium chloride method before the heat shock procedure. LB medium contained 10 g tryptone, 10 g NaCl, and 5 g yeast extract per 1 liter. LB agar was also prepared with the same recipe but with 17 g agar addition. The mediums were sterilized by autoclave for 15 min at 121℃. Antibiotic stocks of ampicillin (Amp) and chloramphenicol (Chl) were prepared as 100 mg/ml and 35 mg/ml and IPTG as 0.4 mM concentrations, respectively. Buffer preparations were performed according to Table 2.1, then all buffers were filtered by 0.22 nm filter and vacuum pump. All chemicals in the buffer solution were obtained from IsoLab (Wertheim, Germany) and Sigma (Darmstadt, Germany).

Table 2.1: His A, His B, and Lysis buffer ingredient concentration as per liter.

	Lysis Buffer	**His A**	**His B**
Substance	**Concentration**		
NaCl	500 mM	200 mM	200 mM
Tris-HCl	50 mM	20 mM	20 mM
Imidazole	-	-	250 mM
Glycerol	0,05	-	-
Triton-X100	0.1%	-	-

2.2 Bacterial transformation

For bacterial transformation first BL21 Rosetta-2 *E. coli* cells and gene construct of CbFDH (wild-type and mutants) within pET23a (+) plasmids were placed on ice for 20 min. If the plasmids are in lyophilized form, first they are centrifuged for 3 min at 5000 rpm and then dissolved with freshly taken distilled water as 50 ul. Then, dissolved genes are vortexed for 30 sec and centrifuged at 9.000 rpm for 4 min twice.

50 ul of Rossetta-2 competent cells are aliquoted within sterile 1.5 ml Eppendorf tubes and 2 ul of each plasmid were added on competent cells separately, then mixed by pipetting and placed on ice for 20 min. After incubation, the cell-plasmid mixture was placed on a heat block at 42 °C for 45 sec for heat shock, which causes plasmid uptake of

competent cells. Then, heat-shocked cells were placed on ice for 2 more min and 500 ul of sterile LB medium are added to them to incubate cells for 90 min with gently stirring.

After cell incubation, for antibiotic resistance selection, cells were centrifuged for 2 min at 1500 rpm and the supernatant was discarded. The remaining cell pellet and medium dissolved and added to sterile antibiotic Amp and Chl contained LB agar (1:1000 ratio, antibiotic: medium) as 70 ul. The medium is spread with sterile glass beads with a gentle shaking of an agar plate. The agar plate was placed at 37°C for overnight incubation. The next day, the colonies were obtained on agar plates. Multiple single colonies are picked with sterile pipette tips or spread tips and inoculated within LB medium with (1:1000 ratio, antibiotic: medium) Amp and Chl antibiotics for growing starter and glycerol stock culture and incubated at 37°C and 110 rpm shaker. After 4 hours of incubation, 750 ul of culture is mixed with 750 ul sterile %80 glycerol within sterile cryo-tubes and after gently vortexing, the tubes are placed at -80°C for further usage.

2.3 Small-scale expression test

First, started culture from glycerol stock at -80°C is prepared within 150 ml LB medium with (1:1000 ratio, antibiotic: medium) Amp and Chl antibiotics and the addition of glycerol stock culture picked with the sterile pipette tip. The culture was incubated at 37°C and 110 rpm overnight. The next morning, 10 ml of culture is taken and centrifuged at 3500 rpm and 4°C. for 25 min to get before induction sample and stored at -40°C. Then the rest of the culture was induced with a 1:1000 ratio of 0.4 mM IPTG as 130 ul and incubated for 4 hours at 37°C and 110 rpm as described previously (Bulut et. al, 2020). After IPTG induction cells were harvested with 3500 rpm centrifugation for 25 min. The pellet which contained cells remains while the supernatant is discarded.

The cell pellets are sonicated after the addition of 20 ml Lysis Buffer, with max %25 amplitude, 1 sec on 1 sec off 30 sec time interval, and 3-4 rounds until cell pellet completely dissolved. After sonication solutions are transferred to centrifuged tubes and balanced. Avanti J26-XP centrifuge is used for precipitation of cell debris and getting protein at clean supernatant. The centrifugation is performed with JA20 rotor, at 15.000 rpm for 1h at 4°C. After sonication and centrifugation, whole cell samples for before and after induction, pellet, and supernatant samples are taken as 40 ul and 10 ul loading dye added for SDS-page electrophoresis. After sample taking, 300 ul equilibrated Ni-NTA

beads (with His A buffer) were added into supernatant samples and incubated between 3 hours and overnight at 4°C. After incubation, the supernatant solutions are centrifuged for 1 min, at 4°C and 1000 rpm. The supernatant is discarded after centrifugation and samples were washed twice with 25 ml His A buffer by placing the samples on ice for 20 min with inversion of tubes in 5 min intervals and centrifugation at 1000 rpm and 4°C for 1 min. After washing, His B buffer is added as 1 mL for the elution step, and the sample is placed on ice for 5-10 mins. After protein samples were removed from Ni-NTA beads 40 ul of pulldown samples are taken and 10m ul loading dye was added for the SDS page. As result, SDS-Page gel electrophoresis is performed. The rest of the Ni-NTA beads are kept within the used Ni-NTA bead tube. For pulldown assay, samples are kept on ice or 4°C for preventing heat degradation.

2.4 Large-scale protein production and purification

Starter cultures are prepared within LB medium with Amp and Chl antibiotics (1:1000 ratio, antibiotic: medium) and the addition of glycerol stock culture similar to the small-expression check. This culture was incubated at 37°C and 110 rpm overnight. The next morning starter cultures are taken into 1.5 L LB medium with Amp and Chl antibiotics (1:1000 ratio) and incubated until the O.D.590 value reaches at least 0.6. After incubation, the cultures were induced by 1.5 ml (1:1000 ratio) 0.4 mM IPTG and incubated for 4 hours at 37°C and 110 rpm. The cells are harvested by 3500 rpm centrifugation for 25 min and collecting pellet within 50 mL sterile falcon tubes. The harvesting step must be on ice preventing protease degradation of proteins within cells.

After cell harvest, cells are sonicated within 40 mL Lysis Buffer in each pellet contained tubes with setting 1 sec on 1 sec off pulse for 30 sec at %45 amplitude for 3-4 rounds. The Sonication process is applied until the solution becomes clear, and viscosity disappears caused by DNA impurity. After sonication, the cell solution is taken to ultracentrifuged and balanced carefully. The centrifugation is performed with a Beckman Coulter Optima L-XP ultracentrifuge device and Ti-45 rotor type for 1 hour at 4°C and 35.000 rpm. After centrifugation, 40 ul supernatant and pellet samples are taken and mixed with 10 ul SDS loading dye. Supernatant separated as load sample for Ni-NTA column purification.

For Ni-NTA column affinity chromatography Akta-Go system (Sigma-Aldrich Darmstadt, Germany) is used. For this purpose, first, the column is placed in the Akta-system, and capillaries washing by dH_2O pumping. Then, the flow rate of the pump is fixed to 2.5 mL/min and the Ni-NTA column is equilibrated with 3x volume column His A buffer, after 1 M imidazole buffer cleaning, until the column reaches equilibration. After equilibration load sample is injected by the sample capillary and the flowthrough sample was collected from "Outlet1" after UV280 reaches the maximum point until the decrease. After loading sample column is washed with 3x volume of His A buffer for removing unbound proteins. The washing step is performed until the UV280 line reaches the bottom line and remains constant. After washing, elution of His-tagged CbFDH is made by His B buffer injection as 3x volume of the column. After loading 1x volume of column His B, UV280 starts to increase and His-tagged CbFDH elution collected as 15% His B contained and 100% His B contained samples. Flowthrough and elution samples are taken as 40 ul and mixed with 10 ul SDS loading dye as well. After purification, the column is washed with 3x volume 1M imidazole buffer and capillaries are washed by pumping dH20. The device is cleaned and closed at the end of the process. The purification samples are investigated by SDS-Page gel electrophoresis.

2.5 Protein verification via SDS-Page gel electrophoresis

For SDS-Page gel electrophoresis gel, casting is performed first. For this clean glass, plates are placed on a trace and gel casting stand and checked for leaks. After that, separating gel for 15% polyacrylamide gel is prepared according to the recipe as mentioned in Table 2.2 in the same order as the list.

After TEMED polymerization of acrylamide starts, thus prepared solution is quickly poured into the pool between glass plates. For linearization of separating gel and limiting oxygen exposure, pure ethanol is poured until the top of the glass plates and gels have waited for 30 min. After gel solidification, the rest of the ethanol is removed, and stacking gel is prepared in same order in Table 2.2 and quickly poured to the top of the separating gel. After pouring, gel combs are placed on the top and between glass plates for obtaining wells. Gels solidify within 30 mins and are kept at +4°C within dH2O.

Table 2.2: Separating and stacking gel recipe for 15% SDS-Page Gel

Separating Gel Stock Solutions (4, 15% Gels)	**Volumes**
dH_2O	4.4 mL
30% Acrylamide / 0.8% bis-acrylamide	10 mL
1.5 M TRIS-HCl (pH=8.8)	5.2 mL
10% SDS	200 µL
10% APS	200 µL
TEMED	20 µL
Stacking Gel Stock Solution (4 Gels)	**Volumes**
dH_2O	5.86 mL
30% Acrylamide / 0.8% bis-acrylamide	1.34 mL
0.5 M TRIS-HCl (pH=6.8)	2.6 mL
10% SDS	100 µL
10% APS	100 µL
TEMED	10 µL

After casting of gel, previously obtained purification samples are loaded to gel wells as 2 ul for the ladder, 5 ul for pellet, supernatant, and flowthrough samples, and 10 ul for elution samples. Then, the gel is run within x1 Running Buffer (120 mM Tris-base, 1.92 M glycine, and 1% SDS in 1 liter) in a gel running tank and the power supply setting is constant at 120 V electric current for approximately 2 hours. After running, the gel is removed between glass plates carefully and washed with distilled water three times. Then, the staining buffer is poured on the gel and waited for 20 mins, and discarded. After de-staining of gel within distilled water, gel images are taken by the scanner.

2.6 Protein crystallization

Protein crystallizations of CbFDH wild-type and mutant proteins were performed by sitting-drop microbatch underoil screening method. In this method more than three thousand precipitant conditions were screen for CbFHDs. Protein and crystallization condition ratio for screening was 1:1 for this purpose. Basically, first, protein solution was loaded to 96-well Terasaki plate with automated dispersion pipette as 0.83 ul to each well. Then 0.83 ul of crystal screen conditions were loaded to each well separately. After mixing protein and condition 16.6 ul of paraffin oil were added to each well to preventing of evaporation as described at Figure 2.1. Crystal screening plates were kept in both room temperature and at 4°C to increase change to obtain crystals. However, crystals growth was more favorable at 4°C because of salt crystals occurrence at room temperature.

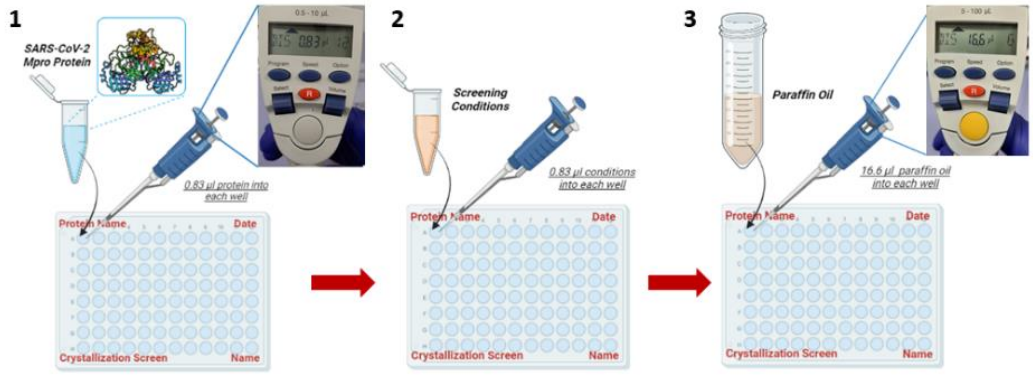

Figure 2.1: Protein crystallization procedure illustration. Plate labelling and pipetting volumes and orders are defined. (Ertem et. al, 2021)

Stereomicroscope is used for following crystal growth. The crystals are checked at 1 week, 1 month and 2 months. Obtained crystals were duplicated to 96-well Terasaki plate again and/or Additive Screen and Cryo-Pro screening conditions were added to optimize those crystals. The wild-type CbFDH protein crystals were obtained with 47#-Wizard III (Hampton research, USA) condition that includes 30% (w/v) PEG 5000 MME,100 mM MES/ Sodium hydroxide pH:6.5 and 200 mM Ammonium sulfate solutions.

