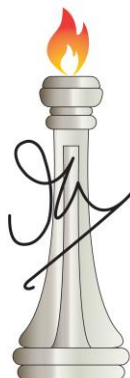

Catalytic mechanism and stereospecificity in class-I fumarate hydratase and L-tartrate dehydratase, and structural characterization of an archaeal IMP dehydrogenase

A Thesis Submitted for the Award of the Degree of

Doctor of Philosophy

By

ASUTOSH BELLUR



J N C A S R

Thesis Supervisor: Prof. Hemalatha Balaram

Molecular Biology and Genetics Unit

Jawaharlaral Nehru Centre for Advanced Scientific Research

(A Deemed University)

Bangalore – 560064, India

May 2022

(This page is intentionally left blank)

DECLARATION

I hereby declare that the matter embodied in this thesis entitled “**Catalytic mechanism and stereospecificity in class-I fumarate hydratase and L-tartrate dehydratase, and structural characterization of an archaeal IMP dehydrogenase**” is the result of the research carried out by me at the Molecular biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India, under the supervision of Prof. Hemalatha Balaram and it has not been submitted elsewhere for the award of any other degree.

In keeping with the general practice of reporting scientific observations, due acknowledgements have been made wherever the work described is based on the findings of other investigators. Any omission which might have occurred by oversight or error in judgement is regretted.



Asutosh Bellur

Place: Bangalore

Date: 18th July 2022

(This page is intentionally left blank)



Hemalatha Balaram, Ph. D.

Professor

CERTIFICATE

I hereby certify that the matter embodied in this thesis entitled “**Catalytic mechanism and stereospecificity in class-I fumarate hydratase and L-tartrate dehydratase, and structural characterization of an archaeal IMP dehydrogenase**” has been carried out by Mr. Asutosh Bellur at the Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India under my supervision and that the results presented in this thesis have not previously formed the basis for the award of any other diploma, degree or fellowship.

Prof. Hemalatha Balaram

(Research Supervisor)

Place: Bangalore

Date: 18th July 2022

(This page is intentionally left blank)

Acknowledgements

The meaning of life is that it is to be lived – Bruce lee

During the course of my Ph.D., I have figured out my place in this world and how I am supposed to live it. Answers to who I am and what I want to be were addressed during my stay here at JNCASR with a large sect of people who understood me without pretention, had little prejudice, and loved me unconditionally for who I am. I am truly indebted to them forever.

The key aspect of a good mentor is to help students become more of who they already are. This very well holds true for my research supervisor Prof. Hemalatha Balaram, who has nurtured my scientific curiosities all throughout my PhD endeavor. I cannot thank her enough for letting me be.

There is only so far an individual can go by oneself. With the right company the distance traversed amplifies. With the collaborations we resorted to, the stretch seemed limitless. My wholehearted gratitude to Prof. Balaram, Prof. Balasubramaniam Sundaram, Dr. Garima Jindal, Soumik, Sudarshan, and Ratul. These people made it possible to build on and complete a story I have been ever so deeply enchanted by.

I feel privileged to have gotten the opportunity to work on protein X-ray crystallography, a technique I will continue to cherish all my life. A special thanks to Prof. B Gopal who was my comprehensive examiner and has always given us machine time for the X-ray diffractometer at the Indian Institute of Science. His constructive criticism has been valuable in me learning the technique better.

Faculty at MBGU have been helpful providing me insights on my work year after year at the work presentations. My sincere thanks to Prof. Anuranjan Anand, Prof. Ranga Uday Kumar, Prof. MRS Rao, Prof. Tapas Kundu, Prof. Maneesha Inamdar, Prof. Kaustuv Sanyal, Prof. Namita Surolia, Dr. Ravi Manjithaya, Dr. James Chelliah, Dr. Kushagra Bansal, and Prof. Sheeba Vasu for letting students learn and grow not just as researchers but also as good presenters.

To where I stand and all that I am today, my teachers have played an impeccable role for which I could never be thankful enough and owe deeply. I thank Dr. N. S. Devaki, Mr. Madan Jagannatha, Sathish Manjunath, Shaurya Kaushal, Chakradhar Thantanapally, Dr. Lakshmeesha K. Nagappa, and Dr. Vijay Jayaraman immensely for shaping my character.

I got the opportunity to interact with the best seniors possible not just at the laboratory but the entire department of MBGU. The essence and flavor of academia has been handed out to me by them in a way I continue to savor the taste to date. I thank Dr. Vijay Jayaraman, Dr. Sanjeev Kumar, Dr. Saurav Roy, Dr. Aparna Dongre, Dr. Arpit Shukla, Dr. Santosh Shivakumaraswamy, Dr. Lakshmeesha K. Nagappa, and Dr. Lakshmi Prasoon for being such wonderful seniors. Current and past lab members, MS students, rotation students and internship students have always created an environment of joy at the workplace. Lab never felt like a workplace. Arpitha, Resmi, Anusha, Nivedita, Chitraleka, Dr. Dhanalakshmi and Bala have tolerated me and loved me like family. Right from the time I joined the laboratory, rotation students or summer trainees have always worked with me. With the interactions I have shared with them I have learnt how hard it is

to teach. These are the people who have made me aspire to be a good teacher. Sarika, Neelakshi, Akshay, Aashisha, Pragya, Kavya, Srijana, and Rohit deserve a mention.

Life was always easy and fun at JNC thanks to these beautiful people who shared similar interests with me and equal passion on topics in and out of science. Aditya Mahadevan, Lakshmi Shreekumar, Avani Mital, Akshay Chandran, Vybhav Yalasi, Divya Chalapathi, Arun Panchapakesan, Anirudha Mirmira, Srijana Dutta and Harshdeep Kaur are some of the people I have thoroughly enjoyed hanging around with.

I would like to thank the staff of X-ray facility at MBU, IISc; transmission electron microscopy imaging, ICMS, JNCASR and mass spectrometry facility, MBU, IISc and JNCASR for extending their support every time we used the facility.

People at Library, Admin section, Academic section, Hostel office, Security, Housekeeping and Dhanavantari are absolutely selfless people. These are the people who run the show so that our life is made easier than it really is. I sincerely thank JNCASR community for providing a conducive scientific environment.

The thing I learnt most from my family is to look at one and all as my very family, empathize with each person equally and most importantly be a human being with human virtues. I am humbled by the fact that as a matter of sheer probability I happened to be a part of them. I express my gratitude to my dad who always emphasized on the importance of “education”.

Synopsis of the thesis entitled

“Catalytic mechanism and stereospecificity in class-I fumarate hydratase and L-tartrate dehydratase, and structural characterization of an archaeal IMP dehydrogenase”

Submitted by

Asutosh Bellur

Molecular Biology and Genetics unit (MBGU), Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Bangalore, India

Thesis supervisor: Prof. Hemalatha Balaram

The thesis deals with 3 enzymes, class-I fumarate hydratase, L-tartrate dehydratase, and inosine-5' monophosphate dehydrogenase, wherein issues of catalytic mechanism, substrate specificity and regulation for the three enzymes are presented in 3 chapters. Each chapter is further divided into 5 sections wherein the first section introduces the topic and concludes by laying down the objectives of the study. This is followed by a section that provides details of the methodology used. The final three sections provide results and discussion concluding with an overall perspective of the study along with future directions. Across the 3 chapters, the thesis has an overall of 15 sections. **Fumarate hydratase** (fumarase, FH, EC 4.2.1.2) catalyzes the stereospecific reversible conversion of fumarate to malate through an *anti*-addition of water molecule to fumarate to form *S*-malate (Jin & Hanefeld, 2011). The conversion of fumarate to malate in aqueous medium has an extremely high activation energy of 36 kcal/mol (Bearne & Wolfenden, 1995). FH are distinctly divided into two classes, class-I and class-II, based on the presence or absence of a [4Fe-4S] cluster (Woods et al., 1988). Both classes of FH are ubiquitous across all kingdoms of life. While, class-II enzymes are of predominant occurrence in eukaryotes, class-I enzymes are widely distributed across archaea and prokaryotes with sparse representation in eukaryotes (Jayaraman et al., 2018). These two enzymes share absolutely no sequence or structural similarity, yet catalyze the same reaction with close to equal efficiencies with a rate enhancement of the order of 10^{15} fold compared to the non-enzymatic reaction at 37°C (Bearne & Wolfenden, 1995; Kronen & Berg, 2015). Many eukaryotic unicellular pathogens including *Plasmodium falciparum* have been found to possess

the class-I type FH that is essential for the organism. However, because of its instability and high sensitivity to oxygen, it is difficult to carry out structural studies on FH from *P. falciparum*. In our study FH from *Methanocaldococcus jannaschii* (Mj) has been used as a model system to understand all class-I fumarate hydratases. With the present available studies on the catalytic mechanism of class I FH still in its nascent state, our understanding is majorly derived from studies carried out on class II FH. Though the biochemical equilibrium of the reaction catalyzed by FH is in the direction of fumarate to malate ($K_{eq} = 4$), the problem has been viewed in the light of dehydration rather than hydration. A serine residue has been singled out as the catalytic base in class-II FH, where it is believed to form an oxyanion hole stabilized by interaction with positively charged neighbors. Extensive kinetic isotope experiments on class-II FH have revealed that the rate limiting step of the reaction does not involve a carbon-hydrogen bond breakage or formation and is proposed to be either the product release or proton exchange alongside a slow conformational isomerization (Rose et al., 1993; Bearne & Wolfenden, 1997). An alternative mechanism that explains the series of covalent changes that take place during the reaction of an enzyme bound substrate is the transition state theory (Porter & Bright, 1980). Rate enhancements in enzymes take place through reducing the activation barrier of the reaction by binding strongly to altered substrate in its transition state compared to the substrate in its ground state (Bearne & Wolfenden, 1995). FH stabilizes this altered substrate in its transition state by a reduction in free energy of at least 30 kcal mol^{-1} with a release in enthalpy of $-24 \text{ kcal mol}^{-1}$ and a gain in entropy of 19 kcal mol^{-1} . It is observed that 3-carbanion intermediates bind FH very strongly, which could reflect the action of a transition state (Porter & Bright, 1980). It is suggested that large gain in entropy is accompanied along with electronic stabilization of the aci-carboxylate/carbanionic intermediate, which acquires greater negative charge as it approaches its transition state. Transition state stabilization is expected to occur by maximizing enzyme-substrate interactions such as hydrogen-bonding, electrostatic and hydrophobic interaction. Thus, it may be a cumulative effect of the active site residues interacting with the substrate to raise it to its transition state that drives the reaction in these largely enthalpically (Wolfenden, 2006) and entropically driven reactions rather than conventional acid/base mediated catalysis. **Chapter one** of the thesis introduces the reader to the literature review on the understanding of the mechanism of catalysis in both class-I and class-II FH and our efforts at understanding the same in class-I FH, using *M. jannaschii* FH as a model system. On this front, we have solved the apo-protein structure of MjFH and with the

already available active site architecture with bound substrate obtained from *Leishmania major* (Lm) FH, we have carried out extensive site-directed mutagenesis of all the active site residues in MjFH. Interestingly, we observe that mutation of the previously predicted catalytic residues did not lead to complete loss of activity. Mutating the catalytic base threonine to alanine led to 196- and 235-fold drop in catalytic efficiency for malate and fumarate, respectively as substrates, while mutating the catalytic acid aspartate to valine showed a 37- and 28-fold reduction for the two substrates. Surprisingly though, mutating the residues forming a hydrogen-bonded network with the carboxylate groups of the substrate led to complete abolishment in enzyme activity. This data is in concordance with the knowledge that transition state stabilization in the fumarase reaction is driven by both enthalpic and entropic factors ($\Delta H = -24 \text{ kcal mol}^{-1}$ and $\Delta S = 19 \text{ kcal mol}^{-1} \text{ K}^{-1}$) (Bearne & Wolfenden, 1995). The large enthalpic stabilization must arise from the electrostatic interactions of the tri-negative aci-carboxylate anion at the substrate binding site, while the positive entropic gain can arise from the liberation of the water solvent shell around the fumarate dianion upon binding to the enzyme active site. Application of both DFT calculations and MD simulation studies in collaboration with Dr. Garima Jindal, department of organic chemistry, IISc, has also revealed the importance of the residues that form electrostatic interactions with the substrate carboxylate group. Natural population analysis has also revealed that the substrate at the active site of the enzyme is charge polarized which could play a major role in catalysis by taking the substrate close to its transition state, thus allowing a better attack of the water molecule on the bound substrate. From the good correlation obtained here, it can be said that non-covalent interactions present at the active site of FH also have an essential part in making the enzyme more efficient towards raising the substrate to its transition state by polarizing it. We believe that the affinity of FH active site for the binding of the transition state far exceeds the affinity for the substrate, thus catalyzing the reaction. Fumarate hydratase is an example of a true Pauling enzyme.