2.7 X-Ray data collection, reduction, and refinement

The data collection of wild-type CbFDH crystals which grown with includes 30% (w/v) PEG 5000 MME,100 mM MES/ Sodium hydroxide pH:6.5 and 200 mM Ammonium sulfate solutions were made by Rigaku's XtaLAB Synergy Flow X-ray diffractometer (Rigaku, Japan) at University of Health Sciences (Istanbul, Turkey). For data collection, flash frozen technique was applied to the crystals by liquid nitrogen after pickup and the data collection was under cryogen temperature 100K. The diffractometer includes PhotonJet-R (Rigaku, Japan) X-ray generator and HyPix-Arc-150° detector (Rigaku, Japan). Data collection was performed with 1.54 Å wavelength and 40 kV, 30 mA, 1200.0 W and 9052 rpm generator power. The detector distance was 47.00 mm, scan width was 0.2 degree and exposure time was 3.0 min with 10% beam intensity. The JetShadow was applied during data collection with α=30.00, β=0.00, jet width at 13.00 and jet distance at 6.00 settings. Total run time was 2 h 58 min 46 sec for data collection. Reflection with bad profiles also rejected during data collection. After data collection, the resolution obtained as 1.8 A and $P_{1\ 1\ 1}$ space group with cell dimensions a=54.41 Å b=68.890 Å c=109.875 Å α=77.99 β= 89.12 γ= 79.92. However, a longer data collection

was performed to 1.0 Å resolution by X-ray diffraction with 0.25 scan width, 47.00 mm detector distance, 10% beam intensity, 10 min exposure time however, after data reduction, the data obtained as 1.4 Å resolution. Unit cell dimensions were obtained as a=54.41 Å b=68.890 Å c=109.875 Å α=77.99 β= 89.12 γ= 79.92 with $P_{1\,1\,1}$ space group. Data reduction was performed by CrysAlisPro version 1.171.42.35a (Rigaku Oxford Diffraction, 2021), and scaling was done by SCALE3 ABSPACK scaling algorithm. The reduction included frame scaling, B-factor correction, and absorption correction.

For data refinement, .mtz file is applied to molecular replacement with PDB ID: 5DNA as model structure (Guo et al., 2016). Molecular replacement was performed by PHASER program in PHENIX (v 1.20.1-4487) (McCoyet al., 2007; Adams et al. ,2010). After initial rigid body refinement, model building was applied by COOT software for placement of water molecules, rotomer fit, rigid body fit, and real space refine of the structure within electron density map (v 0.9.6) (Emsley an0064 Cowtan, 2004). Figure generation was made by PyMOL (v 2.4.1.) (DeLano, 2020). Table for data collection and refinement information can be seen below (Table 2.3 and 2.4).

Table 2.3: The data collection statistic of 1.4 A resolution wt-CbFDH crystal.

Dataset	**Wt-FDH**
PDB ID	-
Data collection[1]	
Beamline	Turkish Delight
Space group	**P1 1 1**
Cell dimensions	
a, b, c (Å)	54.132, 68.959, 109.79
α, β, γ (°)	78.139, 89.482, 81.115
Resolution (Å)[2]	24.75-1.39 (1.51-1.39)
CC (1/2)	99.9 (34.0)
CC*	100.0 (70.8)
I / σI	11.0 (0.00)
Completeness (%)	99.3 (91.5)
Redundancy	151.46 (148.36)

Table 2.4: The data refinement statistic of 1.4 A resolution wt-CbFDH crystal.

Refinement	Wt-FDH
Resolution (Å)	21.83-**1.40** (1.43-1.40)
No. reflections	301717 (18694)
R_{work} / R_{free}	**0.17/0.18** (0.38/0.36)
No. atoms	
Protein	22164
Ligand/Ion/Water	2032
B-factors	
Protein	25.68
Ligand/Ion/Water	33.7
Coordinate errors	0.18
R.m.s deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.77
Ramachandran plot	
Favored (%)	1355 (98.01%)
Allowed (%)	50 (1.99%)
Disallowed (%)	**0 (0.00)**

2.8 Gaussian Network Model (GNM) Analysis

Normal Mode Analysis with GNM was done using ProDy for 1.4 A resolution wild-type apo-CbFDH and holo-CbFDH (PDB ID: 5DN9) (Bakan et al., 2011). Cα atoms of the proteins are used for contact maps. Accordingly, the apo-CbFDH had 709 atoms included in Kirchhoff matrix and the holo-CbFDH structure had 742 since PN, C1, N1, C4, C5, C6, C7 and N7 atoms of NAD+; and azide N2 atom and chloride atom were added in the holo-CbFDH structure's GNM analysis. Cutoff distance was selected as 8.5 Å in both models for calculation of pairwise interactions and default spring constant was 1.0. Maximum correlations at these cut-offs are for Apo-CbFDH 0.577 and Holo-CbFDH 0.727. Since n-1 non-zero modes were calculated in all calculated normal modes by GNM, the apo- structure has 708 and holo- structure has 741 non-zero modes.

The experimental B-factors were examined with the theoretical fluctuations calculated with all GNM modes, to understand reliability of the analysis. Overall GNM modes were employed to obtain cross-correlations between residue fluctuations. The differences in cross-correlations between holo- and apo-CbFDH structures were calculated by subtraction of cross-correlation each chain of apo-CbFDH from holo-CbFDH. Both cross-correlation and cross-correlation differences results were shown by heat-maps. Apo- and holo- structure weighted squared fluctuations were examined and compared to understand CbFDH dynamics. The global motions were calculated by the ten slowest modes while local motions were analyzed by the ten fastest modes.

2.9 Alphafold Analysis

Alphafold structure prediction was performed for E47K mutation and wt-CbFDH by AlphaFold2.ipynb GitHub open-source script. The sequence is used for E47K-CbFDH mutation was taken from 2FSS and for wt-CbFDH 5DNA sequence was performed with Alphafold2. Query sequence was entered twice with “colon (:)” to indicate homodimer structure of CbFDH. The template mode was selected as “"none" = no template information is used”. Multiple sequence alignments (Msa) mode and pair mode were chosen as “MMseqs2 (UniRef+Environmental)” and “unpaired+paired”, respectively. Model type was selected as "AlphaFold2-ptm" and complex prediction was made by "AlphaFold-multimer-v2" (Jumper et al., 2021). The prediction run cycle was three to obtain sequence-based structure prediction. The prediction structures were compared with 2FSS structure to understand efficiency of prediction of mutation effect on structure.

.

Chapter 3:

RESULTS

Pre-cloned CbFDH gene was in pET23a (+) plasmid, and this constructed transformed into BL21-Rosetta2 E. coli strain. The transformation success was observed with multiple colonies on agar plates after overnight incubation (Data not shown). After transformation, efficiency of gene construct for recombinant protein production was checked via small scale expression check. Accordingly, wild-type CbFDH and all mutant forms of CbFDHs construct express CbFDH protein after IPTG induction (Figure 3.1).

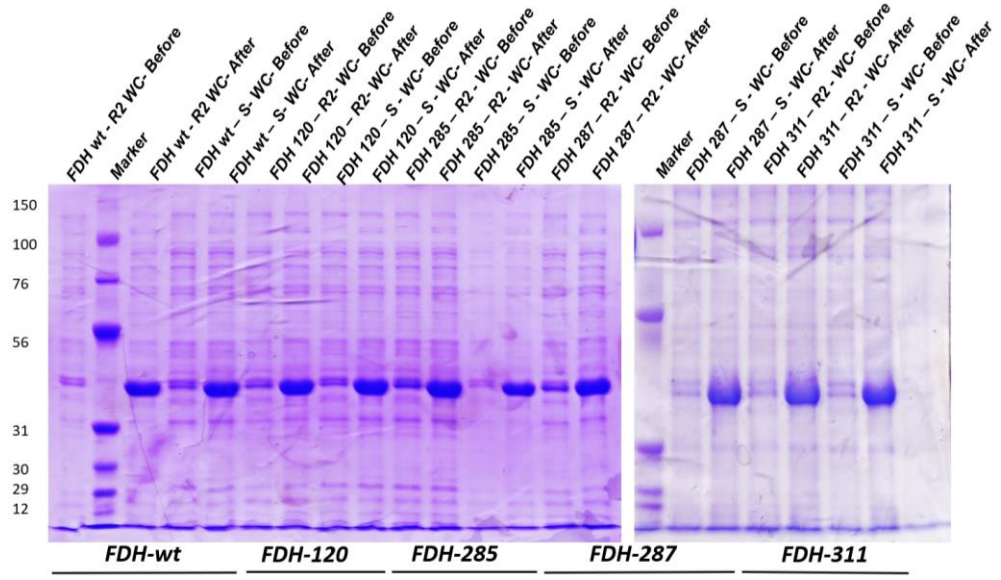

Figure 3.1: Small scale expression verification of wild-type and mutants CbFDH enzyme for 2 hours induction at 37°C.

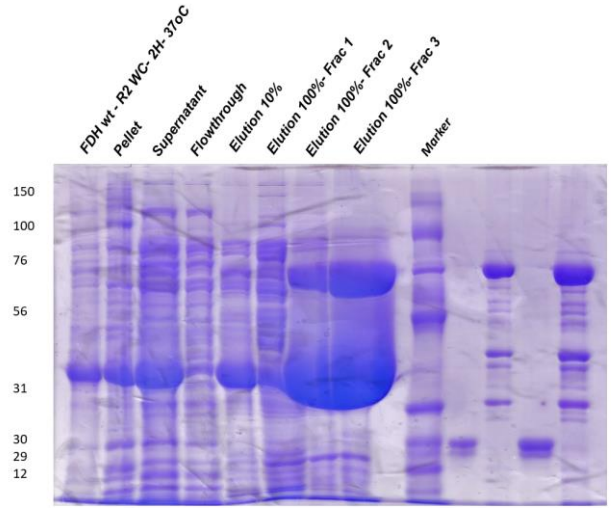

Figure 3.2: Large scale protein production and purification verification of wt-CbFDH at 37oC for 3h.

After small scale expression check. Large scale protein production for wt-FDH was performed as 6 L. Purification has more than 90% purity, which is required for crystallization process, verified with a tick band at 41 kDa as CbFDH molecular weight (Figure 3.2). Because of high-protein condition as 35 mg/mL there was also dimeric band of wt-CbFDH protein at 84 kDa.

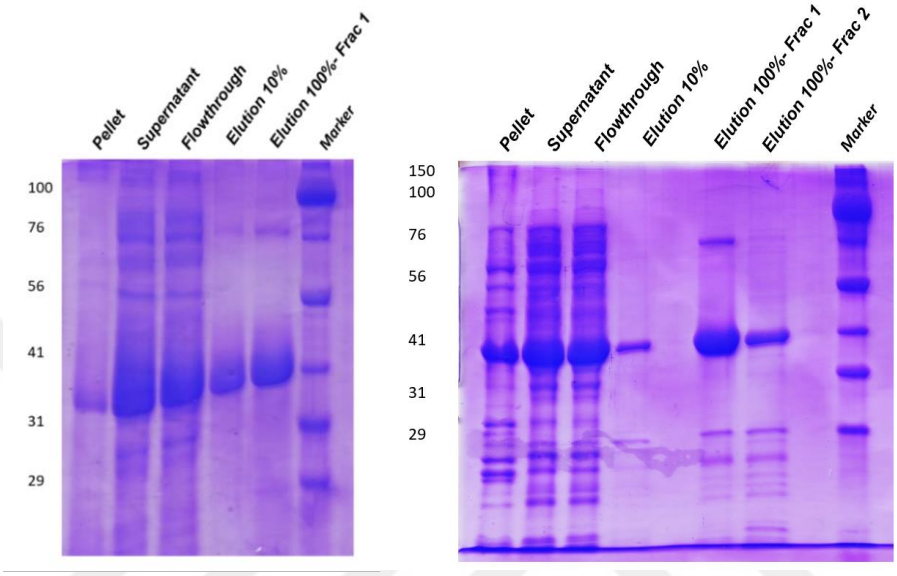

Figure 3.3: SDS-page gel electrophoresis for protein purification verifications of CbFDH-120* (left) and CbFDH-285* (right). Elution fractions have 41 kDa CbFDH protein band.