Class-I FHs share very high similarity with another hydrolyase, **L-tartrate dehydratase** (L-TTD) both of which catalyze reactions of opposite stereoisomers, L/D-tartrate to oxaloacetate. While working on MjFH, we noticed that these two proteins, share such a high degree similarity that it in fact, makes it difficult to discern the two enzymes apart which has led to mis-annotation of the two proteins in the protein database. Although the two enzymes catalyze reactions of opposite stereoisomers, interestingly we found that they have very well conserved and identical active site residues. Comprehensive understanding of enzyme chemistry and catalysis is lacking in both class-

I FH and L-TTD. Biochemical characterization of two-subunit class-I FH has been carried out in only *Pelotomaculum thermopropionicum* (Shimoyama et al., 2007) and *Pyrococcus furiosus* (van Vugt-Lussenburg et al., 2009) thus far, while L-TTD has been biochemically characterized from *Pseudomonas putida* (Kelly & Scopes, 1986), *E. coli* (Reaney, Begg, et al., 2009) and a L-tartrate fermenting anaerobic bacteria (Schink, 1984). While, recently structures of class-I FH have been solved and efforts are underway to understand the role of catalytic residues in this class of enzymes, L-TTD has very limited studies with regard to structure and catalysis with only one structure being predicted from primary sequence using constraints derived from sequence co-evolution and modelling (Ovchinnikov et al., 2015). This structure does not have a bound 4Fe-4S cluster and even the three cysteine residues known to coordinate the cluster are not very well poised to accommodate it, making it difficult to consider the structure with confidence. Stereospecificity is one of the key properties of enzymes. By nature, active sites of proteins are chiral and hence prefer specific enantiomers of chiral substrates and inhibitors (Adamu et al., 2018). This aspect of the enzyme especially in hydrolases are of interest in pharmaceuticals and chemical industries as biocatalysts because of their potential in synthesis of pure enantiomers and chiral intermediates (Patel & Patel, 2010; Singh et al., 2016). Although the molecular basis of enantioselectivity in many enzymes has not been completely understood, studies targeted at enhancing stereospecificity in enzymes have been successful. This has been achieved in enzymes such as lipase (Ema et al., 2005; Reetz et al., 2001; Rotticci et al., 2001), halohydrin dehalogenase (Tang et al., 2003) and epoxide hydrolase (Rui et al., 2005) from multiple organisms. Although increase in stereospecificity can be achieved using single amino-acid substitutions, inversion in stereospecificity involves a more complicated procedure that involves exchange in position of substituents or a complete switch in catalytic residue location through multiple mutations. Such examples of enzyme engineering have been achieved through rational design and/or directed evolution in enzymes such as lipase (Koga et al., 2003), hydrolase (Van Loo et al., 2004), esterase (Bartsch et al., 2008; Ivancic et al., 2007) and decarboxylase (Terao et al., 2007) from different sources. A multiple sequence alignment of class-I FH and L-TTD revealed that both these enzymes share identical active site residues with all the residues involved in binding the substrate in case of FH being conserved in both these proteins. To the best of our knowledge this is the first example where two proteins catalyze reactions of opposite stereospecificity with an active site possessing identical residues. For a reaction to get catalyzed at the active site, according to the widely accepted

three point attachment model (TPA), the substrate stereoisomer has to be anchored at three separate points by the catalytic residues in the enzyme of which one of the interactions break for the opposite stereoisomer thus disallowing its binding/catalysis (Abrol & Sundaresan, 2002). In this example, the two sets of residues anchoring the carboxyl group of the substrate are the same. Fe-S cluster is the third anchoring point in the protein the orientation of which could confer the stereospecificity to the enzyme. Subtle variations during evolution that contributed to structural reorganizations must have caused the switch in stereospecificity in these proteins. There are no previous reports on protein cofactors acting as stereospecificity centres and studies aimed at evolving one protein to catalyze the opposite stereospecific reaction will throw more light on the molecular basis of enantioselectivity in these enzymes. Introduction to the understanding on the stereospecificity in enzymes and that of FH and L-TTD in particular are provided in **Chapter two** along with the studies directed at understanding the stereospecificity in these two classes of enzymes, class-I FH, and L-TTD. A phylogenetic tree with the aim of identifying conserved motifs between these two classes of proteins to distinguish them apart and their relationship with class-I FH has been constructed. L-TTD has been kinetically characterized for the first time. Opposite stereoisomers have been tested for binding/catalysis/inhibition for both MjFH and *E. coli* L-TTD. Both the enzymes were found to have exactly opposite stereoselectivity with L-TTD active on L-tartrate and D-malate while FH was active on D-tartrate and L-malate. All class-I FH have a highly conserved motif 'KGXGS' close to the active site. This motif is present as an 'AGGGC' motif in L-TTD. Several attempts at swapping the motif and also using binding energy guided mutations of MjFH for L-tartrate binding have been made and tested for switch in stereospecificity. Close inspection of the active site in class-I FH structure has also been carried out to affirm the speculation that subtle variations in residues at the active site must contribute to the difference in stereospecificity in these two enzymes.

Inosine-5'-monophosphate dehydrogenase (IMPDH) (EC 1.1.1.205) catalyzes the first decisive step in the formation of guanine nucleotide, which is the oxidation of inosine 5'-monophosphate (IMP) to xanthosine 5'-monophosphate (XMP) with the concordant reduction of NAD^+ to NADH. This reaction is at the branch point of adenine and guanine nucleotide biosynthesis and is the rate limiting step in guanosine 5'-monophosphate (GMP) biosynthesis. GMP reductase (GMPR) catalyzes the NADPH dependent reduction of GMP to IMP and together with IMPDH constitutes

a family of proteins that control the guanine nucleotide pools in the cell and thus regulate many cellular processes. The genes *guaB* coding for IMPDH and *guaC* coding for GMPR share high level of identity which has led to the misannotation of this family of genes. IMPDH and GMPR have a core catalytic domain made of $(\beta/\alpha)_8$ barrel and a regulatory cystathionine- β -synthase (CBS, also called Bateman domain) projecting out of the core domain. CBS domain is highly dispensable (Baykov et al., 2011; Ereño-Orbea et al., 2013; Ignoul & Eggermont, 2005). Both proteins largely exist as homotetramers and homooctamers with polymerization of IMPDH into filaments called cytoophidia being reported recently (Keppeke et al., 2018). Kinetic and structural studies have been carried out in great detail on *Tritrichomonas foetus* IMPDH and it serves as the prototype for demarcating protein features. The enzyme possesses five key structural motifs where substrate binding and catalysis take place. The catalytic loop carries the active site cysteine residue. Adjacent to the active site loop is the phosphate binding region that contributes six residues to the active site. A finger domain composed of twisted β -sheet interacts with the adjacent monomers in the octameric complex and has been observed to be a key player in transmission of allosteric signal from CBS to the catalytic domain (Buey et al., 2015). Around ten residues from the distal end of the finger domain constitute the mobile flap that moves in and out of the active site during catalysis. Finally, a C-terminal segment of the protein couples the active site to a monovalent cation. IMP binding site is contributed by residues from the cysteine loop, phosphate binding motif and mobile flap, and are identified to be invariant while NAD binding site residues contributed by catalytic loop and mobile flap are highly divergent (Hedstrom, 2009). All IMPDHs reported to date show a requirement of monovalent cation (preferably K^+) for optimal activity. Majority of IMPDHs show Michaelis-Menten (MM) kinetics for IMP and display substrate inhibition by NAD^+ . Of the more than 100 X-ray structures deposited in the RCSB database, more than half the structures have CBS domain deleted and only close to 10 structures have the CBS domain mapped going to show that this domain of the protein is highly disordered. Removal of this subdomain is reported to facilitate crystallization (Hedstrom, 2009). Allosteric regulation in IMPDH takes place through the CBS domain and this mode of regulation has been found to be significantly different in prokaryotes and eukaryotes where they evolved differently allowing adaptation to metabolic requirements of each organism. Prokaryotic IMPDHs have two canonical nucleotide binding sites in the CBS domain where only ATP binds, while IMPDHs from eukaryotes have an additional non-canonical nucleotide binding site and guanine nucleotides (GDP and GTP) also bind to the nucleotide

binding site apart from ATP (Buey et al., 2015). GMP has been reported to be a competitive inhibitor competing for IMP at the active site of both prokaryotic and eukaryotic IMPDH and ADP and AMP do not affect the catalytic activity significantly in both cases. Two classes of bacterial IMPDH have been reported, class-I are kept auto-inhibited *in-vitro* in the absence of nucleotides and require ATP to achieve full catalytic activity, while class-II are active in the absence of nucleotides and do not require ATP (Alexandre et al., 2015). In case of eukaryotic IMPDHs, ATP is not required for achieving full catalytic activity and is inhibited by GDP and GTP unlike the prokaryotic counterparts. Recently it has been proposed that a conformational switch exists in IMPDH where, bacterial IMPDH are kept inhibited in the cell in the absence of ATP while eukaryotic IMPDH are kept inhibited when bound to GDP/GTP (Buey et al., 2017). This inhibition is relieved by binding of ATP to the CBS domain, which in turn causes structural changes in the protein that allow flexibility between the catalytic and Bateman domains, which were stapled into an inhibited fixed conformation. GTP/GDP induced rigidity in both eukaryotic IMPDH and apoprotein of prokaryotic IMPDH is evident from the octameric *Pseudomonas aeruginosa* and *Asbhya gossippii* IMPDH structures where the octamers are around 90 Å in width along the 4-fold axis (Anthony et al., 2017; Labesse et al., 2013). Binding of ATP in both cases leads to a more relaxed and expanded structure of 110 Å width. Human IMPDH has been reported to form micron length fibers through association of octamers in the presence of ATP although the role of this association has not been understood (Anthony et al., 2017). The CBS, catalytic loop, mobile flap, and C-terminal segment are all found to display varying degree of disorder in all the structures solved to date making it difficult to study the structural basis of modulation in this enzyme. Modulation in archaeal IMPDH mediated by purine nucleotides is yet to be reported with biochemical and kinetic characterization still lacking in this class of organisms. To understand the modulation aspects of the enzyme from archaea and carry out its kinetic characterization, IMPDH from *Methanocaldococcus jannaschii* (MjIMPDH) has been cloned, purified, and characterized earlier in the laboratory. Although MjIMPDH shows higher sequence similarity with homologs from prokaryotes, the archaeal enzyme is sensitive to both purine nucleotides (AMP, ADP, GMP, GTP, but not ATP) a feature similar to that exhibited by eukaryotic counterparts. MjIMPDH displayed a very weak dependence on K⁺ ions for its activity, a feature that has never been observed in any IMPDH before. **Chapter three** provides a brief introduction on the catalysis in IMPDH and similarities/differences in modulatory aspects of the same in prokaryotes and eukaryotes along

with the efforts made at solving the structures of IMPDH full length and CBS domain deleted proteins to understand some of the aforementioned peculiar features of MjIMPDH. The structure of MjIMPDH TIM domain with the CBS deleted was solved. Transmission electron microscopy (TEM) imaging has also been carried out and this reveals an ATP mediated cytoophidia formation similar to what has been observed in the case of eukaryotic IMPDH. This confirms the kinetics results of MjIMPDH as well which clearly describes that MjIMPDH has features similar to eukaryotic rather than prokaryotic IMPDH. There has been reports of cytoophidia filaments formed by CTP synthase in archaea (Zhou et al., 2020). This is the first report of an archaeal IMPDH forming cytoophidial filaments.

List of publications

1. Structural basis for the hyperthermostability of an archaeal enzyme induced by succinimide formation. Dongre AV, Das S, Bellur A, Kumar S, Chandrashekarmath A, Karmakar T, Balaram P, Balasubramanian S, Balaram H. *Biophys J.* 2021, 120, 3732-3746.
2. Mechanistic insights from biochemistry, crystal structure and cross-linking mass spectrometry into the functioning of a two-subunit GMP synthetase. Santosh Shivakumaraswamy, Sanjeev Kumar, Asutosh Bellur, Satya Dev Polisetty, Hemalatha Balaram. (Manuscript under revision)
3. Catalysis through electrostatics mediated substrate polarization: A unique mechanism of catalysis in class-I fumarases. Asutosh Bellur *et al.*, (Manuscript under preparation).
4. Understanding stereospecificity in *Escherichia coli* L-tartrate dehydratase. Asutosh Bellur *et al.*, (Manuscript under preparation).
5. Biochemical, Kinetic, and structural characterization of inosine 5'-monophosphate dehydrogenase from hyperthermophilic archaeon, *Methanocaldococcus jannaschii*. Thota LP, Asutosh Bellur *et al.*, (Manuscript under preparation)



(This page is intentionally left blank)



List of Abbreviations

ACE: Acetylcholine esterase
ADP: Adenosine diphosphate
AMP: Adenosine monophosphate
ARL: Arginosuccinate lyase
ASL: Adenylosuccinate lyase
ATP: Adenosine triphosphate
BSA: Bovine serum albumin
CAN: Carbonic anhydrase
CBS: Cystathionine β -synthase
CCP: Cytochrome c peroxidase
CTD: C-terminal domain
CTPS: Cytidine triphosphate synthetase
DFT: Density functional theory
DHFR: Dihydrofolate reductase
DNA: Deoxyribonucleic acid
dNTPs: Deoxy nucleoside triphosphates
DTT: Dithio thritol
ESI: Electrospray ionization
FEP: Free energy perturbation
Fe-S: Iron-Sulphur
FH: Fumarate hydratase
FRT: Flippase recognition target
FUM: fumarase
GDP: Guanosine diphosphate
GMP: Guanosine 5'-monophosphate
GMPR: Guanosine 5'-monophosphate reductase
GTP: Guanosine triphosphate
HGT: Horizontal gene transfer
IMP: Inosine 5'-monophosphate
IMPDH: Inosine-5'-monophosphate dehydrogenase
IPRO: Iterative Protein Redesign and Optimization
IPTG: Isopropyl-thiogalactoside
kb - Kilobases
kDa- Kilo Dalton
KSI: Ketosteroid isomerase
LC: Liquid chromatography
LGK: Levo-glucosan kinase

Lm: *Leishmania major*
MCS: Multiple cloning sites
MD: Molecular dynamics
Mg: Magnesium
Mj: *Methanocaldococcus jannaschii*
MM: Michaelis-Menten
MR: Molecular replacement
MS: Mass spectrometry
MSA: Mercaptosuccinic acid
NAD: Nicotinamide adenine dinucleotide
NADP: Nicotinamide adenine dinucleotide phosphate
Ni-NTA: Nickel-nitrilotriacetic acid
NPA: Natural population analysis
NTD: N-terminal domain
OD: Optical density
PAGE: Polyacrylamide gel electrophoresis
PCR: Polymerase chain reaction
PEG: Polyethylene glycol
PEI: Polyethylenimine
Pf: *Plasmodium falciparum*
PMSF: Phenylmethyl sulfonyl fluoride
ppm: parts per million
QMMM: Quantum mechanics/Molecular mechanics
RMSD: Root-mean-square deviation
RT: Retention time
Rubisco: Ribulose-1,5-bisphosphate carboxylase oxygenase
SAD: Single wavelength anomalous diffraction
SDS: Sodium dodecyl sulphate
SOD: Superoxide dismutase
TB: Terrific broth
TCA: Tricarboxylic acid
TEM: Transmission electron microscopy
TI: Thermodynamic integration
TIM: Triosephosphate isomerase
TPA: Three point attachment
TTD: Tartrate dehydratase
WT: Wild type
XMP: Xanthosine 5'-monophosphate

List of figures

1.1. Distribution of enzyme kinetic parameters	3
1.2. Structure and domain organization in class-I and class-II FH.....	8
1.3. Mechanisms for dehydration of malate to fumarate	10
1.4. Proposed reaction mechanism for class-I FH.....	13
1.5. Active site of LmFH with bound L-malate (5l2r).....	14
1.6. MjFH protein crystal morphology.....	21
1.7. Kinetics of WT MjFH and IC ₅₀ plot for MSA on substrates malate and fumarate	24
1.8. Structure of MjFH.....	25
1.9. Differences between apo-MjFH and holo-LmFH (5l2r) structures	27
1.10. Active site comparison between apo-MjFH and holo-LmFH (5l2r).....	28
1.11. MD simulation reveals class-I FH adopts two different states based on presence or absence of a bound Fe-S cluster and substrate	30
1.12. Motif KGXGS is identical and conserved across all class-I FH.....	31
1.13. Circular dichroism spectra of apo- and Fe-S reconstituted wild type and mutants of MjFH.....	32
1.14. A conserved water molecule could potentially take over the role of catalytic base in the absence of the threonine residue	33
1.15. Substrate is charge-polarized at the active site with more charge-interactions on one side compared to other.....	34
1.16. Mutation of Gly147 to Ala will lead to the side chain methyl group coming in close contact with the carboxylate group of the bound substrate.....	35
1.17. Three pathways of interconversion between malate and fumarate in aqueous medium.....	36
1.18. Stepwise mechanism of reversible (de)hydration in the protein environment proceeding through the carbanion pathway.....	38
1.19. Plausible mechanistic pathway of reversible (de)hydration in T467A mutant (pathway M1) and D135A mutant (pathway M2).....	39
1.20. Contacts of Lys213 of the conserved KGXGS motif with Gln195 and Glu267 that interact with malonate bound at the tunnel in LmFH (5l2r).....	41
1.21. Electronic energy profiles obtained from different model systems	45
1.22. Active site architecture of class-II FH taken from <i>M. tuberculosis</i> FH reveals a larger number of H-bonds at one end of the substrate compared to other.....	47
1.23. Mechanism of enzyme catalysis in FH.....	47
1.24. MjFH enzyme activity on mesaconate	50
2.1. Multiple sequence alignment of class-I FH and L-TTD highlighting the active site residues involved in substrate binding and catalysis.....	56
2.2. Multiple sequence alignment of biochemically characterized class I FH and L-TTD.....	66

2.3. Phylogenetic tree (cladogram) for fumarate hydratase Class-I and L-tartrate dehydratase protein sequences.....	67
2.4. Cloning, expression, and purification of <i>E. coli</i> L-TTD and <i>fumA</i>	69
2.5. Plot of initial rate vs L-tartrate concentration using reconstituted L-TTD.....	70
2.6. Stereoisomers of tartrate.....	71
2.7. Substrate specificity of class-I FH.....	72
2.8. Substrate specificity of L-TTD.....	74
2.9. Total ion chromatogram (TIC) and mass spectra of fumarate and malate standards.....	75
2.10. Total ion chromatogram (TIC) and mass spectra of fumarate incubated with L-TTD.....	75
2.11. Swapping the residues in the conserved motif from L-TTD onto MjFH does not lead to switch in stereospecificity.....	77
2.12. L-tartrate docking to WT LmFH and in silico mutant design of LmFH.....	78
2.13. Comparison of class-I FH with modelled structures of L-TTD to gain insights into difference in substrate specificity between the two enzymes.....	80
2.14. L-tartrate, D-tartrate and meso-tartrate docked active site in the structure of LmFH.....	81
2.15. The three-point attachment model.....	85
2.16. Attempts at L-TTD structure solution.....	86
2.17. Genotyping of <i>E. coli</i> Δ L-TTD and Δ <i>fumACB</i> / Δ L-TTD strains.....	87
2.18. Phenotyping of <i>E. coli</i> Δ L-TTD and Δ <i>fumACB</i> / Δ L-TTD strain.....	88
2.19. Multiple sequence alignment of biochemically characterized class I FH and L-TTD.....	91
2.20. Pull down of <i>E. coli</i> <i>fumA</i> CTD with MjFHa subunit.....	91
3.1. Domain architecture and spatial arrangement of IMPDH.....	96
3.2. Mechanism of IMPDH reaction.....	97
3.3. MjIMPDH ^{ΔCBS} protein crystal morphology.....	101
3.4. Multiple sequence alignment of IMPDH from various organisms show that MjIMPDH has all the catalytic residues conserved.....	105
3.5. Multiple sequence alignment of Bateman domain of IMPDH from various organisms highlighting the non-canonical nucleotide binding site observed in eukaryotic IMPDH.....	106
3.6. MjIMPDH protein purification and MjIMPDH ^{ΔCBS} structure analysis.....	108
3.7. Negative-stain electron microscopy imaging of purified MjIMPDH.....	109
A.1. LC-ESI-MS of selenomethionine incorporated MjFH subunits.....	113
B.1. Representative time course curves for MjFH WT and active mutants on A. 1 mM malate and B. 500 μ M fumarate.....	114
C.1. Time course curves for conversion of OAA to L-malate by MDH in the presence of L-malate in the assay mixture (pH 7.4).....	115
D.1. The crystals of MjIMPDH obtained from multiple conditions were small and not of diffraction quality.....	117

List of tables

1.1. <i>List of eukaryotic unicellular organisms having class-I FH and the respective disease they cause in humans</i>	16
1.2. <i>List of oligonucleotide sequences used for cloning M. jannaschii WT FH and generating site-directed mutants</i>	17
1.3. <i>Data collection and refinement statistics of MjFH$\alpha\beta$ protein structures solved by Se-SAD</i>	22
1.4. <i>Kinetic parameters of MjFH WT and mutant enzymes</i>	32
1.5. <i>NPA charges on bound fumarate for different mutants of LmFH</i>	40
2.1. <i>Oligonucleotide sequences used for cloning E. coli WT L-TTDα and β genes and fumA in pET-Duet vector and generating site-directed mutants</i>	58
2.2. <i>Oligonucleotide sequences used for knockout of L-TTD orfz1 and orfz2 genes in WT BW25113 E. coli and ΔfumACB strain</i>	62
2.3. <i>List of substrates, inactive analogues and inhibitors for class-I, class-II FH, and L-TTD</i>	89
3.1. <i>Data collection and refinement statistics of MjIMPDHΔCBS protein structure</i>	102
3.2. <i>Comparison of sequence similarity and identity of MjIMPDH with IMPDH from organisms for which structures are available</i>	103
C.1. <i>Conditions under which full-length MjIMPDH crystals were obtained</i>	114
C.2. <i>Conditions in which MjIMPDHΔCBS crystals were obtained</i>	115



(This page is intentionally left blank)



Contents

Declaration.....	iii
Certificate.....	v
Acknowledgements.....	vii
Synopsis.....	ix
List of Publications.....	xvii
List of Abbreviations.....	xix
List of Figures.....	xxi
List of Tables.....	xxiii
Table of Contents.....	xxv

1. Chapter 1: Understanding the mechanism of catalysis in class-I fumarate hydratase

1.1. Introduction

1.1.1. Moderately efficient enzymes and exceptional enzymes. Why fumarate hydratase is exceptional.....	2
1.1.2. Evolutionary dynamics of enzyme catalytic mechanism and kinetic parameters.....	3
1.1.3. Common set of underlying physical principles in enzyme catalysis.....	5
1.1.4. Class-I and class-II FH are enzymes with different folds catalyzing the same reaction with equal efficiencies.....	7
1.1.5. Understanding of catalysis in class-II fumarate hydratase.....	8
1.1.5.1. Transition state stabilization.....	8
1.1.5.2. Steps involved in the reaction probed by kinetic isotope studies.....	9
1.1.5.3. Conventional acid/base catalysis.....	10
1.1.6. Understanding of catalysis in class-I fumarate hydratase.....	11
1.1.6.1. Fe-S cluster hydrolyases.....	11
1.1.6.2. L. major class-I FH.....	13
1.1.7. C-I and C-II FH show differences with regard to enzyme promiscuity.....	14
1.1.8. Physiological and clinical significance.....	15
1.1.9. Objectives.....	16

1.2. Experimental procedures

1.2.1. Chemicals, strains, and molecular biology reagents.....	17
1.2.2. Generation of plasmid constructs.....	17
1.2.3. Protein expression and purification.....	18
1.2.4. Reconstitution of Fe-S cluster and confirmation through CD spectroscopy.....	19
1.2.5. Enzyme activity.....	20
1.2.6. Crystallization and data collection.....	21
1.2.7. Structure determination and refinement.....	22

1.3. Results and discussion

1.3.1. Enzyme activity and kinetic characterization of MjFH.....	23
1.3.2. Structure of MjFH.....	24
1.3.2.1. Features of the enzyme structure.....	24
1.3.2.2. Global structural differences between Mj and LmFH.....	26
1.3.2.3. Differences in active site architecture between Mj and LmFH.....	28

1.3.2.4. Binding of cluster and substrate compacts quaternary fold.....	29
1.3.3. Mutational analysis of the active site residues.....	30
1.3.3.1. Catalytic acid and base mutants retain activity.....	32
1.3.3.2. Mutation of residues involved in electrostatic interactions with the substrate led to loss of activity.....	33
1.3.3.3. The conserved motif – KGXGS has a role in maintaining the environment around the Fe-S cluster.....	35
1.3.4. QMMM studies gives insights into the mechanism of catalysis in class-I FH.....	35
1.3.4.1. Fumarate hydration in aqueous solution proceeds through the carbanion intermediate.....	36
1.3.4.2. Reaction mechanism at the enzyme active site.....	37
1.3.4.3. Alternative mechanism in the absence of catalytic residues.....	38
1.3.5. Natural population analysis reveals that the enzyme is indeed charge polarized at the active site.....	39
1.4. Summary	
1.4.1. MjFH has two binding sites for substrates.....	41
1.4.2. Binding of Fe-S cluster and substrate compacts FH quaternary structure.....	42
1.4.3. Insights into catalytic mechanism of Class-I FH from site-directed mutants of MjFH.....	42
1.4.4. QM/MM supports the dominant role of substrate charge polarization in catalysis.....	43
1.4.5. Proposed mechanism of catalysis in class-I FH.....	46
1.4.6. Conclusion.....	48
1.5. Future directions	
2. Chapter 2: Understanding stereospsecificity in class-I FH and L-tartrate dehydratase	
2.1. Introduction	
2.1.1. Types of specificity offered by enzymes.....	52
2.1.2. How do enzymes deal with stereoisomers?.....	52
2.1.3. Significance of enzyme stereospecificity.....	53
2.1.4. Methodologies adapted to switch stereospecificity.....	54
2.1.5. Class-I FH and L-TTD, a unique case of stereospecificity.....	55
2.1.6. Objectives.....	56
2.2. Experimental procedures	
2.2.1. Chemicals, strains, and molecular biology reagents.....	57
2.2.2. Sequence alignment and phylogenetic tree construction.....	57
2.2.3. Cloning, protein expression and purification.....	58
2.2.4. Reconstitution of Fe-S cluster.....	60
2.2.5. Enzyme activity.....	60
2.2.6. Kinetics of inhibition by substrate analogs.....	61
2.2.7. Generation of L-TTD null <i>E. coli</i> strains and validation.....	61
2.2.8. Computational design for switching stereospecificity in class I FH to catalyze L-tartrate dehydratase activity.....	62
2.3. Results and discussion	
2.3.1. Multiple sequence alignment.....	64

2.3.2. Phylogenetic analysis.....	66
2.3.3. Cloning, expression, and purification.....	68
2.3.4. Kinetic characterization of L-TTD.....	70
2.3.5. Substrate specificity in C-I FH and L-TTD.....	71
2.3.5.1. Class-I FH substrate specificity.....	71
2.3.5.2. L-TTD substrate specificity.....	73
2.3.6. Motif swap to switch specificity in class-I FH and L-TTD.....	76
2.3.7. Energy guided mutants to switch specificity in C-I FH.....	78
2.3.8. Examination of active site of class-I FH and predicted L-TTD structures.....	79
2.4. Summary	
2.5. Future directions	
2.5.1. L-TTD structure solution.....	86
2.5.2. Directed evolution to switch substrate specificity.....	86
2.5.3. Wider substrate screen for better understanding of enzyme active site.....	88
2.5.4. Interface analysis to understand the tight interaction between subunits of class-I FH and L-TTD.....	90
3. Chapter 3: Understanding modulation of IMPDH from archaea <i>M. jannaschii</i>	
3.1. Introduction	
3.1.1. IMPDH is a key player for nucleotide regulation within cells.....	94
3.1.2. General features of the enzyme IMPDH.....	94
3.1.3. Understanding IMPDH enzyme modulation in prokaryotes and eukaryotes.....	97
3.1.4. Cytoophidia and its role in physiology.....	98
3.1.5. Unique regulatory features of an archaeal MjIMPDH.....	99
3.1.6. Objectives.....	99
3.2. Experimental procedures	
3.2.1. Chemicals, strains, and molecular biology reagents.....	100
3.2.2. Sequence alignment and phylogenetic tree construction.....	100
3.2.3. Protein expression and purification.....	100
3.2.4. Crystallization, data collection and structure solution.....	101
3.2.5. Transmission electron microscopy imaging.....	102
3.3. Results and discussion	
3.3.1. Sequence analysis.....	103
3.3.2. Structure of IMPDH TIM domain.....	106
3.3.3. Transmission electron microscopy imaging of MjIMPDH.....	108
3.4. Summary	
3.5. Future directions	
Appendix A	113
Appendix B	114
Appendix C	115
Appendix D	116
Appendix E	118
Rights and permissions	125
References	129



(This page is intentionally left blank)



Chapter 1

Understanding the mechanism of catalysis in class-I fumarate hydratase

Class-I fumarate hydratase (FH) is present in many eukaryotic pathogens such as *Plasmodium*, *Cryptosporidium*, *Toxoplasma*, *Trypanosoma*, *Leishmania*, to name a few and its study is difficult considering the labile nature of Fe-S cluster. *Methanocaldococcus jannaschii* FH (MjFH) has been taken as a model to understand class-I FH owing to its stability. This chapter is aimed at understanding the mechanism of catalysis in class-I FH through biochemical, kinetic, and structural characterization of the enzyme. Mutational studies have been carried out to probe into the role of specific residues in catalysis. Structure of apo-MjFH has portrayed two different states adopted by the protein from its comparison with the previously solved structure of *Leishmania major* FH (LmFH). Results obtained from MD simulation studies support the two different structural states adopted by the protein (*Studies carried out by Sudharshan Behera from Prof. Balasubramanian Sundaram's laboratory at the Chemistry and Physics of Materials Unit, JNCASR as part of a collaborative project*). Mutational studies have disclosed that the enzyme does not follow a conventional acid-base catalytic mechanism, as both mutation of catalytic acid and catalytic base does not lead to abolished activity. A novel mechanism of catalysis which involves polarization of the substrate at the active site aiding easy addition and removal of water from the substrate by the enzyme has been proposed. Results obtained through QMMM studies have revealed that in the absence of catalytic residues, a water molecule could take over the role of catalysis, and Natural Population Analysis (NPA) has endorsed the possibility of substrate charge polarization at the active site both of which are discussed in this chapter (*QMMM studies and NPA studies were carried out by Soumik Das from Dr. Garima Jindal's laboratory at the Department of Organic Chemistry, Indian Institute of Science*). Our studies show that the class-I FH is a true Pauling enzyme, one which lures the substrate into its transition state to catalyze the reaction forward. Finally, future directions regarding this part of the study have been discussed.