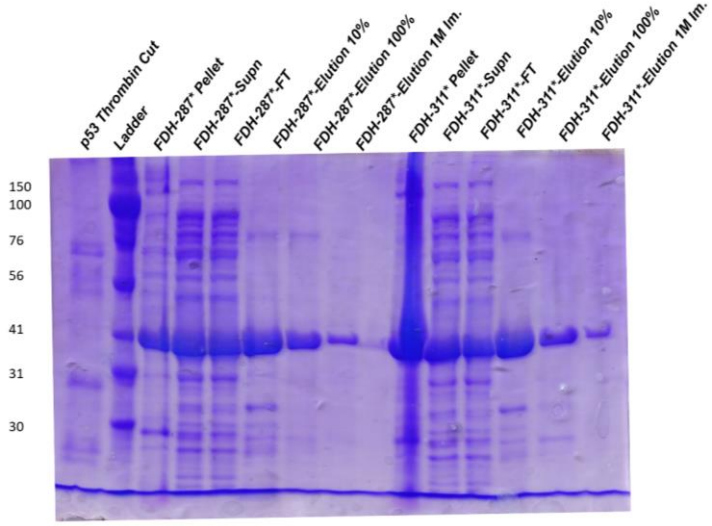

Figure 3.4: CbFDH-311* and -287* mutants large scale protein purification verification by SDS-page gel electrophoresis.

Then, 1.5 L of mutant CbFDHs large scale protein productions were made, and the proteins were purified successfully with more than 90% purity (Figure 3.3 and 3.4). Both 10% and 100% His B buffer included elution samples included high amount of

recombinant protein but not impurity. The his-tag of CbFDH proteins could not cut via restriction enzymes since there was no specific cut-site within the gene construct.

The crystallization screening was performed for all wild-type and mutant for of CbFDH by sitting-drop microbatch under oil method. The crystal formation was followed each 2 weeks. The micro crystals were seen at the end of first two weeks, however first crystals obtained six weeks after crystallization. All crystals were obtained at 4°C incubation and first crystals obtained at the beginning of wintertime. Most of the mutant protein crystals were obtained with similar crystal morphology. Because of late setup of XRD device first crystals degraded after 3rd month. The found crystals were indicated at Figure 3.5 and 3.7 for wt and each mutant form of CbFDH proteins.

Figure 3.5: Wt-CbFDH protein crystals with different conditions that obtained by sitting-drop microbatch under oil method. The crystals obtained after two months of crystallization.

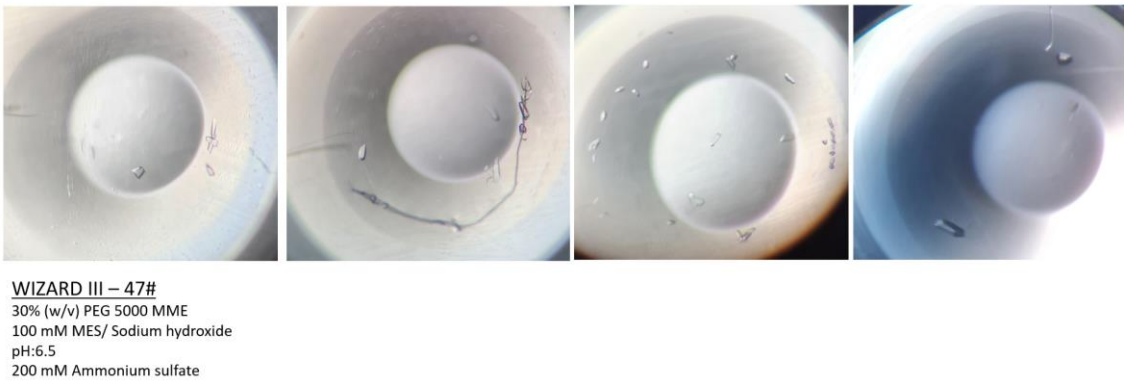

Figure 3.6: Wt-CbFDH protein crystals replicas gave result with Wizard III-47# condition that used for data collection.

The condition replicas of crystals were made to 96-well Terasaki plate and incubated at room temperature. But there was no protein crystal observation after 2 months incubation. The new replicas were incubated at 4°C again and first wt-CbFDH crystals

obtained in replica. One of those crystals diffracted by X-ray diffractometer and data collection performed (Figure 3.6).

Figure 3.7: Several crystals conditions produce protein crystals with mutant-CbFDH proteins.

The data collection parameters were 1.54 Å wavelength and detector distance was 47.00 mm, scan width was 0.2 degree and exposure time was 3.0 min and beam intensity was 10%. After data collection, the resolution was 1.8 A and cell dimensions were

obtained as a=54.41 Å b=68.890 Å c=109.875 Å α=77.99 β= 89.12 γ= 79.92 with $P_{1\,1\,1}$ space group. Total run time was 2 h 58 min 46 sec. However, a longer data collection time and different detector range applied with 0.25 scan width, 47.00 mm detector distance, 10% beam intensity, 10 min exposure time parameters. The diffraction pattern had separated and clear stops without clash between 12.5 A and 1.0 resolution (Figure 3.8). 1.0 A resolution data of wt-CbFDH was collected with total 15h 43m 19s run time. P_1 space group with a=54.13 Å b=68.95 Å c=109.76 Å α=78.14 β= 89.48 γ= 81.11-unit cell parameters were observed for 1.4 A data. The higher resolution data was reduced to 1.4 A according to completeness and refined as described at methodology section. The refined data has R_{work} 0.171 and R_{free} 0.187 after the third refinement cycle. Structure was homodimer and secondary structure of enzyme contained Roseman-folds as expected (Figure 3.9).

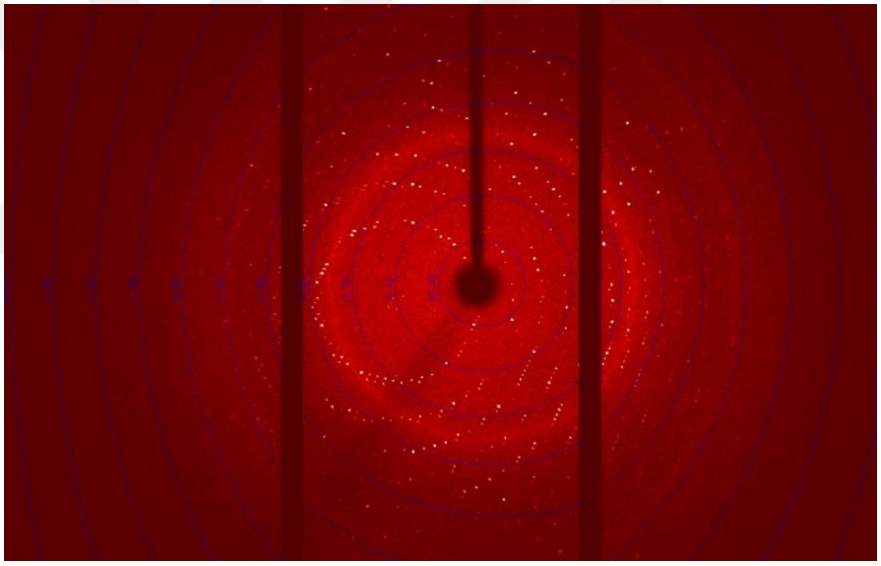

Figure 3.8: The diffraction pattern that is obtained while 1.0 A resolution data collection process.

The wt-CbFDH data with 1.4 A is apo-form and it was compared previous apo-form CbFDH structure with 1.75 A resolution. Accordingly, those structures aligned well after ce-aling comment without outlier rejection and RMSD value after alignment was 0.149 (740 atoms). However, there was conformational differences at some of charged amino acid side chain that expose to water. Asn-51, Lys-109, Lys-246, Lys-291, Glu-325 residues at chain A and Lys-12, Glu-18, Lys-19, Asp-36, Asn-147, Arg-174, Glu-213, Glu-217, Lys-246 and Glu-325 at chain B were observed different side chain conformations (Figure 3.9).

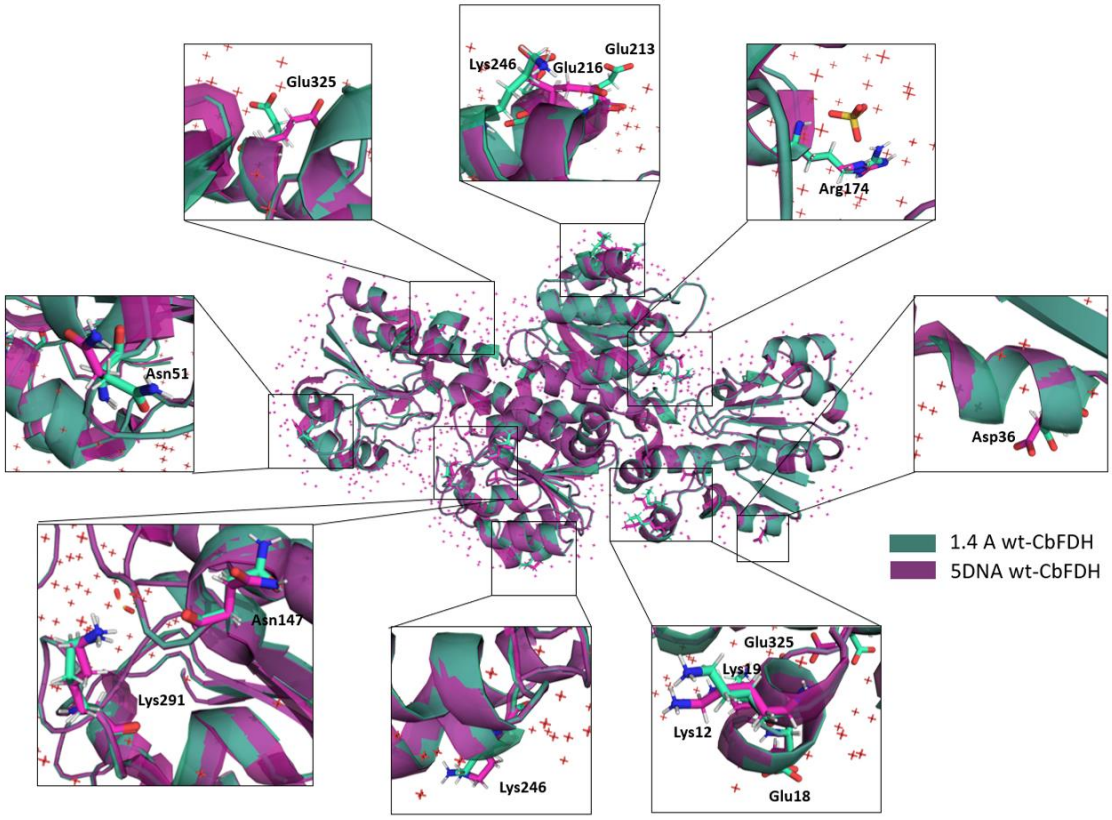

Figure 3.9: 1.4 A apo-CbFDH structure and 5DNA apo-CbFDH structure comparison. Novel structure and 5DNA is colored green cyan and light magenta, accordingly. The residues with conformational changed were indicated by sticks. RMSD was 0.149 after ce-alignment.