1.1. Introduction

1.1.1. Moderately efficient enzymes and exceptional enzymes. Why fumarate hydratase is exceptional

Enzymes are biological catalysts that speed up metabolic reactions within the cells to rates fast enough to sustain life (Stryer L, Berg JM & Tymoczko JL 2002). As such, the concept of cell coming to life is through complex network of metabolite conversions enabled through biochemical reactions catalyzed by enzymes (Martin, 2011; Ullusu, 2015). Enzymes are known to catalyze as many as 5000 biochemical reactions (Schomburg et al., 2013), which has led to their commercial exploitation, considering that uncatalyzed reactions occur slowly at room temperature. Unlike the metal catalysts that are used in industries, enzymes offer another component of extreme importance which is specificity. This has allowed enzymes to attain the crown of remarkable catalysts.

Catalytic rates offered by enzymes is perhaps best expressed by its kinetic parameters, especially k_{cat} , referred to as turnover number which represents the number of molecules of substrates that can be converted to product by a single molecule of enzyme per unit time. Although enzymes are characteristically portrayed in textbooks as kinetically superior catalysts, it has been observed that most 'average enzymes' exhibit a k_{cat} of roughly 10 s^{-1} , which is much lower than the diffusion limit (**Fig. 1.1**) (Bar-even et al., 2011). The theoretical limit for k_{cat} of an enzyme is $10^6 - 10^7$ and the values for k_{cat}/K_M cannot exceed $10^8 - 10^9$. Very few enzymes are considered perfect, where the catalytic rates could reach close to that above values of theoretical limit. The rate limiting step of these reactions is the diffusion of the substrate into and out of the active site. Acetylcholine esterase, carbonic anhydrase, cytochrome c peroxidase, fumarase and triosephosphate isomerase are some of the well-studied enzymes with catalytic rates reaching that of diffusion limit (Bar-even et al., 2011).

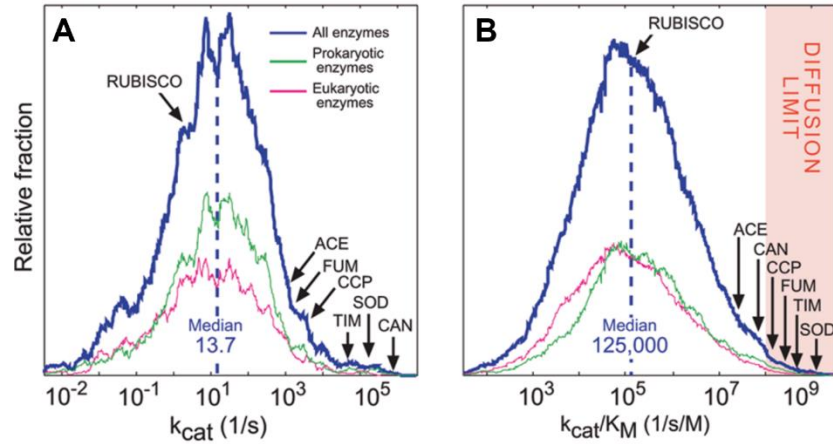


Fig. 1.1. Distribution of enzyme kinetic parameters adapted from Arren Bar-Even et al., 2011. **A.** k_{cat} values ($N=1942$) and **B.** k_{cat}/K_M values ($N=1882$). All values were considered for enzymes' natural substrate. Green and magenta lines correspond to the prokaryotic and eukaryotic enzymes, respectively. Location on the curve for several well-studied enzymes are highlighted. ACE, acetylcholine esterase; CAN, carbonic anhydrase; CCP, cytochrome *c* peroxidase; FUM, fumarase; Rubisco, ribulose-1,5-bisphosphate carboxylase oxygenase; SOD, superoxide dismutase; TIM, triosephosphate isomerase.

Enzymes function by reducing the free energy barrier of the reaction they catalyze or in other words, help the substrate overcome the transition state barrier. The feature that distinguishes most 'average' or moderately efficient enzymes from the "exceptional" enzymes is the level of reduction in free energy barrier of the reaction catalyzed. The level of reduction in transition state energy in case of fumarate hydratase (FH) is 30 kcal/mol which is substantially greater than some of the other exceptional enzymes such as triosephosphate isomerase (19 kcal/mol) and ketosteroid isomerase (21 kcal/mol) with the only other enzyme with comparable degree of stabilization being OMP decarboxylase (32 kcal/mol) (Bearne & Wolfenden, 1995). Fumarate is an extremely stable dicarboxylic acid, meaning if placed at room temperature, its hydration would take 700,000 years for completion without a catalyst. Hence, fumarate hydratase offers rate enhancements of the order of 10^{15} , a stellar number distinguishing it amongst some of the other exceptional catalysts.

1.1.2. Evolutionary dynamics of enzyme catalytic mechanism and kinetic parameters

Considering their ability to reach physically achievable maxima, most enzymes are rather sloppy with their kinetic parameters spanning over a large range of values. It is believed that kinetic parameters of an enzyme tend to evolve with a strong directional force up to a point beyond which they tend to evolve under near-neutrality (Labourel & Rajon, 2021). Enzyme evolution is strongly

linked to organismal fitness and a more accurate and holistic understanding of enzyme evolutionary process can only be understood in the light of organismal fitness. This is best exemplified by the understanding that, small functional changes in poor enzymes which catalyze rate-limiting metabolic step will lead to large changes in flux of the pathway and is reflected in cell growth. However, large changes to enzymes that already have high activity, reflect negligible effects on the metabolic pathway or the organismal fitness. Perhaps this is best explained with an example; in some bacteria TrpF that catalyzes an isomerization reaction in tryptophan biosynthesis, upon small improvements in enzyme performance showed large improvements in cell growth early in evolutionary trajectory. Later, mutations that had greater improvements in enzyme function however, had little effect on fitness (Klesmith et al., 2015). A similar phenomenon has been observed in levo-glucosan kinase, LGK.1 that is involved in the pathway for levoglucan utilization in *E. coli* (Newton et al., 2018).

Role of evolutionary pressures, although enigmatic and unclear seems to impose some changes in kinetic parameters. An attempt at understanding the evolutionary parameters that shape up the kinetic parameters of the enzyme shows that a number of factors may influence kinetic parameters. It has been observed that enzymes operating in the central metabolism compared to secondary metabolism are 30 fold faster in catalytic rates (Bar-even et al., 2011). This seems sensible considering the average high flux. For enzymes functioning in secondary metabolism, evolutionary pressures may be weaker. Many studies have also shown that several other factors may have a weak effect on enzyme evolution such as crowding effects in solution, physical constraints, reaction reversibility, enzyme gene expression and also protein folding (Labourel & Rajon, 2021). The above studies however fail to explain why inefficient enzymes having stronger control on metabolic flux and organismal fitness do not evolve higher efficiencies. Correlation between enzyme kinetic parameters and physicochemical constraints have also been observed (Bar-even et al., 2011). For smaller substrates of up to ~350 Da, the K_M decreases with increasing substrate mass and hydrophobicity. Also, addition of large substrate modifiers to small substrates considerably lowers the K_M value of enzymes utilizing the substituted substrates (Bar-even et al., 2011).

Role of evolution in the chemical step catalyzed by enzymes again is enigmatic considering that chemistry is not the rate-limiting step in many optimized enzymes. Presumably in their primitive enzymes, the step of chemical conversion could have been rate-limiting. As enzymes became faster along with evolving chemical step(s), kinetic events such as substrate binding, product release, pre-arrangement of proteins active site, etc., became rate-limiting. This would have again led to a lack of evolutionary pressure on the chemical step of enzyme catalysis (Klinman & Kohen, 2014).

With the knowledge that beyond a critical point, the kinetic parameters k_{cat} and k_{cat}/K_M of enzymes do not hamper the metabolic flux of a pathway or organismal fitness (Newton et al., 2018), the widely held view that to classify an enzyme as a perfectly evolved catalyst, the value of k_{cat}/K_M must approach that of diffusion limit ($10^9 \text{ M}^{-1}\text{s}^{-1}$) is questionable (Pettersson, 1989). Instead, evolutionary trajectory of the enzyme is dictated by a combination of biochemical, biophysical, and regulatory factors explaining why most enzymes are far from perfect catalysts. Evolution is not in pursuit of the perfect catalysts (Newton et al., 2018). Amidst this perplexity, fumarate hydratase which is at the core of the central metabolism and displays a k_{cat}/K_M of the order of diffusion limit, hints that the enzyme must be highly evolved.

1.1.3. Common set of underlying physical principles in enzyme catalysis

Enzymes catalyze an array of chemical transformations through many different types of mechanisms utilizing a common set of underlying physical principles. Enzymes are broadly grouped into seven classes based on the chemistry of the reaction catalyzed as oxidoreductases, transferases, hydrolases lyases, isomerases, ligases and translocases. One or many fundamental factors including proton donor-acceptor transfer motions, hydrogen bonding, hydrogen tunneling, electrostatics, pK_a shifting, preorganization, reorganization and conformational motions can contribute to the above chemical reactions catalyzed by enzymes (Hammes-Schiffer, 2012).

Hydrogen tunneling is known to reduce the free energy barrier of an enzyme catalyzed reaction by 2-3 kcal/mol. Nuclear quantum effects studies on dihydrofolate reductase (DHFR) have indicated that the free energy barrier for the hydrogen tunneling reaction decreased by ~ 2 kcal/mol (Agarwal et al., 2002). The motion of hydrogen donor and acceptor in the enzyme also plays a critical role in hydrogen transfer reactions. As the distance between the acceptor and the donor increases, the barrier to proton transfer decreases and vanishes completely at extremely short distances (Borgis

& Hynes, 1993). Hydrogen donor-acceptor distance has been observed to reduce from $\sim 3.3 \text{ \AA}$ to $\sim 2.7 \text{ \AA}$ in the transition state for both DHFR and ketosteroid isomerase (KSI) (Agarwal et al., 2002; Chakravorty et al., 2009). It has been observed that a larger network of hydrogen bonds is formed during the intermediate state of catalysis compared to the reactants and the product bound states in KSI. The strength of hydrogen bonding interactions between the dienolate intermediate and enzyme strongly influences free energy barrier for proton transfer reactions in KSI (Chakravorty & Hammes-Schiffer, 2010). The local environment in enzymes shift the pK_a of residues, which could have consequence on catalysis (Amyes et al., 2008; Harris & Turner, 2002). In aqueous solution, aspartate side chain carboxylate has a pK_a of ~ 4 and is more acidic than tyrosine ($\text{pK}_a \sim 10$). Mediated by electronic inductive effects and a large hydrogen-bonded network, catalytic tyrosine (Tyr16) hydroxyl group is more acidic than the catalytic aspartate (Asp103) in KSI (Hanoian et al., 2010). Electrostatic interactions between ligand and enzyme, similar to hydrogen-bonding is much larger in the intermediate state than in the substrate binding or product release state, signifying the importance of electrostatics in chemical reaction. Electrostatics have been shown to reduce the free energy barrier of KSI catalysed reaction by 7.3 kcal/mol of the total 11.5 kcal/mol by the enzyme (Fried & Bagchi, 2014). Local conformational changes on the enzyme usually happen on the millisecond time scale to ensure that configurations are conducive for chemical reaction. This makes conformational changes the rate limiting step with the actual reaction being virtually instantaneous (Hammes-Schiffer & Benkovic, 2006). This probability of sampling the transition state configurations by local conformational changes in the enzyme has been argued out to be an integral part in lowering the free energy barrier of the reaction by the enzyme (Hammes-Schiffer, 2012). Preorganization and reorganization, both involve conformational motions to some extent during a catalytic cycle. Preorganized active site requires conformational changes, such as lid loop opening or closing. While reorganization usually involves relatively small conformational changes during chemical reaction. In spite of having a preorganized active site, conformational changes do occur in KSI to bring down the proton donor-acceptor distance to $\sim 2.7 \text{ \AA}$ (Chakravorty et al., 2009). Mutations can in principle disrupt one or all of the above factors. Hydrogen-bonding capabilities, electrostatics, and conformational sampling are particularly harmed.

1.1.4. Class-I and class-II FH are enzymes with different folds catalyzing the same reaction with equal efficiencies

Fumarate, a symmetric olefinic dicarboxylic acid, is an important intermediate in the tricarboxylic acid (TCA, Krebs') cycle. In living cells fumarate is efficiently and stereospecifically converted to L-malate in a hydration reaction, catalysed by the enzyme fumarate hydratase (fumarase, FH, E.C. 4.2.1.2). FH are distinctly divided into two classes, class-I and class-II, based on the presence or absence of an [4Fe-4S] cluster (Woods et al., 1988). Both classes of FH are ubiquitous across all kingdoms of life. While, class-II enzymes occur predominantly in eukaryotes, class-I enzymes are widely distributed across archaea and prokaryotes, with sparse representation in eukaryotes (Jayaraman et al., 2018). Class-I and class-II FH share no sequence or structural similarity, yet catalyze the same reaction with close to equal efficiencies (Bearne & Wolfenden, 1995; Kronen & Berg, 2015). Class-II FH are homotetramers, whereas class-I FH are further divided into single-subunit and two-subunit types, depending on the number of genes coding for the functional enzyme (**Fig**). Two-subunit proteins are found only in prokaryotes and archaea, but not in eukaryotes. The alpha (α) and beta (β) subunits of the two-subunit type class-I FH correspond to the N-terminal and C-terminal domains (NTD and CTD) of single-subunit class I FH (Shimoyama et al., 2007). Single-subunit and two-subunit FH have high sequence similarity and a small polypeptide insertion between the two subunits is the only distinguishing feature between them. A few organisms including bacteria are found to have both classes of FH as exemplified by *Escherichia coli* (Woods et al., 1988). As to why the oxygen sensitive class-I FH is needed in the presence of a more robust class-II gene is not understood. It has been suggested that the class of FH the organism possesses may determine its lifestyle with regard to oxygen sensitivity (Shimoyama et al., 2007).

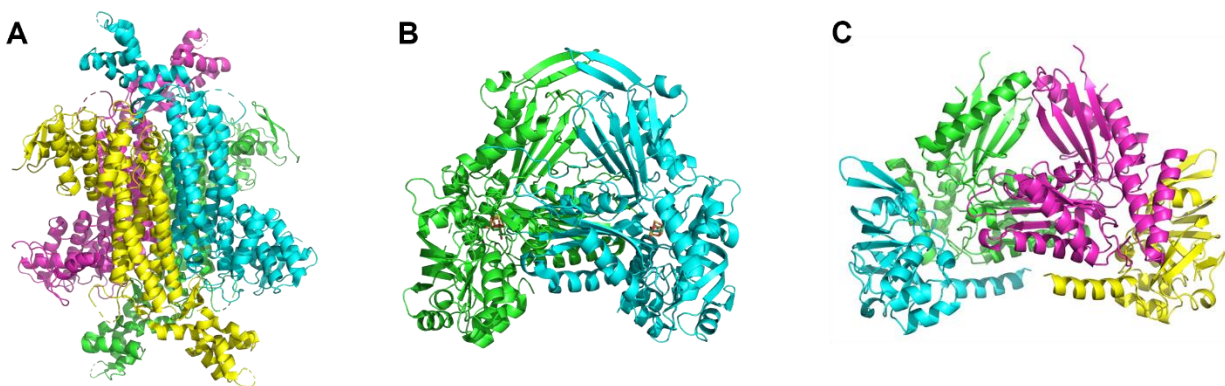


Fig. 1.2. Structure and domain organization in class-I and class-II FH. **A.** Structure of class-II FH from *Mycobacterium tuberculosis* (4apa). Each subunit of the tetramer is colored differently; yellow, magenta, cyan and green. Residues from 3 chains contribute to the active site at 4 locations on the enzyme. **B.** Structure of class-I single-subunit FH from *Leishmania major* (5l2r). Each subunit of the homodimer is colored in green and cyan and is further divided into NTD and CTD. Active site containing the Fe-S cluster is located in each monomer at the interface of NTD and CTD. **C.** Structure of class-I two-subunit FH from *Methanocaldococcus jannaschii* (this study). Highlighted in green and cyan are heterodimers of α - and β -subunits equivalent to the NTD and CTD of single-subunit FH. The functional form of the enzyme is a dimer of a heterodimer with the active site containing Fe-S cluster located at the interface of the heterodimers.

1.1.5. Understanding of catalysis in class-II fumarate hydratase

Class-II FH have been studied extensively from the angle of both the structural and functional aspects of the enzyme considering physiological relevance as they are present in humans. The enzyme has been a focus of study since early 1950s by multiple groups who have put forth different mechanistic perspectives on the rate enhancements offered by the enzyme.

1.1.5.1. Transition state stabilization

According to the transition state stabilization theory postulated by Linus Pauling in 1948 (Pauling, 1946; Pauling, 1948), large rate enhancements in reactions offered by enzymes occur due to the protein catalyst specificity to bind the substrate in an activated complex or the ‘transition state’. Comparing two substrates with similar chemical reactivity but differing enzymatic activity, the more reactive substrate is simply the one for which the enzyme possesses a stronger force of attraction to the transition state. Thus, it is suggested to be meaningless to distinguish binding sites and catalytic sites, as catalysis can only take place with an enhanced degree of binding (Schramm, 1998; Wolfenden, 2002). Richard Wolfenden and coworkers proposed that large rate enhancements observed by FH occur through transition state stabilization, where the greater negative charge developed on fumarate is stabilized by electrostatic interactions (Bearne &

Wolfenden, 1995). The apparent enthalpy and entropy calculated for binding of transition state showed that the enzyme is both entropically and enthalpically favored. It was suggested that binding of the substrate in the altered state corresponding to the transition state, would require a larger enzyme-ligand interactions through hydrogen-bonding, electrostatic and hydrophobic effects. Usually, these interactions tend to be accompanied by a gain in entropy. During the time of this study, the structure of FH was not available and Wolfenden predicted that a large network of bonding residues has to surround the ligand at the active site and the nature of the enzyme-ligand interactions would throw valuable light on this remarkable catalyst (Bearne & Wolfenden, 1995).

1.1.5.2. Steps involved in the reaction probed by kinetic isotope studies

There is support for the chemical mechanism in class-II fumarases taking place through all three pathways; carbonium (E1)(Hansen et al., 1969), concerted (E2)(Blanchard & Cleland, 1980) and carbanion (E1cB)(Rose, 1998) pathway (**Fig. 1.3**), with the weight of evidence tilting in favor of the carbanion pathway (Porter & Bright, 1980). Extensive kinetic isotope experiments in class-II FH have revealed that the rate limiting step of the reaction does not involve a carbon-hydrogen bond breakage or formation. Irwin Rose and coworkers have shown that the chemical interconversion of fumarate to malate takes place through carbanion mechanism and is the most rapid step of the reaction cycle (Rose & Weaver, 2004). Changes have to occur at the active site of the enzyme when the substrate is converted to product to accommodate the next substrate, for the reaction to occur again. Through an isotope counterflow method called the Britton counterflow effect, where labelled reaction products rebound to substrate when unlabeled substrate is added showed that the product form of the enzyme is slow to recycle (Britton, 1973; Rose et al., 1993). The rate limiting step is thought to occur right after product release, where a proton exchange alongside a slow conformational isomerization of the enzyme takes place (Rose et al., 1993).

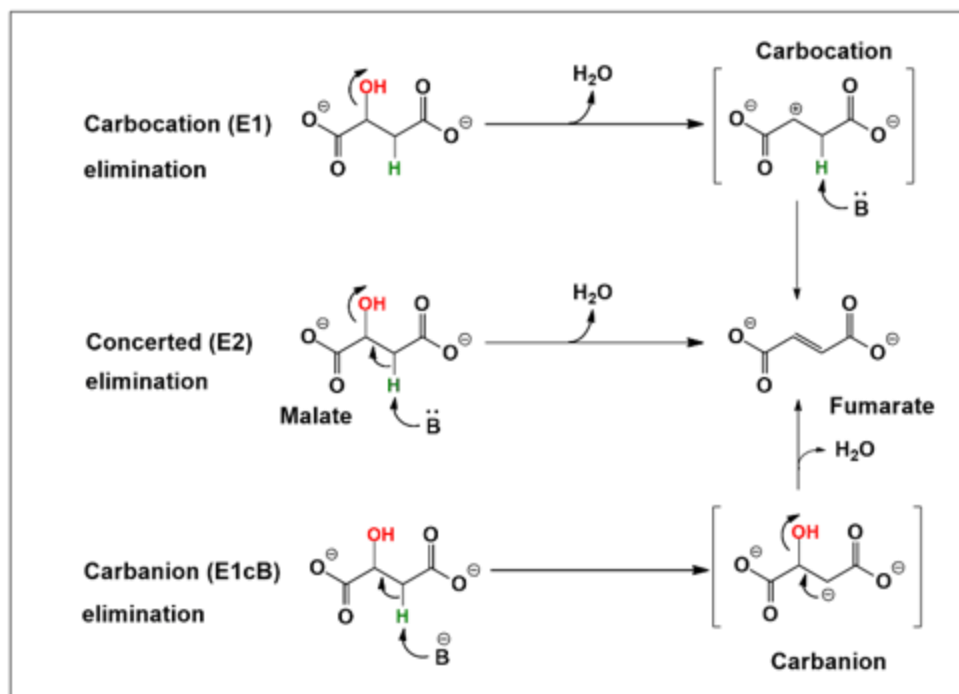


Fig. 1.3. Mechanisms for dehydration of malate to fumarate. E1 and E1cB mechanisms proceed through a two-step reaction involving a carbocation and a carbanion intermediate, respectively, while the E2 mechanism is a one-step reaction that proceeds without an intermediate with simultaneous removal of proton and elimination of water molecule.

1.1.5.3. Conventional acid/base catalysis

In the past decade, structural and mechanistic studies on class-II FH have focused on the malate dehydration reaction, with discussions centered on general acid-base catalysis. Class-II FH has been classified under the aspartase/fumarase family of enzymes which are known to share a similar catalytic fold and carry out a similar dehydration reaction (Ajalla Aleixo et al., 2019; Puthan Veetil et al., 2012). Aspartase, fumarase, arginosuccinate lyase (ARL) and adenylosuccinate lyase (ASL) are members of this class of enzymes. A conserved serine residue is believed to be the catalytic base while the catalytic acid has remained elusive in this class of enzymes (Puthan Veetil et al., 2012). An oxyanion hole is expected to form where a substantial perturbation of the pK_a of serine at the active site occurs through interaction with backbone amides (Tsai et al., 2007). Although sharing a similar fold, the rate limiting step in fumarate hydratase catalyzed reaction does not involve a catalytic base mediated proton abstraction, and the catalytic mechanism could be different in FH compared to other enzymes of the family. Moreover, FH is mostly a hydratase rather than a dehydratase. The biochemical equilibrium in the reversible hydration reaction in the

TCA cycle lies in the direction of malate ($K_{eq}=4$, first measured by Krebs) (Goldberg & Tewari, 1995; Krebs, 1953), suggesting that the rates of hydration are 4-fold higher than the rates of dehydration (Erickson et al., 1959). The question arising from studies on FH is, how does the enzyme lower a formidable activation barrier for a rather deceptively simple reaction? While kinetic and thermodynamic analyses have provided valuable insights, the molecular mechanisms remain to be clarified (Gajewski et al., 1985; Erickson et al., 1959).

1.1.6. Understanding of catalysis in class-I fumarate hydratase

Unlike class-II fumarases, a comprehensive understanding of enzyme chemistry and catalysis is lacking for class-I FH due to their thermolabile and oxygen-sensitive nature. Single-subunit class-I FH have been biochemically characterized in a handful of organisms (De Pádua et al., 2017; Feliciano et al., 2012, 2016; Jayaraman et al., 2018; Kronen et al., 2015; Shibata et al., 1985; Van Kuijk & Stams, 1996; van Vugt-Lussenburg et al., 2013). Thus far, two-subunit Class-I FH from only *Pelotomaculum thermopropionicum* (Shimoyama et al., 2007) and *Pyrococcus furiosus* (van Vugt-Lussenburg et al., 2009) have been biochemically characterized. Much of the understanding on mechanism of catalysis in class-I FH comes from the well-studied Fe-S cluster hydrolyase, aconitase. Catalysis in class-I FH has been modelled on that followed in aconitase (Flint & Allen, 1996). The recent reports of the structures of a substrate and inhibitor bound class-I FH from *Leishmania major* is a major step forward in probing the molecular mechanism of the fumarase reaction (Feliciano et al., 2016, 2019; Feliciano & Drennan, 2019).

1.1.6.1. Fe-S cluster hydrolyases

Enzymes in the hydrolyase class [EC 4.2.1.x] catalyze a reaction in which a water molecule is removed from a carbon-carbon bond, thus converting an alcohol into a vinyl group. Enzymes of this class are further subdivided into ones that contain metal, lack a metal, or use Fe-S cluster for catalysis. A number of hydrolyases utilize Fe-S cluster for catalysis, such as aconitase, homoaconitase, isopropyl malate isomerase, fumarase, maleate hydratase, mesaconase, tartrate dehydratase, dihydroxy-acid dehydratase, phosphogluconate dehydratase, serine dehydratase and quinolinate synthase (Flint & Allen, 1996). Only a few enzymes of the hydrolyase class containing [4Fe-4S] are well studied, with biochemical studies and structural information available only for aconitase, fumarase, serine dehydratase and quinolinate synthase (Beinert et al., 1996; Cherrier et al., 2014; Feliciano et al., 2016; Lloyd et al., 2008; Saunders et al., 2008; Thoden et al., 2014).