In addition to conformational changes, electron density maps of both structures were examined via Pymol (Figure 3.10). The mesh of electron density maps was doubled, and mesh level was selected as 1.0 A for both electron density map. According to electron density map, 1.4 A structure does not include disordered residues, while 5DNA structure has some disordered residues on some of the charged (especially on Lys) residues (Figure 3.10). Those residues are Lys12, Glu18, Lys19, Asn51, Gln56, Lys83, Lys109, Glu242, Lys206, Arg174, Lys246, Lys189 Ile215, Lys236, Glu332, Lys328 Lys338. The new structure has also partial electron cloud loss at Asp36 (Diff conf) Lys201, Lys24, Arg178, Lys219 and Lys249 residues. However, when this electron cloud loss checked by COOT, missing clouds do not affect side-chain conformation (data not shown).

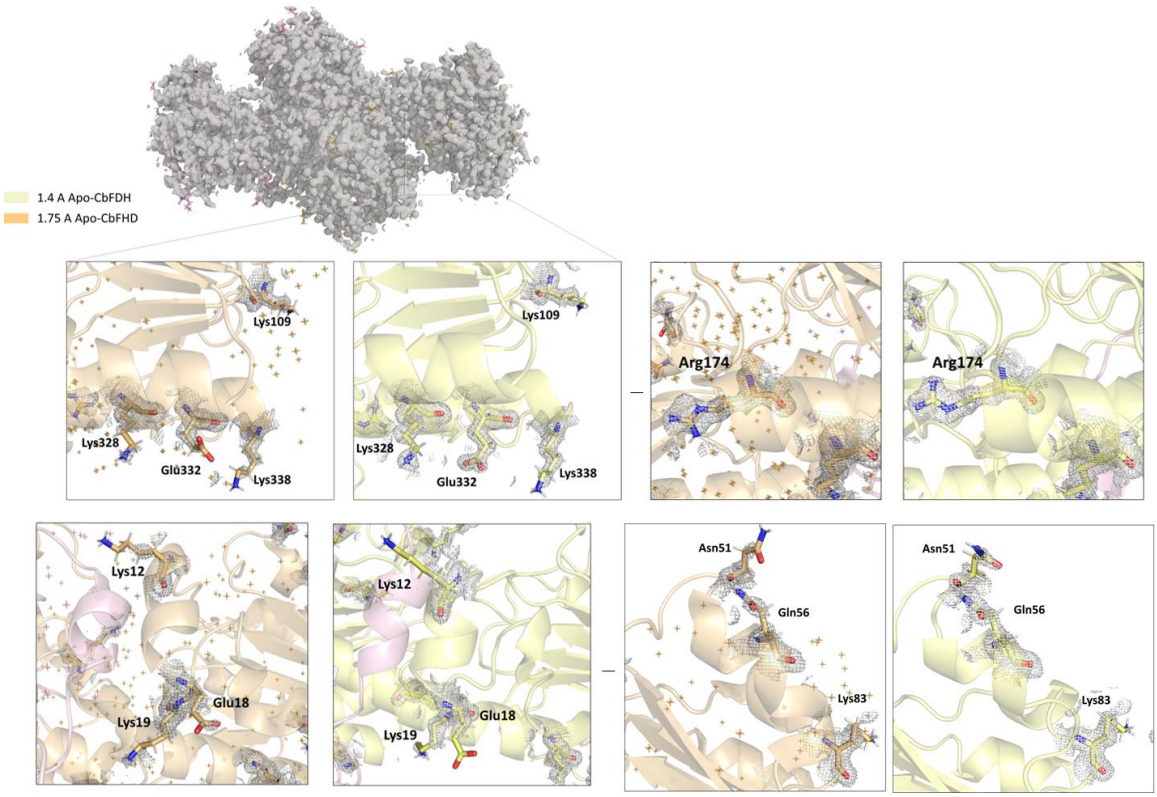

Figure 3.10: Electron density map of new 1.4 A resolution apo-CbFDH structure with 1σ level 2Fo-Fc simulated annealing-omit map that are presented with grey mesh. The electron clouds on some of hydrophilic and charged residues are improved.

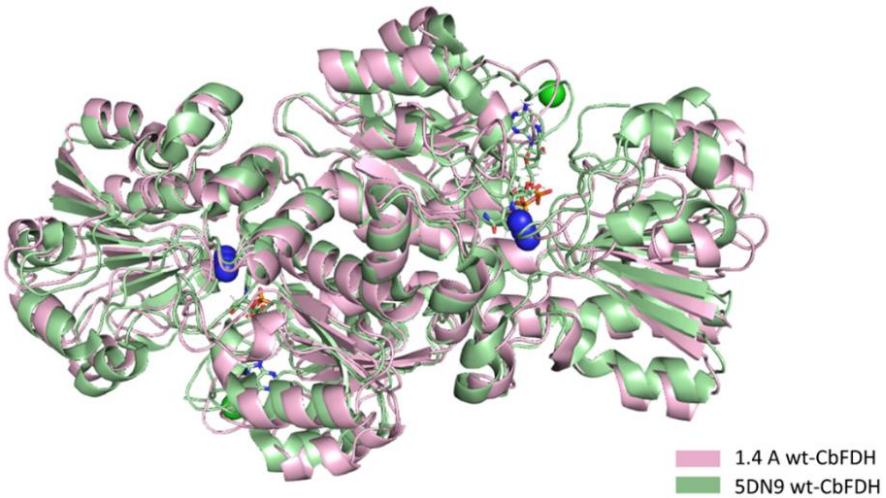

Figure 3.11: Alignment of 1.4 A resolution apo-CbFDH and 5DN9 coded holo-CbFDH structures were obtained with 1.968 A RMSD value.

1.4 A apo-CbFDH structure and holo-CbFDH structure (5DN9) were also compared after alignment. RMSD value of this alignment was found as 1.968 A. Even carbon-backbone of secondary structure did not fit well and was observed with tilt (Figure

3.11). Thus, morph was applied to understand the differences of holo- and apo-state of the enzyme.

Accordingly, morph analysis revealed that, CbFDH close the binding pocket with substrate and co-factor binding by major conformational change (Figure 3.12). The binding of substrate and NAD+ brings catalytic domain and binding domain together. Moreover, the catalytic residues Arg258 and Gln287 were observed significant conformational change since side chains of those residues were shifted to upper part of binding pocket.

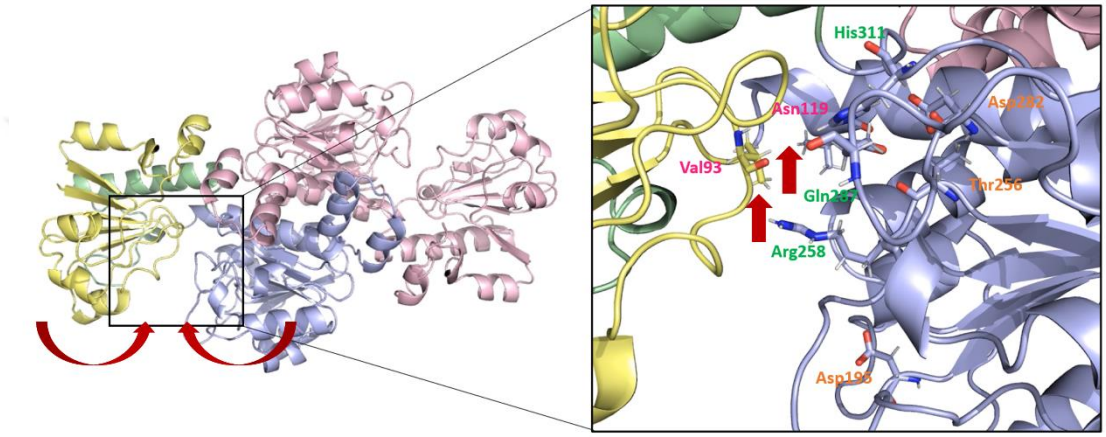

Figure 3.12: Morph analysis between apo- and holo- state of CbFDH reveal major conformational change and possible motion of the enzyme catalytic activity. The analysis was performed by Pymol.

The GNM analysis was performed with apo- and holo- CbFDH structures were investigated for binding and activity structural dynamics. The cross-correlation by GNM analysis represents each chain for homo-dimer enzyme, CbFDH, and with three domains (Figure 3.13a and 3.13b). First and second domain of each residues has high intrachain cross-correlation with each other. In contrast, third domain has high cross-correlation with itself.

For interchain cross correlation, third domain has high cross-correlation by residues around 160th and 140th position for 1-40th with 310-330th and 280-310th residues of neighbor chain, respectively. Since those residues have close positions to each other the result was expected for interchain cross-correlations. The cross correlation of Apo-CbFDH has higher correlation values according to Holo-CbFDH results (Figure 13.3b).

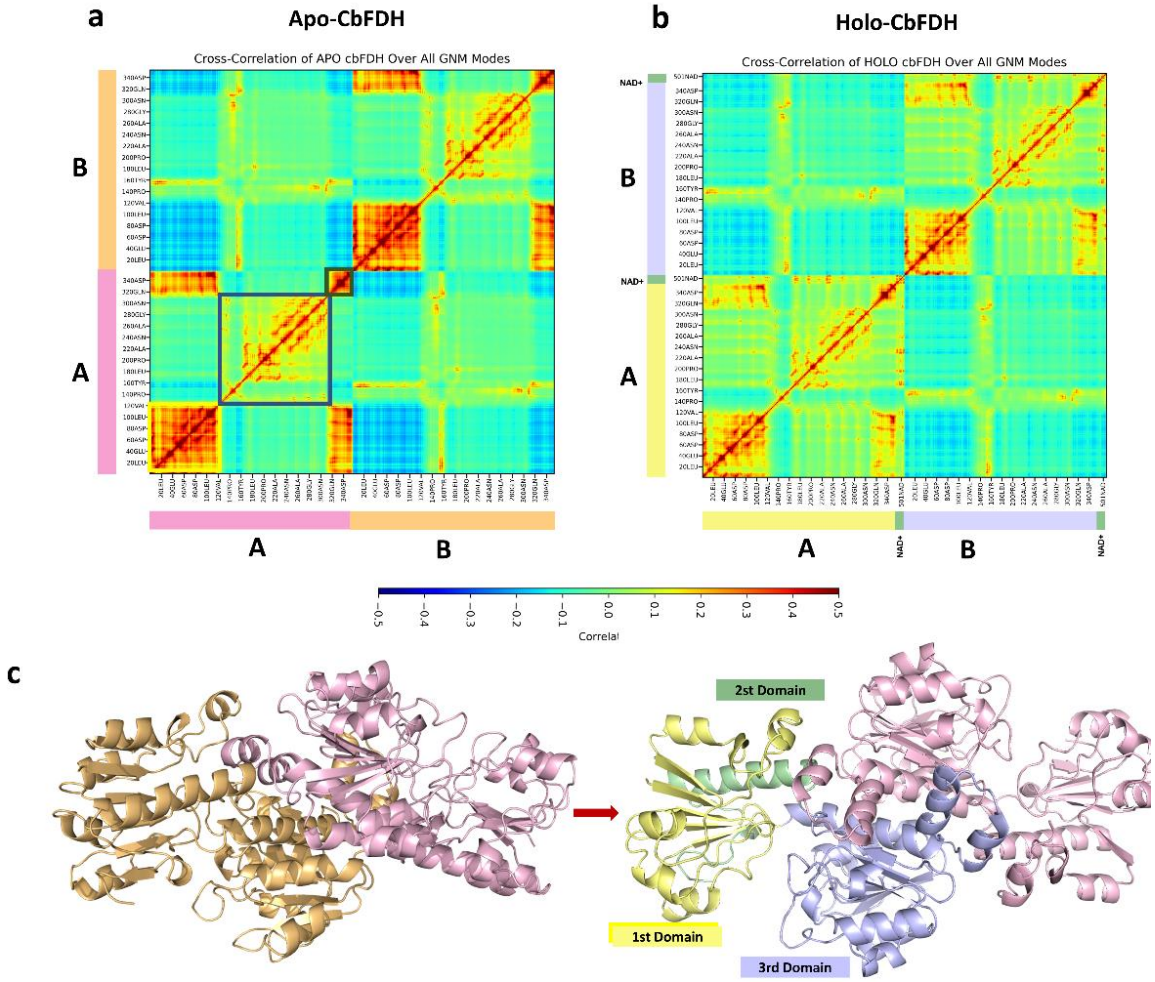

Figure 3.13: Cross-correlation maps of apo- (a) and holo-CbFDH (b) enzyme. Apo form of enzyme has higher correlation but similar pattern with holo-CbFDH. c) Homodimer structure color coded according to Chain A and B (left) as light pink and light orange. The GNM result revealed three domains with high cross-correlation (right) as light yellow, pale green and light blue for first domain, second domain and third domain, respectively. While high cross-correlation was indicated with red, low cross correlation was mentioned with blue between 0.5 and -0.5 range.

Cross-correlation differences map also confirmed this difference with higher correlation between first and second domain and less correlation for third domain at apo-form of intrachain correlations. On the other hand, it can be clearly seen, apo-form has higher correlation between interchain residues through third domain (Figure 3.14). As morph analysis showed that the catalytic (first and second) and binding (third) domains has more approximate position in holo-state CbFDH. This conformational change also

observed with higher correlation of third domain with others and lower cross correlation between first two domains in holo-state CbFDH cross-correlation.

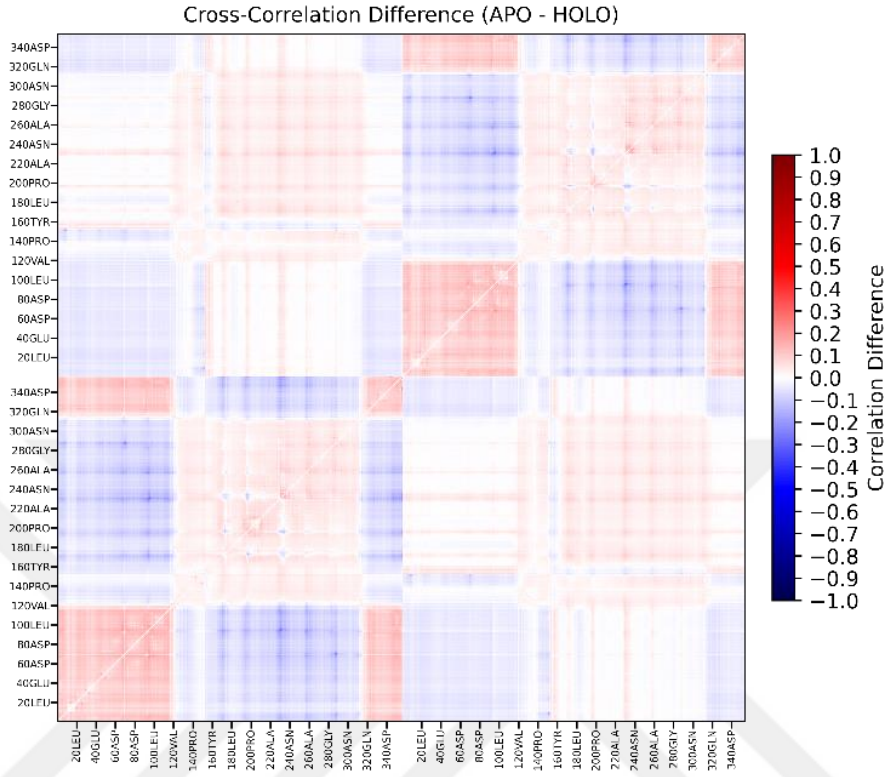

Figure 3.14: Cross-correlation differences apo-CbFDH over holo-CbFDH. Correlation differences indicated with red color for higher correlation and blue for lower correlation between 1.0 and -1.0 range.

GNM analysis also revealed square fluctuations of CbFDH for weighted 10 slowest and fastest mode. The apo-state CbFDH has higher fluctuations around binding pocket residues than holo-CbFDH at 10 slowest mode analysis that gives information about global motions (Figure 3.15). The global motions of GNM analysis that are provided by the ten fastest modes were verified with GNM mode variance that shows first ten modes have the highest variance (Figure 3.16). Moreover, the ten fastest weighted mode of square fluctuations may indicate local motions as hotspots or hinge residues that important for folding. Accordingly, some important residues for catalytic activity and NAD+ and formate binding was observed with localized motion (Figure 3.15). On the other hand, apo- and holo- state have different hotspot residues in this manner. While holo-state hotspot residues are Val93, Asn119 which are important for formate binding, Thr256 which is indicated as fundamental for NAD+ binding; apo-state hotspot residues are the non-binding related residues which are around binding pocket residues.

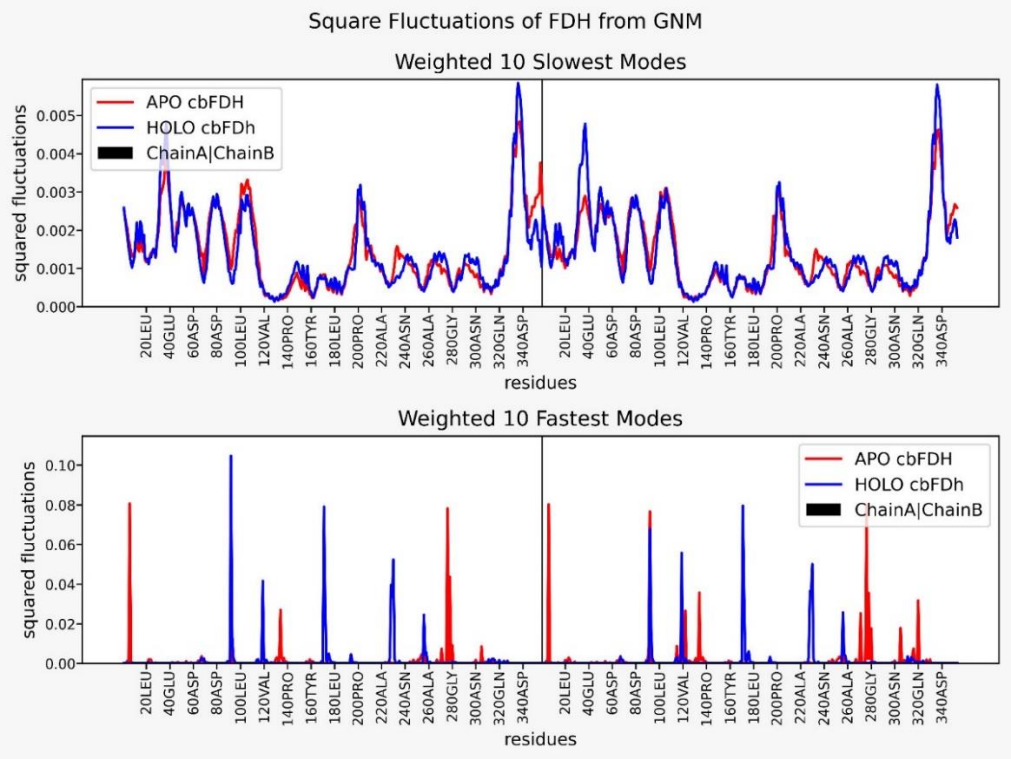

Figure 3.15: Squared fluctuations of CbFDH for apo- and holo- forms. The 10 slowest modes represent global motions (above) while 10 fasters modes indicate local motions (below).

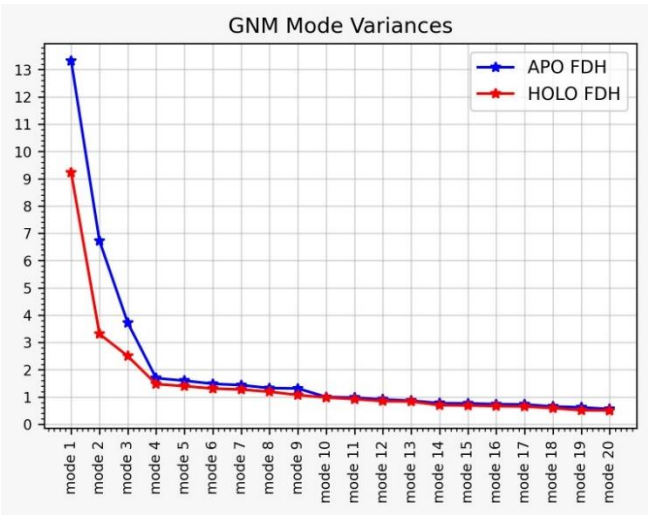

Figure 3.16: GNM mode variance represent that the first ten fastest modes are highest.

Finally, the co-enzyme NAD+ and azide inhibitor atoms and chloride ion were tested for cross-correlation with overall residues by GNM analysis. The cross-correlation data present that, the approximate residue to those molecules gives high correlation at intrachain as expected (Figure 3.17a) for indicated residues with blue color. On the other hand, the residues that shown by light orange has high cross-correlation with NAD+

atoms while only has approximate position with azide atoms. However, interchain cross-correlation interestingly were observed with high correlation between residues 135-160th (indicated with brown color) and NAD+ and azide selected atoms (Figure 3.17b), while there is not direct residues interactions or polar contact.

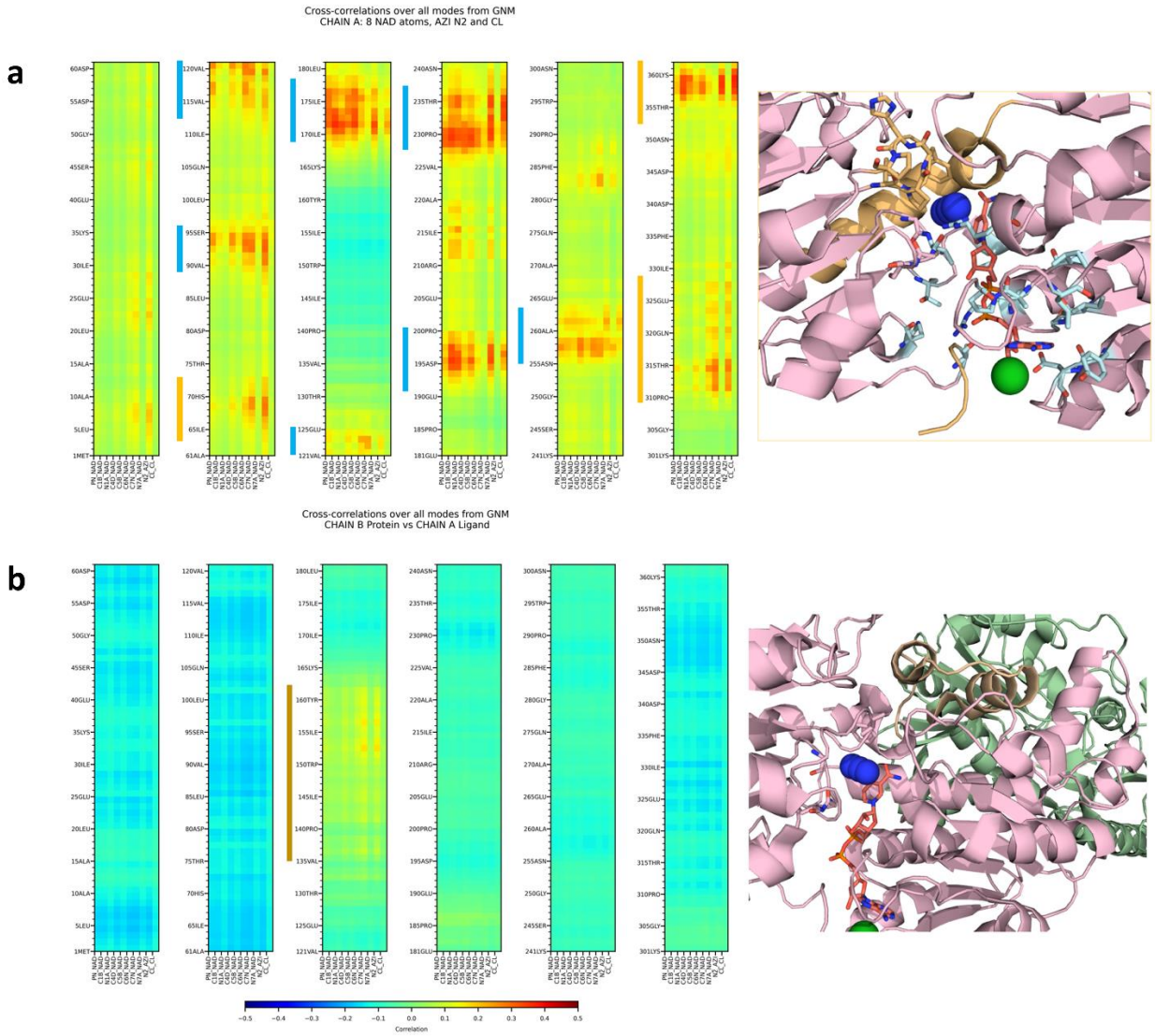

Figure 3.17: Cross-correlation of intra- (a) and inter-chain (b) residues with NAD+, azide and chloride molecules. Heat maps represent high and low cross-correlation with red and blue between -0.5 and 0.5 range, respectively.