Understanding of catalysis adapted by this class of enzymes comes majorly from studies on aconitase. The [4Fe-4S] containing hydrolyase enzymes adapt a common catalytic strategy where the cluster is primarily involved in activating the substrate hydroxyl group by acting as a Lewis acid to make it a better leaving group. In aconitase, a serine residue acts as a catalytic base to abstract a proton from the substrate and an aspartate and a histidine residue form an ion pair to eliminate substrate hydroxyl group (Zheng et al., 1992). Based on the catalytic mechanism adopted by aconitase, a similar mechanism has been proposed to occur in class-I FH (**Fig. 1.4**) (Flint & Allen, 1996). Apart from aconitase majorly being a dehydratase and FH a hydratase, a number of differences between the two have been observed, suggesting the mechanism may not be all the same. One of the important differences is the sensitivity of [4Fe-4S] cluster to oxygen in these enzymes. Upon exposure to air, the labile iron atom (Fe_a) which is not ligated to cysteine in aconitase falls off to form an oxidized [3Fe-4S] cluster which is inactive and is stable in air for several hours. In FH however, the cluster is extensively degraded with the enzyme losing almost all activity within two minutes of exposure to air (Flint, 1993; van Vugt-Lussenburg et al., 2013). While reconstitution of Fe-S cluster takes hours in FH, it takes only 5 minutes for aconitase suggesting that the cluster is more stable in aconitase which makes it easier for studying. The relative “stickiness” of substrates to the cluster as demonstrated by K_M/K_d ratio is 100 times larger in aconitase compared to FH (Flint et al., 1992). While there are only 11 residues observed in the active site of FH, aconitase has as many as 23 residues contributing to the active site (Lauble et al., 1992). These observations hint that environment around the cluster in the two active sites might be different for these two classes of enzymes. From biochemical and structural studies in aconitase, we know that the labile iron atom Fe_a is tetradentate and bound to a hydroxyl group ($FeOH$) in the absence of substrate. Upon entry of substrate, the hydroxyl group gets protonated to a water molecule still remaining bound to Fe_a (FeO_w) which in turn expands its coordination sphere to become hexadentate (Lauble et al., 1992). However, substrate bound structure of *Leishmania major* class-I FH has Fe_a forming a pentadentate bond, having lost its initial bound water molecule (5l2r). Further work must reveal differences in enzymes within the Fe-S cluster hydrolyase class of enzymes.

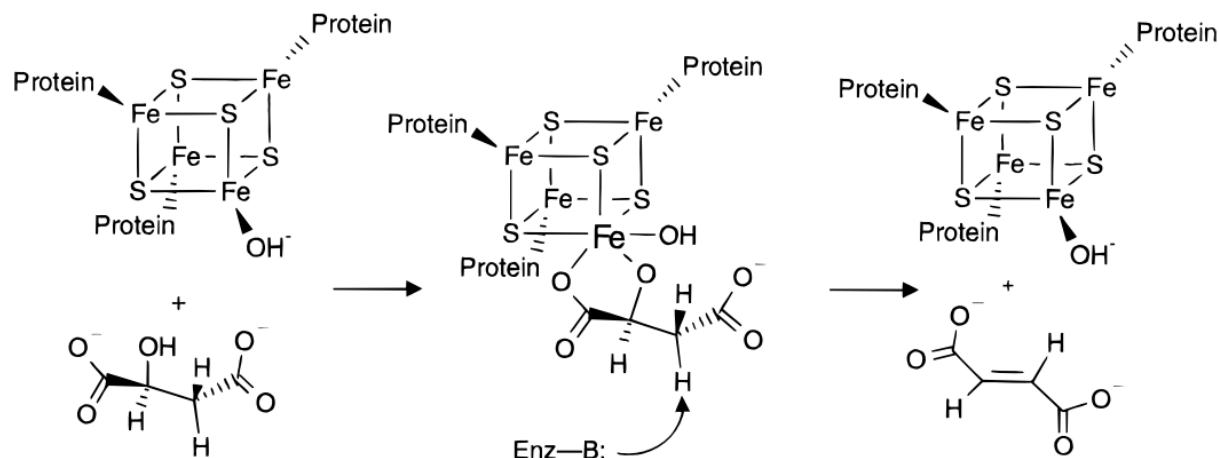


Fig. 1.4. Proposed reaction mechanism for class-I FH, adapted from Flint & Allen 1996. The Fe-S cluster is ligated to the protein through three cysteine residues. The fourth iron which is unbound is called the labile iron (Fe_a). In the absence of the substrate, Fe_a is bound to a hydroxyl group to form a tetradentate structure. Upon entry of substrate malate, the hydroxyl group on the substrate along with its carboxylate group binds to the labile iron to form a hexadentate structure. The Fe-S cluster functions as a Lewis acid to pull the lone pair of electrons on the hydroxyl group of the substrate to render it prone for breakage. This is followed by a catalytic base abstracting a proton from the C3 carbon of malate subsequently leading to the formation of product with the release of a water molecule.

1.1.6.2. *L. major* class-I FH

Recently, the first full length structure of class-I FH has been solved from *Leishmania major*, a single-subunit type enzyme (Feliciano et al., 2016). Subsequently, structures of the enzyme bound to various substrate analogs and inhibitor thiomalate has also been solved and deposited (Feliciano et al., 2019; Feliciano & Drennan, 2019). As the LmFH structures have both Fe-S cluster and substrate bound at the active site, critical information with regard to active site residues interacting with the substrate can be deciphered (**Fig. 1.5**). A total of 7 residues from the CTD and 5 residues from the NTD are in hydrogen-bonding distance with the substrate, with a threonine residue (Thr467) closest towards the C3 carbon of the substrate (3.2 Å) and an aspartate residue (Asp135) closest to the C2 hydroxyl group of the substrate (2.9 Å). Mutational studies (carried out in subsequent studies) of active site residues of threonine (T467A), aspartate (D135A) and three arginine residues (R173A, R421A, and R471A) in hydrogen-bonding distance with the substrate led to a 2,000 to 16,000 fold drop in k_{cat} (Feliciano & Drennan, 2019). The two arginine residues Arg421 and Arg471 are in hydrogen-bonding distance with Thr467 (2.5 – 3.0 Å), either of which could accept a proton from C3 carbon of substrate via Thr467 side chain. It is proposed that the three residues act as a catalytic base triad to abstract the proton from the substrate in a base

catalyzed reaction. Asp135 which is closest to the C2OH group of the substrate along with its hydrogen-bonded partner His334 (2.5 Å) is proposed to act as the catalytic acid dyad to donate a proton to the hydroxyl group to aid its removal as a water molecule. Arg173 has been proposed to play an important role in positioning of the substrate at the active site to allow efficient catalysis to take place. Mutational studies of some of the other residues that are in hydrogen-bonding distance with the substrate have not been carried out.

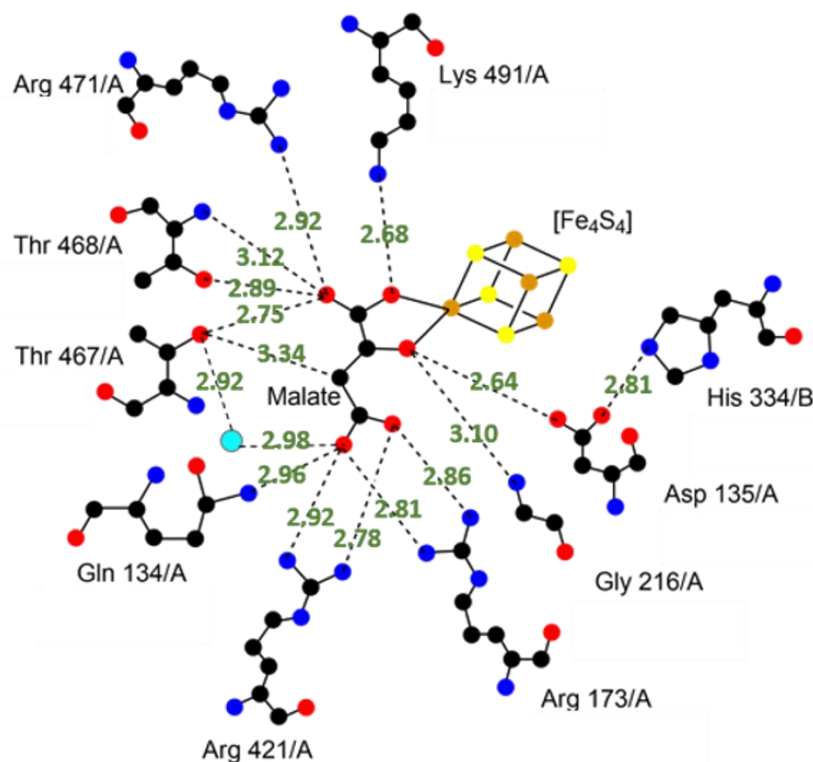


Fig. 1.5. Active site of LmFH with bound L-malate (5I2r). Active site of the protein includes an iron-sulfur cluster, bound L-malate, and substrate binding residues. The residues involved in malate binding and their corresponding distance from the different atoms of malate are shown. Image was created using LigPlot (Laskowski & Swindells, 2011). Water molecule at the active site is shown in cyan; C, N, O, Fe, and S atoms are shown in black, blue, red, orange, and yellow, respectively. Hydrogen-bonds are shown as black dashed lines with distances highlighted in green.

1.1.7. C-I and C-II FH show differences with regard to enzyme promiscuity

Apart from fumarate – malate interconversion, FH has promiscuous activity on substrate analogs like D-tartrate and acetylene dicarboxylate. Although, both class-I and class-II FH have activity on D-tartrate and acetylene dicarboxylate, there exists other similarities and differences between the two class of enzymes (Flint, 1994; Teipel et al., 1968) (summarized in **Table. 2.3** in chapter 2). For example, mesaconate is a substrate of class-I FH while, it is a competitive inhibitor of

class-II FH. Succinate and L-tartrate are inactive analogs for class-I FH activity but, inhibit class-II FH competitively. Mercaptosuccinic acid or thiomalate is shown to be a competitive inhibitor of class-I FH (Jayaraman et al., 2018), but the same is an inactive analog of class-II FH. Substrate analogs such as malonate, trans-aconitate and meso-tartrate inhibit both class-I and class-II FH. The fact that both these enzymes are able to act on and are inhibited by some mutually exclusive substrate analogs suggests there may be some differences with regard to substrate binding and catalysis. No differences between the class-I single-subunit and two-subunit enzymes have been observed and a wider substrate screen must give more information on mechanistic differences in catalysis adapted by the two class of enzymes.

1.1.8. Physiological and clinical significance

Class-II FH have been found to be localized in both mitochondria and cytosol in many eukaryotes including yeast (Knox et al., 1998), rats (Tuboi et al., 1986), plants (Pracharoenwattana et al., 2010), and humans (Dik et al., 2016). In the mitochondria, the primary role of the enzyme is to act in congruence with the TCA cycle, while diverse roles have been associated to the cytosolic counterpart. Cytosolic class-II FH have been found to be involved in urea cycle in mammalian cells (Adam et al., 2013), nitrogen assimilation (Nunes-Nesi et al., 2010), seed germination (Eprintsev et al., 2018) and carbon storage (Zell et al., 2010) in plants and more recently, DNA repair pathway of yeast and mammalian cells (Jiang et al., 2018). Mutation in, or inactivation of FH has pathological consequences. Biallelic inactivation of FH is known to cause genetic disorders such as hereditary leiomatosis, renal cell carcinoma and encephalopathies (Tomlinson et al., 2002).

With the physiological importance associated with class-I FH in all the organisms, the enzyme has been characterized for only its role in the TCA cycle. However, the promiscuous activity of fumarases on mesaconate in *Burkholderia xenovorans* (Kronen et al., 2015) and D-tartrate in *Escherichia coli* (Kim et al., 2007) suggests the role of class-I FH in aiding the organisms' survival in certain conditions. Similar to class-II FH, class-I FH have recently been shown to be essential components in DNA damage response in *Bacillus subtilis* (Silas et al., 2021). Some of the unicellular eukaryotic pathogens are found to solely possess class-I FH. Since humans have only class-II FH, the perturbation of which could lead to diseases, designing class-I FH specific

inhibitors will help treat some of the largely prevalent diseases caused by unicellular eukaryotic pathogens listed in **Table. 1.1**.

Table. 1.1. List of eukaryotic unicellular organisms having class-I FH and the respective disease they cause in humans.

Organism	Disease	FH type
<i>Plasmodium</i>	Malaria	Class-I
<i>Babesia</i>	Babesiosis	Class-I
<i>Cryptosporidium parvum</i>	Cryptosporidiosis	Class-I
<i>Cyclospora cayetanensis</i>	Cyclosporiasis	Class-I
<i>Isospora belli</i>	Isosporiasis	Class-I
<i>Toxoplasma gondii</i>	Toxoplasmosis	Class-I
<i>Trypanosoma brucei</i>	Sleeping sickness	Class-I
<i>Trypanosoma cruzi</i>	Chagas disease	Class-I
<i>Leishmania</i>	Leishmaniasis	Class-I

1.1.9. Objectives

Owing to the difficulty in biochemical and structural characterization of thermolabile and oxygen sensitive class-I FH, a thermostable homolog of the enzyme from *M. jannaschii* was selected for the current study to serve as a model to understand class-I FH. The primary objective of the study was to kinetically and structurally characterize the class-I FH from *M. jannaschii*. With the structure solved, analyze the active site of the enzyme to draw information on functioning of the highly efficient catalyst. Carry out mutational analysis to understand the role of active site residues in catalysis and finally arrive at the mechanism of catalysis adapted by class-I FH in general.

1.2. Experimental procedures

1.2.1. Chemicals, strains, and molecular biology reagents

Restriction enzymes, Phusion polymerase and T4 DNA ligase were obtained from New England Biolabs, USA. Primers were custom synthesized from Sigma Aldrich, Bengaluru. Media components for growing *E. coli* cultures were from Himedia, Mumbai, India. Akta HPLC and Q-Sepharose resins were from GE Healthcare Life Sciences, USA. Selenomethionine was obtained from Cayman chemicals Co., Michigan. Crystallization cocktails were from Hampton Research, Aliso Viejo, CA. *RS-2*-thiomalate was obtained from Sigma Aldrich, USA. All enzyme assays were monitored using Hitachi U2010 (Hitachi High Technologies, Japan) spectrophotometer fitted with water-circulated cell holder.