The obtained mutant CbFDH crystals were degraded after second and third round of crystallization replicas since there was technical problem with X-ray diffractometer. Thus, there is no novel structure for mutant form of CbFDH proteins within our constructs. However, a recently hot topic, Alphafold2 algorithm is estimated as a potential for mutant structure prediction. Since a single residue mutant structure (PDB: 2FSS)

CbFDH with K47E mutant and the Alphafold2 prediction of this sequence were examined.

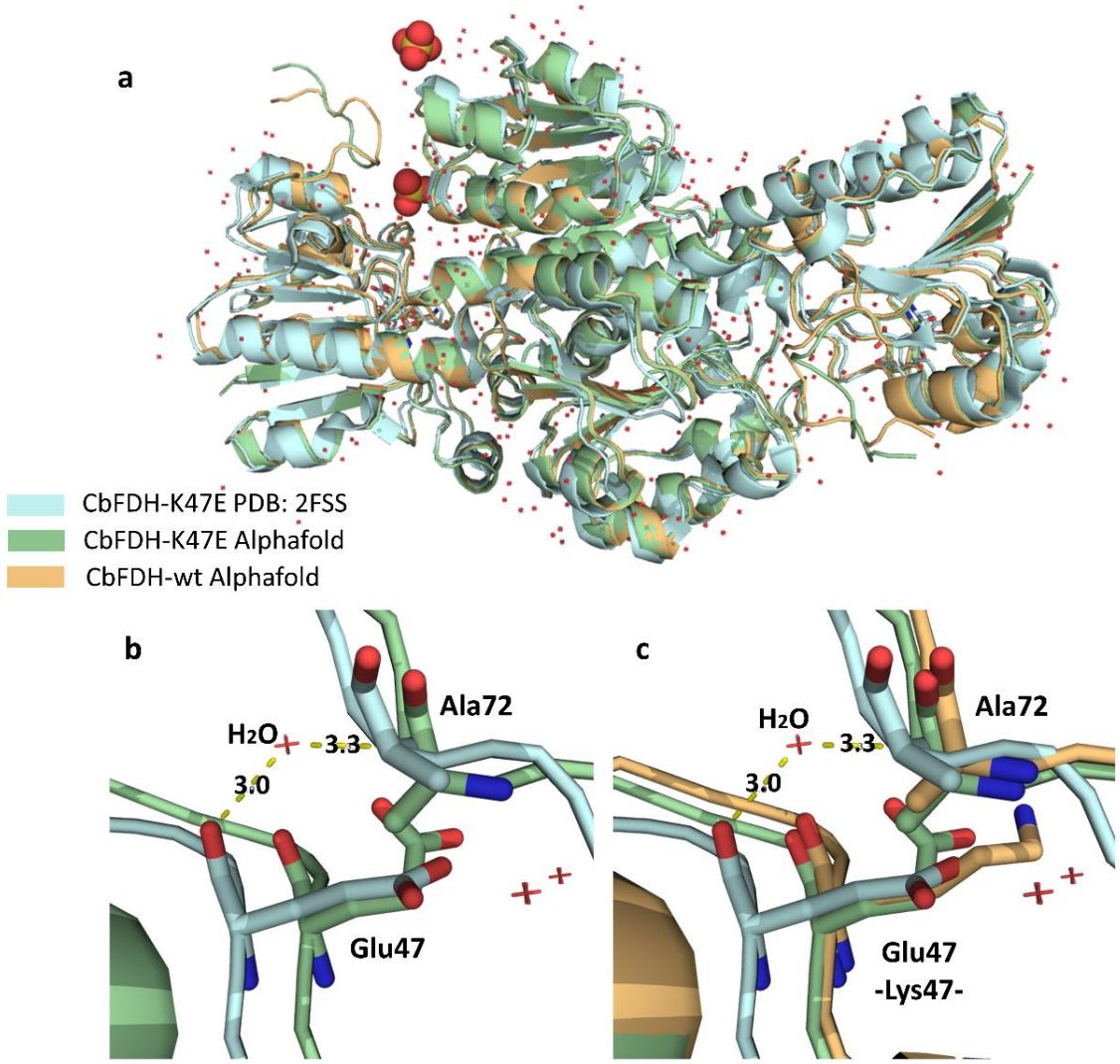

Figure 3.18: Alphafold2 prediction of CbFDH K47E mutant is compared with X-ray structure (PDB: 2FSS). RMSD of alignment was found as 1.018 RMSD between prediction and X-ray structures. But wt-CbFDH and K47E-CbFDH Alphafold2 prediction structures has high alignment with 0.251 RMSD. The mutant and wt-CbFDH Alphafold2 structures were colored as pale green and light orange respectively. 2FSS for mutant-CbFDH structure was colored as aquamarine. The H-bonds are demonstrated yellow dashed-lines.

Both structures were aligned and had 1.018 RMSD (Figure 3.18a). Moreover, mutation residue was observed with major conformational differences (Figure 3.18b). The E-47 residue and polar contacts were observed with 3.0 and 3.3A H-bond with a coordinated water molecule and Ala-72. However, Alphafold2 prediction did not

observed with those interactions. Also, conformation of mutant residue was significantly different. For understanding prediction efficiency for seconder structure alteration by mutation wt-CbFDH sequence was also modelled by Alphafold2, however the alignment with K47E mutant prediction was unexpectedly high as 0.251 RMSD. Moreover, the E-47 and K-47 residues was only conformational change because of having different side-chain group but the rest of the structure had high conformational similarity (Figure 3.18c).

Chapter 4:
DISCUSSION

The obtained 1.4 A resolution new apo-CbFDH structure was obtained with higher resolution and P1 unit cell rather than the previous apo-structure (PDB: 5DNA). The new structure provided a better electron cloud, especially in solvent exposed charged residues. Thus, new conformations with a 1.4 A resolution structure were observed in Lys12, Glu18, Lys19, Asp36 Asn51, Lys109, Asn147, Arg174, Glu213, Glu217 Lys246 Lys291, Glu325 residues (Figure 3.9). In addition, Glu332, located next to His311, which is an important residue for catalytic activity has better electron cloud which may provide valuable information for future biosynthetic and rational design applications.

Since these residues Arg and Lys are positive, Glu and Asp (Asn) negatively charged amino acids, the pH of the condition they are in may have caused this situation. The condition of the 1.4 A structure is pH 6.5, while the condition in which the old structure is solved is 7.5 pH. However, this pH range does not change the protonation states of charged or polar residues such as Lys and Arg. Thus, it can be suggested that pKa of neighbor residues such as His may alter its protonation state and cause different electrostatic interactions and residues conformations. On the other hand, in the precipitant species of the conditions, sodium acetate trihydrate is an organic and weak acid for the 5DNA structure, while sodium hydroxide is a strong base. There is also HEPES in the condition of 5DNA, but (morpholino) ethane sulfonic acid (MES) can be evaluated as a conjugate to HEPES with the sulfate it contains. Therefore, not only the better electron cloud, but also pH or the ionization potential of the condition may have created a difference in the conformations of these charged residues (Figure 3.9)

The fact that CbFDH holo- and apo-structures made GNM analysis possible to gain insight into the dynamics of the enzyme. In this analysis, we had the opportunity to examine the differences in inter-residual correlations and fluctuations, as well as the major difference (RMSD 1.968), especially with the tilts of the secondary structures, which arise when the structures of the enzyme are compared with alignment.

According to the GNM cross-correlation heat map, there are 3 domains that make up each monomer. Considering these domains, it is an expected result that the first (green) and second (yellow) domains have a high correlation in terms of proximity (Figure 13.2). The third domain is located close to the neighboring chain and contains a separate

Roseman-fold. This third domain's residues between 140-160th position, which gives interchain cross-correlation. Those residues that interacted with the residues of neighboring chain are in close position as expected from GNM results (data not shown).

For cross-correlation of residues with NAD+, azide and chloride ion, residues between 65-70th positions have an interestingly high cross-correlation, especially close to the azide molecule and do not have a direct polar contact with NAD+. Similarly, the residues in the 150-160st position in the neighboring chain gave a high correlation, even though they did not have a direct polar contact or proximity to NAD+. This may be the result of residual and inter-chain allosteric effects, needs further investigation. For interchain residues cross-correlation with NAD+, azide and chloride ion residues around 140-160th position has slight correlated motion with NAD+ C7 and azide N2 atoms makes those residues invaluable candidate for dimeric cooperativity of the enzyme that may arise by interchain correlated motions. However, those residues must investigate by mutational studies about dimerization and effect on catalytic activity to understand exact mechanism on cooperativity.

The decrease in the correlation in the first and second domains in the holo- form seen in the cross-correlation difference map and the increase in correlation of these domains with the third domain were confirmed by morph analysis. According to the morph analysis, the first and second domains approach to the third domain and the binding pocket is closed (Figure 3.12). This result proves that the correlation difference is seen as expected, because of closer position of all domains. However, the first and second domains do not diverge in binding to substrate and co-enzyme. In addition, the squared fluctuation of the residues belonging to those domains decreased, that shows that its mobility has decreased. In this case, flexible residues in apo-form may provide correlated motion, especially in the first and second domains, while this correlation and flexibility may decrease when transitioning to holo-form, leading to the closure of the binding domain. We can attribute as the apo-form may be more flexible domains to provide the necessary flexibility to the binding pocket opening for the entry of the substrate and the large NAD+ coenzyme, and to the fact that the holo-form substrate and co-enzyme are in a certain conformation.

Finally, the prediction efficiency of Alphafold2, which has been popular in recent years, on single-residue mutants has been investigated. As can be seen from the Figure 3.18, Alphafold2 estimates were made by creating a solid carbon backbone for both wt-

and K47E-CbFDH, giving only the side-chains of the residual differences in random conformations. In addition, when the K47E-CbFDH structure (PDB: 2FSS) and its prediction were compared, both the structure was generally not aligned (RMSD 1.018) and the conformation of mutant residues was different. At the same time, the presence of H-bond interactions and coordinated water for Glu47 residue in 2FSS could not be seen in Alphafold2 prediction. This shows that although Alphafold2 predictions are an important tool for estimating the secondary structures of various unstructured proteins, they are not sufficient to show important details especially for enzyme and binding structural studies such as interaction, coordinated water molecules or mutation alteration.

For future studies, the high-resolution structure we obtained, and the simple mechanism of the enzyme show that CbFDH is a candidate protein for real-time imaging of the mechanism by time-resolved serial X-ray crystallography studies. In addition, the data of mutant-CbFDHs, whose diffraction data cannot be collected even though their crystals can be obtained due to technical limitations, will be collected as soon as possible, and a solution can be found at the molecular level to the questions about how these mutants increase enzymatic activity and thermal stability by affecting their properties such as allostery and cooperativity on the enzyme. That invaluable information may pave a way to design more stable proteins in extreme condition or increase half-life of protein for commercializing and active enzymes to cost effective formate detection application for methanol toxicity and vitamin B deficiency.