1.2.2. Generation of plasmid constructs

The genes for the subunits of MjFH α and MjFH β cloned into two tandem multiple cloning sites of pET-Duet vector (Novagen, Merck), under the restriction sites BamHI and SalI for MjFH α subunit and NdeI and XhoI for MjFH β subunit for co-expression from a single plasmid (pETduet_MjFH $\alpha\beta$) was already available in the laboratory. For expressing the individual subunits separately, the same primers and restriction sites were used (**Table. 1.2**). Site-directed mutants were generated using the PCR-driven overlap extension method using pETduet_MjFH $\alpha\beta$ as a template. DNA fragments with the required mutations generated using appropriate oligonucleotides (**Table. 1.2**) were then assembled into the plasmid by ligation-dependent cloning (Sambrook & Russell, 2001). All the clones were confirmed by DNA sequencing.

Table. 1.2. List of oligonucleotide sequences used for cloning *M. jannaschii* WT FH and generating site-directed mutants.

Primer name	Primer sequence (5' to 3')
MjFH α BamHI_up	CGCGGATCCGAAAATCTCCGATGTTGTTGTTGAATTATTTAG
MjFH α SalI_down	ACGCGTCGACTTATAATTTAGCATCCAATTTTATTCTTTTAATTGCC
MjFH β NdeI_up	CTAATTCCATATGGAATATACATTTAACAAATTAACAAAAAAGATG
MjFH β XhoI_down	CCGCTCGAGTTATAATCCTATCAATTCATTAAGCTTTTCATAAAC
MjFH_D62V_FP	GTCCTCTATGTCAAGTACTGGTGTCCCAATAG
MjFH_D62V_RP	CTATTGGGACACCAGTAACTTGACATAGAGGAAC
MjFH_H257N_FP	GAGATTGCTGGATGCAACACAGCTTCTTTACCTGTAGG
MjFH_H257N_RP	CCTACAGGTAAAGAAGCTGTGTTGCATCCAGCAATCTC
MjFH_T80V_FP	GGTTTGTGTTTCTATAGGCCAGTAACATCTGCAAGGATGAATG
MjFH_T80V_RP	CATTCATCCTTGCAGATGTTACTGGGCCTATAGAAACACAAACC
MjFH_T80C_FP	GGTTTGTGTTTCTATAGGCCATGTACATCTGCAAGGATGAATG

MjFH_T80C_RP	CATTCATCCTTGCAGATGTACATGGGCCTATAGAAACACAAACC
MjFH_G147A_FP	GCATTTCCAAAAGGGGCAGCGAGCGAAAACATGAGTGC
MjFH_G147A_RP	GCACTCATGTTTTTCGCTCGCTGCCCTTTTGGAAATGC
MjFH_T81V_FP	GTGTTTCTATAGGCCCAACAGTTTCTGCAAGGATGAATGATG
MjFH_T81V_RP	CATCATTATCCTTGCAGAACTGTTGGGCCTATAGAAACAC
MjFH_R84L_FP	GCCCAACAACATCTGCATTAATGAATGATGTTGAAGAGG
MjFH_R84L_RP	CCTCTTCAACATCATTCATTAATGCAGATGTTGTTGGGC
MjFH_K104L_FP	GCAATTGTTGGACTGGGAGGAATGAAAAAAGAG
MjFH_K104L_RP	CTCTTTTTTCATTCTCCAGTCCAACAATTGC
MjFH_R102L_FP	GGAAGAGGTTCTTTATTACCTAATGTAGTTCATCC
MjFH_R102L_RP	GGATGAACTACATTAGGTAATAAAGGAACCTCTTCC
MjFH_Q61L_FP	GAAACGCAAGTTCCTCTATGTTTAGATACTGGTGTCCCAATAG
MjFH_Q61L_RP	CTATTGGGACACCAGTATCTAAACATAGAGGAACTTGC GTTTC
MjFH_R32L_FP	GGCAAATATACACTGCGTTAGATGAAGCACATTTAAAAATTATTG
MjFH_R32L_RP	CAATAATTTTTAAATGTGCTTCATCTAACGCAGTGTATATTTTGCC
MjFH_K144L_FP	GATAATTGCATTTCCATTAGGGGCAGGAAG
MjFH_K144L_RP	CTTCCTGCCCTAATGGAAATGCAATTATC
MjFH_S148A_FP	GCATTTCCAAAAGGGGCAGGAGCCGAAAACATGAGTGC
MjFH_S148A_RP	GCACTCATGTTTTTCGGCTCCTGCCCTTTTGGAAATGC
MjFH_R32K_FP	GGCAAATATACACTGCGAAAGATGAAGCACATTTAAAAATTATTG
MjFH_R32K_RP	CAATAATTTTTAAATGTGCTTCATCTTTCGCAGTGTATATTTTGCC
MjFH_R84K_FP	GCCCAACAACATCTGCAAAAATGAATGATGTTGAAGAGG
MjFH_R84K_RP	CCTCTTCAACATCATTCATTTTTGCAGATGTTGTTGGGC
MjFH_R102K_FP	GGAAGAGGTTCTTTAAACCTAATGTAGTTCATCC
MjFH_R102K_RP	GGATGAACTACATTAGGTTTTAAAGGAACCTCTTCC

1.2.3. Protein expression and purification

Recombinant expression of MjFH $\alpha\beta$ was carried out by transforming the *E. coli* strain BL21 (DE3)-RIL with the plasmid pETduet_MjFH $\alpha\beta$, followed by selection on Luria-Bertani (LB) agar plate containing ampicillin (100 $\mu\text{g ml}^{-1}$) and chloramphenicol (100 $\mu\text{g ml}^{-1}$). Multiple colonies on the plate were picked and transferred to 10 ml LB broth. The culture was grown overnight at 37 $^{\circ}\text{C}$ and 1% of inoculum was added to 800 ml of Terrific broth (TB). The cells were grown at 37 $^{\circ}\text{C}$ to an OD₆₀₀ of 0.45, induced with 0.3 mM IPTG and grown further for 18 hours at 16 $^{\circ}\text{C}$. The cells were harvested by centrifugation and resuspended in lysis buffer containing 50 mM Tris HCl, pH 8.0, 5% glycerol, 1 mM DTT and 1 mM PMSF. Cell lysis was achieved by five cycles of French press (Thermo IEC Inc., USA) at 1000 psi, and the lysate was clarified by centrifugation at 30,000 x g for 30 minutes. Supernatant was incubated at 70 $^{\circ}\text{C}$ for 30 minutes to precipitate the bacterial proteins and again clarified by centrifugation at 30,000 x g for 30 minutes. Supernatant was then treated with 0.01% PEI to precipitate nucleic acids and clarified by centrifugation at 30,000 x g for 30 minutes. The clarified lysate was filtered through a 0.44-micron filter and loaded

onto Q-Sepharose anion-exchange column. The protein was eluted using a linear gradient of NaCl in buffer containing 50 mM Tris HCl, pH 8.0, 5% glycerol and 1 mM DTT. Fractions containing co-purified α and β -subunits of the protein were pooled, dialyzed, and stored at -80°C . The purified protein was analyzed by SDS-PAGE (Laemmli, 1970) and protein concentration was determined by the Bradford method (Bradford, 1976) with bovine serum albumin (BSA) as standard. The same protocol was adopted for purification of all the mutants.

Expression and purification of MjFH $\alpha\beta$ with selenomethionine incorporated for single-wavelength anomalous dispersion (SAD) phasing, was carried out following established protocols (Doublé, 1997). Overnight grown culture of *E. coli* strain BL21 (DE3)-RIL containing pETduet_MjFH $\alpha\beta$ was centrifuged, and the cells washed with M9 minimal medium. 0.5% culture was inoculated in pre-warmed M9 minimal medium containing 5 g L⁻¹ glucose as carbon source. Cultures were grown till mid-log phase at 37 °C followed by addition of inhibitory amino acid mix containing phenylalanine, lysine, and threonine at a concentration of 100 mg L⁻¹ and leucine, isoleucine, and valine at a concentration of 50 mg L⁻¹. Cultures were grown for 15 minutes followed by addition of selenomethionine to a final concentration of 100 mg L⁻¹ and grown for further 15 minutes. The cultures were then induced with 0.3 mM IPTG and grown at 20 °C for 18 hours. The cells were harvested, and protein purified following the same protocol as that for the wild type (WT) protein using degassed buffer containing 50 mM Tris HCl, pH 8.0, 5% glycerol and 10 mM DTT. The purified protein was checked on SDS-PAGE and analyzed using LC-ESI-MS for 100% selenomethionine incorporation into both subunits (**Appendix. Fig. A.1**) Purified protein aliquots were concentrated and utilized for setting up X-ray crystal trays.

1.2.4. Reconstitution of Fe-S cluster and confirmation through CD spectroscopy

Purified protein aliquots stored at -80°C were reconstituted with Fe-S cluster following standard protocols (Beinert et al., 1996; Jolla, 1996). The protein solution was incubated in anaerobic chamber (Coy, USA) for 1 h to remove all traces of oxygen. All the subsequent steps were performed under anaerobic conditions in the chamber. Briefly, around 10-50 μM protein was used for reconstitution. The reconstitution procedure was initiated by the addition of 50-fold excess DTT to the protein solution followed by stirring for 30 min. Subsequently, 10-fold molar excess of ferrous ammonium sulfate and sodium sulfide were added and stirred for another 3.5 hours. Metal-thiolate charge transfer band was detected in the visible region using a 20 μM solution of

[4Fe-4S] cluster-reconstituted protein taken anaerobically in a 1 cm path length cuvette, sealed, and used for acquiring CD spectra in the region 250 to 600 nm using a spectropolarimeter (Jasco J-810).

1.2.5. Enzyme activity

All initial velocity measurements were conducted at 70 °C using a spectrophotometer (Hitachi U2010 spectrophotometer) and initiated by the addition of enzyme. The activity of MjFH was found to be maximal at pH 7.2 and hence all assays were carried out at this pH in 50 mM Tris-HCl. Conversion of fumarate to malate was measured spectrophotometrically as a decrease in absorbance caused by the depletion of fumarate. The enzymatic conversion was monitored at different wavelengths as follows: 240 nm ($\epsilon_{240} = 2440 \text{ M}^{-1}\text{cm}^{-1}$) (Flint, 1994) for fumarate concentrations up to 750 μM ; 250 nm ($\epsilon_{250} = 1345 \text{ M}^{-1}\text{cm}^{-1}$) for 1 mM; 260 nm ($\epsilon_{260} = 805 \text{ M}^{-1}\text{cm}^{-1}$) for 1.5 mM; 270 nm ($\epsilon_{270} = 475 \text{ M}^{-1}\text{cm}^{-1}$) for 2 mM; 280 nm ($\epsilon_{280} = 240 \text{ M}^{-1}\text{cm}^{-1}$) for 3 mM; 290 nm ($\epsilon_{290} = 110 \text{ M}^{-1}\text{cm}^{-1}$) for 5 mM; 300 nm ($\epsilon_{300} = 45 \text{ M}^{-1}\text{cm}^{-1}$) for 10 and 20 mM; 305 nm ($\epsilon_{305} = 18 \text{ M}^{-1}\text{cm}^{-1}$) for concentrations of 30 and 40 mM in a quartz cuvette of 1 cm path length.

Inhibition kinetics of the enzyme with *RS*-2-thiomalate was carried out using similar assay conditions with fixed concentrations of fumarate (500 μM) and malate (1000 μM) in 50 mM Tris-HCl, pH 7.2 at varying concentrations of *RS*-2-thiomalate. Apparent inhibitory constant K_i (app) was calculated from the IC_{50} -to- K_i web-server (Cer et al., 2009) for *RS*-2-thiomalate for MjFH with malate and fumarate as substrates.

All initial velocity measurements were conducted in duplicate with data points in the plot and values derived being mean \pm S.E. The initial rate vs substrate concentration plots were generated using GraphPad Prism version 5.0 (Graphpad software Inc) and non-linear regression method was used to fit the data points for WT and mutants using the equation, $v = V_{\max}[\text{S}]/[K_M + [\text{S}](1 + [\text{S}]/K_I)]$ (**eq. 1**) that incorporates substrate inhibition. IC_{50} was determined by fitting the data to the equation $v = V_0/[1 + (I/\text{IC}_{50})]$ (**eq. 2**), where v is the initial velocity, V_{\max} is maximum initial velocity, $[\text{S}]$ is substrate concentration, K_I is the dissociation constant for equation 1 and v is the observed velocity, V_0 is the uninhibited velocity, $[\text{I}]$ is the inhibitor concentration and IC_{50} in equation 2 is the 50% inhibitory concentration of the inhibitor.

1.2.6. Crystallization and data collection

MjFH α , MjFH $\alpha\beta$ apo-protein and holo-protein were set up for crystallization using the Microbatch method (Chayen et al., 1992). A 72 well multi-well plate from Grenier-Bio was used. Conditions for crystallization were obtained by using all the conditions in the commercially available crystallization kit from Hampton (Hampton Research, USA). In case of MjFH $\alpha\beta$ holo-protein, reconstituted protein was passed through a ZebaTM Spin Desalting Column, 7K MWCO, to remove unbound iron and sulphur and set up for crystallization within the anaerobic chamber. For MjFH $\alpha\beta$ apo-protein, to ensure a homogenous population, the protein was subjected to apo-protein preparation. Apo-protein was prepared by addition of 50x excess EDTA and 20x excess K₃[Fe(CN)₆] followed by incubation for 30 minutes. The protein was concentrated by precipitation with 100% ammonium sulfate and dissolved in a minimum volume of 20 mM potassium phosphate, pH 6.4 and dialysed. Various concentrations ranging from 3 mg/ml to 24 mg/ml of the protein were used for setting up crystal screens. The crystallization droplet contained 3 μ l of protein and 3 μ l of buffer from different conditions placed under a 50% mix of paraffin and silicone oil at room temperature. Crystals of MjFH $\alpha\beta$ apo-protein were obtained in 0.2 M magnesium chloride hexahydrate, 0.1 M BIS-TRIS, pH 5.5, 25% w/v polyethylene glycol 3,350. Selenomethionine incorporated MjFH $\alpha\beta$ (Se-MjFH $\alpha\beta$) also crystallized in the same condition but were not well formed. Well-formed crystals were obtained in the presence of an additive, non-detergent sulfobetaines (0.3 M NDSB-195) obtained from Hampton additive screenTM (**Fig. 1.6**). Crystals were cryo-protected by addition of 20% (v/v) glycerol to initial conditions. Selenium-SAD data was collected on Se-MjFH $\alpha\beta$ crystals at a wavelength of 0.9537 at the BM-14 beamline of European Synchrotron Radiation facility (ESRF).