Chapter 5:

CONCLUSION

1.4 A new wt-CbFDH structure reveal new conformational changes on charged amino acid with novel and better electron clouds obtained. Conformational changes may be caused by difference of protonation state neighbor residues by pH but this suggestion requires further investigation for protonation state of those amino acids by computational aspects.

GNM analysis and morph provide a new insight for substrate and co-enzyme binding. Major conformational changes were observed between apo- and holo-CbFDH and each monomer of CbFDH has major 3 domains about correlated motion. While first and second domain (included in catalytic domain) has higher correlation between each other than third (binding) domain as expected. On the other hand, cross-correlation differences revealed that substrate and co-factor binding decrease correlated motion and flexibility within catalytic domain while increase correlation between catalytic and binding domains. According to cross-correlation of apo- and holo-state structures with NAD+ and azide molecules, residues around 140^{th} and 160^{th} position may provide cooperativity between dimer by interchain allosteric effect on binding and/or catalytic activity that can be investigated by mutation studies.

Finally, Alphafold prediction is not sufficient to observe mutations' effects on structure because the electrostatic interactions, allosteric effect of mutation and conformations still absent in Alphafold prediction.

BIBLIOGRAPHY

[Adams et al., 2010] Adams, P. D., Afonine, P. V., Bunkoczi, 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 Cryst. D*, 66, pp. 213-221.

[Alissandratos et al., 2013] Alissandratos, A., Kim, H. K., & Easton, C. J. (2013). Formate production through biocatalysis. *Bioengineered*, 4(5), 348–350.

[Amao, 2018] Amao Y. (2018). Formate dehydrogenase for CO2 utilization and its application. *Journal of CO2 Utilization*, 26, pp. 623-641, ISSN 2212-9820.

[An and Chen, 2016] An, L., and Chen, R. (2016). Direct formate fuel cells: A review. *Journal of Power Sources*, 320, 127–139.

[Anderson et al., 2002] Anderson RM., Bitterman KJ., Wood JG., Medvedik O., Cohen H., Lin SS., Manchester JK., Gordon JI. and Sinclair DA. (2002). Manipulation of a nuclear NAD+ salvage pathway delays aging without altering steady-state NAD+ levels. *J Biol Chem*., 277(21), 18881-90.

[Bakan et al., 2011] Bakan, A., Meireles, L. M. & Bahar, I. (2011). ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics*.

[Balzer et al., 2013] Balzer, G. J., Thakker, C., Bennett, G. N., & San, K. Y. (2013). Metabolic engineering of Escherichia coli to minimize byproduct formate and improving succinate productivity through increasing NADH availability by heterologous expression of NAD+-dependent formate dehydrogenase. *Metabolic engineering*, 20, 1-8.

[Bansal and Kundu, 2022] Bansal, S., and Kundu, B. (2022). Protein engineering: Methods and applications. *In Advances in Protein Molecular and Structural Biology Methods*, pp. 641-668. Academic Press.

[Bar-Even, 2016] Bar-Even A. (2016). Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways. *Biochemistry*, 55(28), 3851–3863.

[Barin et al., 2018] Barin, R., Biria, D., Rashid-Nadimi, S., & Asadollahi, M. A. (2018). Enzymatic CO2 reduction to formate by formate dehydrogenase from Candida boidinii coupling with direct electrochemical regeneration of NADH. *Journal of CO2 Utilization*, 28, 117-125002E.

[Belenky et al., 2007] Belenky P., Bogan KL. and Brenner C. (2007). NAD+ metabolism in health and disease. *Trends Biochem Sci*., 32(1), 12-9.

[Berríos -Rivera et al., 2002] Berríos -Rivera, S. J., Bennett, G. N., San, K. (2002). Metabolic Engineering of Escherichia coli: Increase of NADH Availability by Overexpressing an NAD+-Dependent Formate Dehydrogenase. *Metabolic Engineering*, 4(3), pp. 217-229, ISSN 1096-7176.

[Billington et al., 2006] Billington RA., Bruzzone S., De Flora A., Genazzani AA., Koch-Nolte F., Ziegler M. and Zocchi E. (2006). Emerging functions of extracellular pyridine nucleotides. *Mol Med*., 12(11-12), 324-7.

[Breen et al., 2006] Breen LT., Smyth LM., Yamboliev IA. and Mutafova-Yambolieva VN. (2006). Beta-NAD is a novel nucleotide released on stimulation of nerve terminals in human urinary bladder detrusor muscle. *Am J Physiol Renal Physiol*., 290(2), F486-95.

[Brosnan and Brosnan, 2016] Brosnan, M. E., and Brosnan, J. T. (2016). Formate: The Neglected Member of One-Carbon Metabolism. *Annual review of nutrition*, 36, pp. 369-388.

[Bulut et al., 2021] Bulut, H., Yuksel, B., Gul, M., Eren, M., Karatas, E., Kara, N., ... & Binay, B. (2021). Conserved amino acid residues that affect structural stability of candida boidinii formate dehydrogenase. *Applied Biochemistry and Biotechnology*, 193(2), 363-376.

[Calzadiaz-Ramirez and Meyer, 2022] Calzadiaz-Ramirez, L., and Meyer, A. S. (2022). Formate dehydrogenases for CO2 utilization. *Current Opinion in Biotechnology*, 73, 95-100.

[Carugo and Argos, 1997] Carugo O. and Argos P. (1997). NADP-dependent enzymes. I: Conserved stereochemistry of cofactor binding. *Proteins*., 28(1), 10-28

[Castillo et al., 2008] Castillo, R., Oliva, M., Marti, S., & Moliner, V. (2008). A theoretical study of the catalytic mechanism of formate dehydrogenase. *The Journal of Physical Chemistry B*, 112(32), pp. 10012-10022.

[Cedrone et al., 2000] Cedrone F., Ménez A. and Quéméneur E. (2000). Tailoring new enzyme functions by rational redesign. *Current Opinion in Structural Biology*, 10, 405–410.

[Chen, 2001] Chen, R. (2001). Enzyme engineering: rational redesign versus directed evolution. *Trends in Biotechnology*, 19, 13-14.

[Choe et al., 2014] Choe H, Joo JC, Cho DH, Kim MH, Lee SH, et al. (2014) Efficient CO2-Reducing Activity of NAD-Dependent Formate Dehydrogenase from Thiobacillus sp. KNK65MA for Formate Production from CO2 Gas. *PLOS ONE*, 9(7).

[Crable et al., 2011] Crable, B. R., Plugge, C. M., McInerney, M. J., & Stams, A. J. (2011). Formate formation and formate conversion in biological fuels production. *Enzyme research*, 532536.

[Crameri et al., 1998] Crameri A., Raillard S. A., Bermudez E. and Stemmer W. P. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature*, 391, 288-291.

[Curry, 2015] Curry S. (2015) Structural Biology: A Century-long Journeyinto an Unseen World, *Interdisciplinary Science Reviews*, 40:3, pp. 308-328.

[Dawson, 1985] Dawson R.B. (1985). Data for biochemical research Oxford. Clarendon press., 122.

[DeLano, 2020] DeLano, W. L. (2020) The PyMOL Molecular Graphics System, Version 2.3, Schrödinger LLC.

[Dobson, 2019] Dobson, C. M. (2019). Biophysical Techniques in Structural Biology. *Annual Review of Biochemistry*, 88(1).

[Drenth, 2007] Drenth, J. (2007). Principles of protein X-ray crystallography. pp. 1-8 Springer Science & Business Media

[Du et al, 2022] Du, C., Li, Y., Xiang, R., Yuan, W. (2022) Formate Dehydrogenase Improves the Resistance to Formic Acid and Acetic Acid Simultaneously in Saccharomyces cerevisiae. *Int. J. Mol. Sci*., 23, 3406.

[Durnin et al., 2012] Durnin L., Dai Y., Aiba I., Shuttleworth CW., Yamboliev IA. and Mutafova-Yambolieva VN. (2012). Release, neuronal effects and removal of extracellular β-nicotinamide adenine dinucleotide (β-NAD^+) in the rat brain. *Eur J Neurosci*., 35(3), 423-35.

[Emsley and Cowtan, 2004] Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr*.

[Ertem et al., 2022] Ertem, F. B., Guven, O., Buyukdag, C., Gocenler, O., Ayan, E., Yuksel, B., ... & DeMirci, H. (2022). Protocol for structure determination of SARS-CoV-2 main protease at near-physiological-temperature by serial femtosecond crystallography. *STAR protocols*, 3(1), 101158.

[Farinas et al., 2001] Farinas ET., Bulter T. and Arnold FH. (2001). Directed enzyme evolution. *Current Opinion in Biotechnology*, 12 (6), 545-551.

[Fauque and Barton, 2012] Fauque, G. D., and Barton, L. L. (2012). Hemoproteins in Dissimilatory Sulfate- and Sulfur-Reducing Prokaryotes. *Advances in Microbial Physiology*, 60, 1–90.

[Felber, 2001] Felber S. (2001). Optimierung der NAD-abhängigen Formiatdehydrogenase aus Candida boidinii für den Einsatz in der Biokatalyse, Ph.D. Thesis, Düsseldorf University.

[Foster and Moat, 1980] Foster JW. and Moat AG. (1980). Nicotinamide adenine dinucleotide biosynthesis and pyridine nucleotide cycle metabolism in microbial systems. *Microbiol Rev*. 44(1), 83-105.

[Galkin et al., 2002] Galkin A.G., Kutsenko A.S., Bajulina N.P., Esipova N.G., Lamzin V.S., Mezentzev A.V., Shelukho D.V., Tikhonova T.V., Tishkov V.I., Ustinnikova T.B. and Popov V.O. (2002). *Biochim.Biophys.Acta*, 1594, 136.

[Getzoff et al., 1992] Getzoff E. D., Cabelli D. E., Fisher C. L., Parge H. E., Viezzoli M. S., Banci L. and Hallewell R. A. (1992). Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature*, 358, 347–351.

[Gopalakrishnan et al., 2019] Gopalakrishnan, B., Khanna, N., & Das, D. (2019). Dark-fermentative biohydrogen production. *In Biohydrogen* (pp. 79-122). Elsevier.

[Gulati and Poluri, 2016] Gulati, K., and Poluri, K.M. (2016) An overview of computational and experimental methods for designing novel proteins. *Recent Patents on Biotechnology*. 10: 235-263.

[Guo et al., 2016] Guo, Q., Gakhar, L., Wickersham, K., Francis, K., Vardi-Kilshtain, A., Major, D. T., Cheatum, C. M., & Kohen, A. (2016). Structural and Kinetic Studies of Formate Dehydrogenase from Candida boidinii. *Biochemistry*, 55(19), 2760–2771.

[Hanukoglu, 2015] Hanukoglu I. (2015). Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochem Mol Biol Educ*., 43(3), 206-9.

[Harden and Young, 1906] Harden A. and Young WC. (1906) The alcoholic ferment of Yeast juice Part II. *Proceedings of the royal Society of London*. Series B., 78 (526), 369-375.

[Henderson, 1983] Henderson LM. (1983). Niacin. *Annu Rev Nutr*., 3, 289-307.

[Hovda et al., 2005] Hovda K.E., Petter U., Dag J. (2005) Increased Serum Formate in the Diagnosis of Methanol Poisoning. *Journal of Analytical Toxicology*, 29(6), pp. 586–588.

[Ilari and Savino, 2008] Ilari, A., and Savino, C. (2008). Protein Structure Determination by X-Ray Crystallography. In: Keith, J.M. (eds) Bioinformatics. *Methods in Molecular Biology*™, vol 452. Humana Press.

[Jameson et al., 1989] Jameson DM., Thomas V. and Zhou DM. (1989). Time-resolved fluorescence studies on NADH bound to mitochondrial malate dehydrogenase. *Biochim Biophys Acta*., 994(2), 187-90.

[Jayathilake et al., 2019] Jayathilake, B. S., Bhattacharya, S., Vaidehi, N., & Narayanan, S. R. (2019). Efficient and selective electrochemically driven enzyme-catalyzed reduction of carbon dioxide to formate using formate dehydrogenase and an artificial cofactor. *Accounts of Chemical Research*, 52(3), 676-685

[Jiang et al., 2016] Jiang W., Lin P., Yang R. and Fang B. (2016). Identification of catalysis, substrate, and coenzyme binding sites and improvement catalytic efficiency of formate dehydrogenase from Candida boidinii. *Applied Microbiology and Biotechnology*, 100, 8425-37.

[Katoh et al., 2006] Katoh A., Uenohara K., Akita M. and Hashimoto T. (2006) Early steps in the biosynthesis of NAD in Arabidopsis start with aspartate and occur in the plastid. *Plant Physiol.*, 141(3), 851-7.

[Kerns et al., 2002] Kerns W. II, Christian Tomaszewski, Kenneth McMartin, Marsha Ford, Jeffrey Brent & META Study Group (2002) Formate Kinetics in Methanol Poisoning, *Journal of Toxicology: Clinical Toxicology*, 40(2), pp. 137-143.

[Koch-Nolte et al., 2011] Koch-Nolte F., Fischer S., Haag F. and Ziegler M. (2011). Compartmentation of NAD+-dependent signalling. *FEBS Lett*., 6, 585(11), 1651-6.

[Labrou and Rigden, 2001] Labrou, N. E., and Rigden, D. J. (2001). Active-site characterization of Candida boidinii formate dehydrogenase. *Biochemical Journal*, 354(2), 455–463.

[Lakowicz et al., 1992] Lakowicz JR., Szmacinski H., Nowaczyk K. and Johnson ML. (1992). Fluorescence lifetime imaging of free and protein-bound NADH. *Proc Natl Acad Sci.*, 89(4), 1271-5.

[Lamarre et al., 2013] Lamarre, S., Morrow, G., Macmillan, L., Brosnan, M. & Brosnan, J. (2013). Formate: an essential metabolite, a biomarker, or more?. *Clinical Chemistry and Laboratory Medicine*, 51(3), 571-578.

[Leisola and Turunen, 2007] Leisola M. and Turunen O. (2007). Protein engineering: opportunities and challenges. *Appl Microbiol Biotechnol*., 75(6), 1225-32.

[Lin and Guarente, 2003] Lin SJ. and Guarente L. (2003). Nicotinamide adenine dinucleotide, a metabolic regulator of transcription, longevity and disease. *Curr. Opin. Cell Biol.*, 15(2), 241-6.

[Ljungdahl, 1980] Ljungdahl, L. G. (1980). Formate dehydrogenases: role of molybdenum, tungsten and selenium. *Molybdenum and Molybdenum-Containing Enzymes*, 463-486.)

[Makar et al., 1975] Makar, A. B., McMartin, K. E., Palese, M., & Tephly, T. R. (1975). Formate assay in body fluids: application in methanol poisoning. *Biochemical medicine*, 13(2), 117–126.

[Martin-Amat et al., 1978] Martin-Amat, G., McMartin, K. E., Hayreh, S. S., Hayreh, M. S., & Tephly, T. R. (1978). Methanol poisoning: ocular toxicity produced by formate. *Toxicology and applied pharmacology*, 45(1), pp. 201–208.

[Maveyraud and Mourey, 2020] Maveyraud, L., and Mourey, L. (2020). Protein X-ray crystallography and drug discovery. *Molecules*, 25(5), 1030.)

[McCoy et al.,] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr*.

[Mutafova-Yambolieva et al., 2007] Mutafova-Yambolieva VN., Hwang SJ., Hao X., Chen H., Zhu MX., Wood JD., Ward SM. and Sanders KM. (2007). Beta-nicotinamide adenine dinucleotide is an inhibitory neurotransmitter in visceral smooth muscle. *Proc Natl Acad Sci*, 9, 104(41), 16359-64.

[Nielsen et al., 2019] Nielsen, C. F., Lange, L., & Meyer, A. S. (2019). Classification and enzyme kinetics of formate dehydrogenases for biomanufacturing via CO2 utilization. *Biotechnology advances*. 37(7), 107408.

[Niks and Hille, 2018] Niks D, and Hille R. (2018). Reductive activation of CO2 by formate dehydrogenases. *Methods in Enzymology*. 613, pp. 277-295.

[Nitta et al., 2018] Nitta, R., Imasaki, T., Nitta, E. (2018). Recent progress in structural biology: lessons from our research history. *Microscopy*, 67(4), pp. 187–195.

[Oizel, 2020] Oizel, K., Tait-Mulder, J., Fernandez-de-Cossio-Diaz, J. et al. (2020). Formate induces a metabolic switch in nucleotide and energy metabolism. *Cell Death Dis.*, 11, 310.

[Parikh and Matsumura, 2005] Parikh M. R. and Matsumura I. (2005). Site-saturation mutagenesis is more efficient than DNA shuffling for the directed evolution of β-Fucosidase from β- Galactosidase. *J. Mol. Biol*., 352, 621-628.

[Pietzke et al., 2020] Pietzke, M., Meiser, J., & Vazquez, A. (2020). Formate metabolism in health and disease. *Molecular metabolism*, 33, 23–37.

[Pinske and Sawers, 2016] Pinske, C., and Sawers, R. G. (2016). Anaerobic formate and hydrogen metabolism. *EcoSal Plus*, 7(1).

[Pitela et al., 2007] Pitela S. B., Chob C. M., Chen W. and Zhao H. (2007). Chapter 3. Directed Evolution Tools in Bioproduct and Bioprocess Development. *Bioprocessing for Value-Added Products from Renewable Resources*, pp. 49-72.

[Pollak et al., 2007] Pollak N, Dölle C. and Ziegler M. (2007). The power to reduce: pyridine nucleotides--small molecules with a multitude of functions. *Biochem J*., 1, 402(2), 205-18.

[Que´me´ neur et al., 1998] Que´me´ neur E., Moutiez M., Charbonnier J. B. and Me´ nez A. (1998). Engineering cyclophilin into a proline-specific endopeptidase. *Nature*, 391, 301–304.

[Rao and Rossmann, 1973] Rao ST. and Rossmann MG. (1973). Comparison of super-secondary structures in proteins. *J Mol Biol.* 15, 76(2), 241-56.

[Rojkova et al., 1999] Rojkova, A. M., Galkin, A. G., Kulakova, L. B., Serov, A. E., Savitsky, P. A., Fedorchuk, V. V., & Tishkov, V. I. (1999). Bacterial formate dehydrogenase. Increasing the enzyme thermal stability by hydrophobization of alpha-helices. *FEBS letters*, 445(1), 183–188.

[Sato and Amao, 2020] Sato, R., and Amao, Y. (2020). Can formate dehydrogenase from Candida boidinii catalytically reduce carbon dioxide, bicarbonate, or carbonate to formate?. *New Journal of Chemistry*, 44(28), 11922-11926.

[Sawer et al., 2004] Sawers, R. G., Blokesch, M., & Böck, A. (2004). Anaerobic formate and hydrogen metabolism. *EcoSal Plus*, 1(1).

[Schink et al., 2017] Schink, B., Montag, D., Keller, A., & Müller, N. (2017). Hydrogen or formate: Alternative key players in methanogenic degradation. *Environmental Microbiology Reports*, 9(3), pp. 189–202.

[Schuchmann and Müller, 2013] Schuchmann, K., and Müller, V. (2013). Direct and reversible hydrogenation of CO2 to formate by a bacterial carbon dioxide reductase. *Science*, 342(6164), pp. 1382–1385. ,

[Singh et al., 2013] Singh, R. K., Tiwari, M. K., Singh, R., Lee J. (2013). From Protein Engineering to Immobilization: Promising Strategies for the Upgrade of Industrial Enzymes. *International Journal of Molecular Sciences*, 14(1), pp. 1232-1277.

[Singh et al., 2018] Singh, R. K., Lee, J. K., Selvaraj, C., Singh, R., Li, J., Kim, S. Y., & Kalia, V. C. (2018). Protein Engineering Approaches in the Post-Genomic Era. *Current protein & peptide science*, 19(1), pp. 5–15.

[Slusarczyk et al., 2000] Slusarczyk, H., Felber, S., Kula, M. R., & Pohl, M. (2000). Stabilization of NAD-dependent formate dehydrogenase from Candida boidinii by site-directed mutagenesis of cysteine residues. *European Journal of Biochemistry*, 267(5), pp. 1280-1289.

[Smyth et al., 2004] Smyth LM., Bobalova J., Mendoza MG., Lew C. and Mutafova-Yambolieva VN. (2004). Release of beta-nicotinamide adenine dinucleotide upon stimulation of postganglionic nerve terminals in blood vessels and urinary bladder. *J Biol Chem.,* 19, 279(47), 48893-903.

[Spadiut et al., 2008] Spadiut O., Pisanelli I., Maischberger T., Peterbauer C., Gorton L., Chaiyen P. and Haltrich D. (2008). Engineering of pyranose 2-oxidase: Improvement for biofuel cell and food applications through semi-rational protein design. *Journal of Biotechnology*, 139, 250–257.

[Stemmer, 1994] Stemmer W. P. (1994). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci*., 91, pp. 1074-1075.

[Subramaniam and Kleywegt, 2022] Subramaniam, S., and Kleywegt, G.J. (2022). A paradigm shift in structural biology. *Nat Methods*, 19, pp. 20–23.

[Thomas et al., 2016] Thomas, S.C., Alhasawi, A., Auger, C. et al. (2016). The role of formate in combatting oxidative stress. *Antonie van Leeuwenhoek*, 109, pp. 263–271.

[Tishkov et al., 1993] Tishkov V.I., Galkin A.G., Marchenko G.M., Egorova O.A., Sheluho D.V., Kulakova L.B., Dementieva L.A. and Egorov A.M. (1993). Catalytic properties and stability of a Pseudomonas sp.101 formate dehydrogenase mutants containing Cys-255-Ser and Cys-255-Met replacements. *Biochem. Biophys. Res. Commun.,* 192, 976.

[Torres et al., 1999] Torres, R. A., Schiøtt, B., & Bruice, T. C. (1999). Molecular dynamics simulations of ground and transition states for the hydride transfer from formate to NAD+ in the active site of formate dehydrogenase. *Journal of the American Chemical Society*, 121(36), pp. 8164-8173.

[Unden and Bongaerts, 1997] Unden G. and Bongaerts J. (1997). Alternative respiratory pathways of Escherichia coli: energetics and transcriptional regulation in response to electron acceptors. *Biochim Biophys Acta*., 1320(3), 217-34.

[VanHook, 2022] VanHook A. M. (2022). Formate for tumor progression. *Science signaling*, 15(736).

[Williams et al., 2004] Williams G. J., Nelson A. S. and Berry, A. (2004). Directed evolution of enzymes for biocatalysis and the life sciences. *Cell. Mol. Life Sci.*, 61, pp. 3034–3046.

[Wong et al., 2004] Wong T. S., Tee K. L., Hauer, B. and Schwaneberg, U. (2004). Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Res*, 56.

[Yoshikuni et al., 2006] Yoshikuni Y., Ferrin T. E. and Keasling J. D. (2006). Designed divergent evolution of enzyme function. *Nature*, 440, pp. 1078–1082.

[Yuan et al., 2005] Yuan L., Kurek I., English J. and Keenan R. (2005). Laboratory-Directed Protein Evolution. *Microbiology and Molecular Biology Reviews*, 69, pp. 373–392.

[Zeymer and Hilvert, 2018] Zeymer, C., and Hilvert, D. (2018). Directed Evolution of Protein Catalysts. *Annual Review of Biochemistry*, 87(1), pp. 131–157.

[Zhang et al., 2021] Zhang, X., Shao, W., Chen, B., & Wang, M. (2021). Cross-linking of carbonic anhydrase and formate dehydrogenase based on amino acid specific recognition: Conversion of carbon dioxide to formic acid. *Enzyme and Microbial Technology*, 146, 109763.

[Ziegler and Niere, 2004] Ziegler M. and Niere M. (2004). NAD+ surfaces again. *Biochem J*., 15, 382, e5-6.