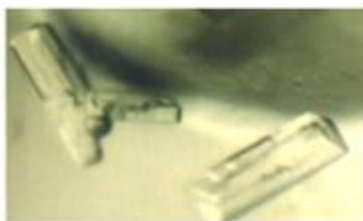


Fig. 1.6. MjFH protein crystal morphology. Crystals obtained under the condition containing 0.2 M magnesium chloride hexahydrate, 0.1 M BIS-TRIS, pH 5.5, 25% w/v polyethylene glycol 3,350 in the presence of the additive, non-detergent sulfobetaines (0.3 M NDSB-195).

1.2.7. Structure determination and refinement

Crystal structure of Se-MjFH $\alpha\beta$ was solved by selenium-SAD, making use of the eleven anomalously scattering selenium atoms per functional dimer. Various packages available in CCP4 suite (Winn et al., 2011) and Phenix modules (Adams et al., 2010) were used. Briefly, iMOSFLM (Battye et al., 2011) was used for data processing and PHENIX AutoSol (Adams et al., 2010) for determining the position of selenium atom sites in Se-MjFH $\alpha\beta$. Automatic model building was performed for Se-MjFH $\alpha\beta$ using PHENIX AutoBuild (Terwilliger et al., 2007). REFMAC 5.0 (Murshudov et al., 2011) and phenix.refine (Afonine et al., 2012) were used for refinement. Manual refinements were carried out using COOT (Emsley et al., 2010). Refinement statistics are summarized in **Table 1.3**. Since the structure has been held for publication during the submission of the thesis, validation report for the structure is attached in **Appendix E**. All the structure-related figures were created using PyMol software (Schrödinger, 2015). Electrostatic surface potential was calculated using Adaptive Poisson-Boltzmann Solver plugin in PyMol (Baker et al., 2001).

Table 1.3. Data collection and refinement statistics of MjFH $\alpha\beta$ protein structures solved by Se-SAD.

PDB ID	7XKY
Wavelength	0.9537, 1.54
Resolution range	52.07 - 2.46 (2.548 - 2.46)
Space group	C 1 2 1
Unit cell	116.949 61.7 83.8796 90 123.9 90
Total reflections	142561 (68284)
Unique reflections	17961 (2578)
Multiplicity	7.9 (7.4)
Completeness (%)	98.10 (97.50)
Mean I/sigma(I)	11.2 (3.9)
Wilson B-factor	30.29
R-merge	0.124 (0.525)
R-meas	0.145 (0.598)
R-pim	0.051 (0.217)
CC1/2	0.976 (0.871)
Reflections used for R-free	934 (91)
R-work	0.205
R-free	0.2439
RMS(bonds)	0.015
RMS(angles)	2.05
Ramachandran favored (%)	96.18
Ramachandran allowed (%)	3.82
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	3.14
Clashscore	5.58
Average B-factor	34.34

1.3. Results and discussion

1.3.1. Enzyme activity and kinetic characterization of MjFH

The activity of reconstituted MjFH $\alpha\beta$ complex was examined at 240 nm as a time-dependent decrease in absorbance with fumarate as substrate and increase in absorbance with malate as substrate. The MjFH apo-protein, β -subunit and reconstituted α -subunit were found to be inactive while, addition of MjFH β to Fe-S reconstituted MjFH α yielded high levels of activity. Kinetic parameters for activity on malate and fumarate were obtained by fitting v vs [S] plots to substrate inhibition equation (eq. 1) (Fig. 1.7A). K_M and V_{max} values derived from the fits are summarized in Table 1.4. The K_M values for MjFH are higher compared to other class-I two subunit FH by 2 to 4 folds for malate and 6 to 8 folds for fumarate. The catalytic efficiency (k_{cat}/K_M) of MjFH for both malate and fumarate is comparable to other reported class-I FH. *RS*-2-thiomalate was found to inhibit MjFH with an IC_{50} of $50.8 \pm 1.2 \mu M$ and $20.6 \pm 0.8 \mu M$ with malate and fumarate as substrates, respectively (Fig. 1.7B). Values of the apparent inhibitory constant K_i (app) for *RS*-2-thiomalate derived from the IC_{50} values with malate and fumarate as substrates were $30.8 \mu M$ and $17.4 \mu M$, respectively, and similar to those previously reported for Class-I FH (de Pádua et al., 2017; Feliciano et al., 2019; Jayaraman et al., 2018).

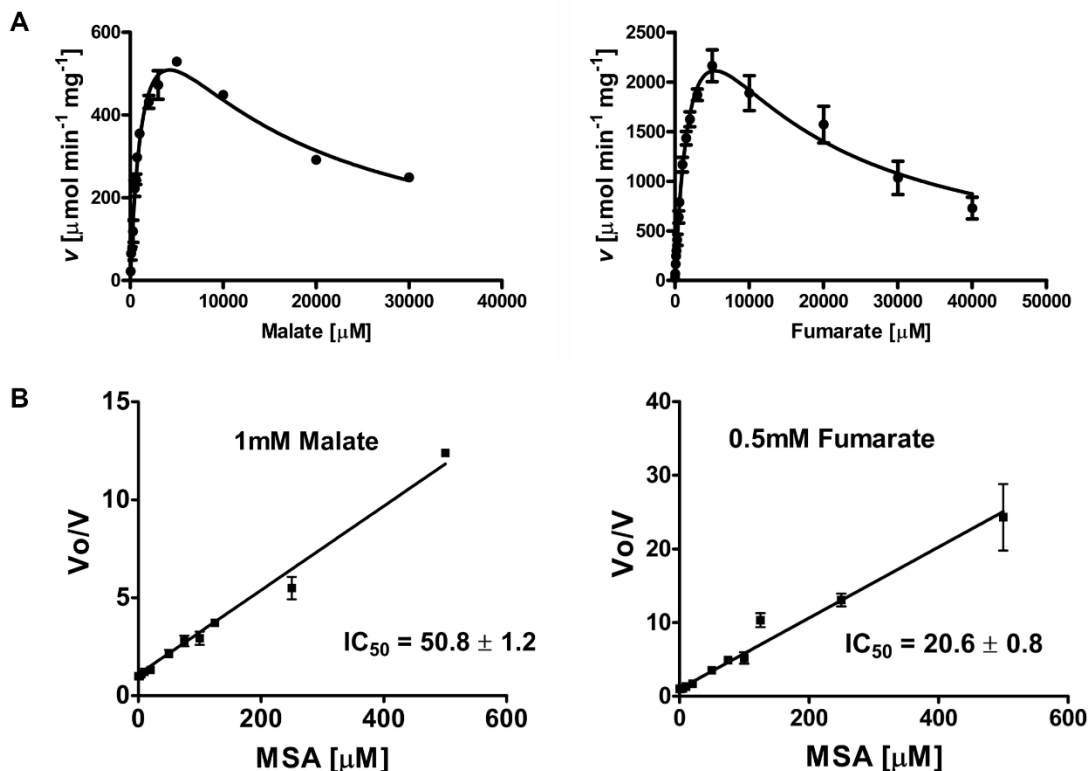


Fig. 1.7. Kinetics of WT MjFH and IC_{50} plot for MSA on substrates malate and fumarate. A. Initial velocity as a function of substrate concentration for malate (left panel) and fumarate (right panel). Data are fit to substrate inhibition equation (Eq. 1). B. IC_{50} plot for RS-2-thiomalate with malate (left panel) and fumarate (right panel) as substrates.

1.3.2. Structure of MjFH

Attempts were made to crystallize individual subunits of MjFH and the complex in both apo- and holo-protein state. In all conditions attempted, only MjFH β and MjFH $\alpha\beta$ apo-protein complex crystallized. MjFH β subunit structure had been solved in the laboratory by the molecular replacement method using the structure of the β subunit of *Archaeoglobus fulgidus* (PDB ID: 2isb) and deposited (5dni). Structure of MjFH $\alpha\beta$ apo-protein complex was solved by single-wavelength anomalous dispersion (SAD) using crystals of the selenium incorporated protein and refined to 2.45 Å resolution (Table. 1.3).

1.3.2.1. Features of the enzyme structure.

The asymmetric unit in MjFH $\alpha\beta$ complex contains one copy each of α - and β -subunits and the functional form of the enzyme, which is a dimer of a heterodimer was generated using a symmetry

related mate from the neighboring unit cell (**Fig. 1.8A**). MjFH β structure (5dni) superposed well on the β -subunit of MjFH $\alpha\beta$ apo-protein complex with an RMSD of 0.528 Å. A 14 residue stretch at the C-terminal end of the protein that was disordered in the MjFH β structure ordered into an α -helix in the MjFH $\alpha\beta$ complex (**Fig. 1.8B**). Ordering of the 14-residue α -helical stretch is aided by a helix-helix interaction with the α -subunit (**Fig. 1.8C**). This 14-residue helical stretch is absent in the class-I single subunit *Leishmania major* FH (LmFH) and may possibly enable tight binding of the two subunits. LmFH structure is shown to have a previously unidentified $\alpha + \beta$ fold and searching MjFH α in DALI server to identify the fold, picked LmFH structure as the only strongest match with an RMSD of 2.1 Å. The SCOP database (Fox et al., 2014) classifies the fold of MjFH β as a “swiveling $\beta/\beta/\alpha$ ” similar to the C-terminal domain (CTD) in LmFH, which is considered to be a highly mobile segment in multi-domain proteins. Surface electrostatic potential of MjFH reveals a large positively charged active site cavity between the α - and β -subunits of MjFH akin to that between N-terminal domain (NTD) and CTD of LmFH (**Fig. 1.8D**). A tunnel originating from a cavity at the top of the protein, observed in LmFH as well, passes through the entire breadth of the protein without connecting to the active site cavity (**Fig. 1.8E**).

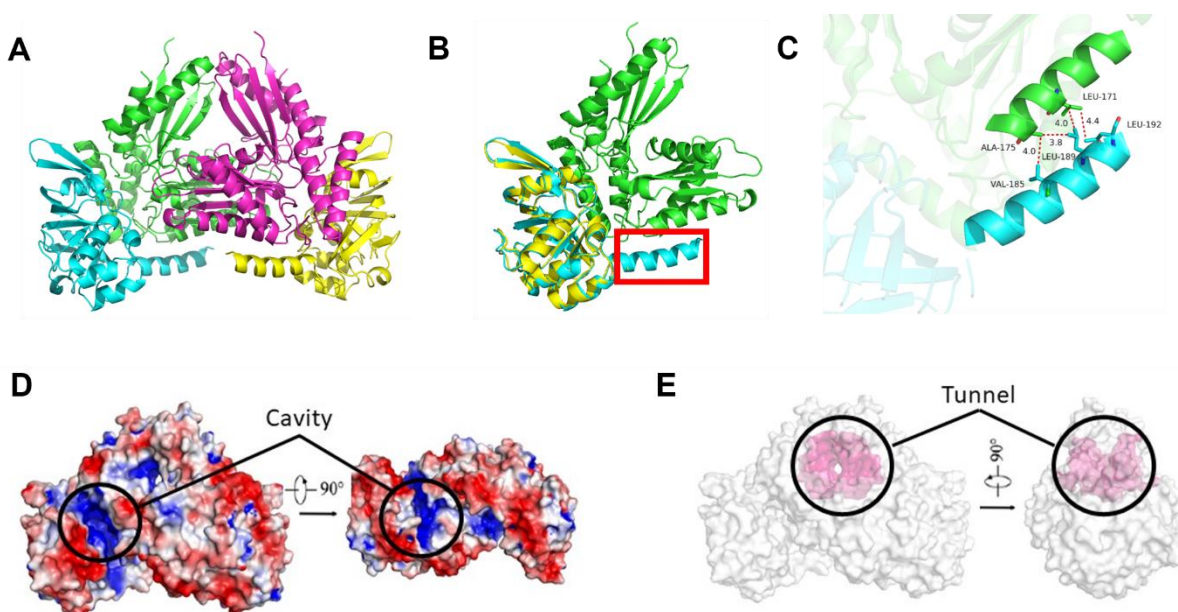


Fig. 1.8. Structure of MjFH. **A.** Structure of MjFH $\alpha\beta$ biological assembly of tetramer. α -subunits are in green and magenta colour, while β -subunits are in cyan and yellow colour. **B.** Superposition of MjFH β structure (5dni) on MjFH $\alpha\beta$ dimer reveals ordering of the C-terminal α -helix in β -subunit of MjFH $\alpha\beta$ complex (enclosed in red box), which is disordered in MjFH β structure. **C.** Helix-helix interaction between the C-terminal helix in β -subunit and helix ($\alpha 4$) from the α -subunit of the protein complex. The β -subunit helix is coloured blue and the α -subunit helix green. **D.** Surface electrostatic potential of MjFH $\alpha\beta$ tetramer.

A positively charged cavity is located between the interface of α and β -subunits creating the active site pocket. E. Protein tunnel. Highlighted in pink is a tunnel that is formed between the α -subunits of MjFH that passes through the entire breadth of the protein.

1.3.2.2. Global structural differences between Mj and LmFH.

To understand structural similarities and differences between MjFH and LmFH, the structures of the two proteins were superposed. Although the MjFH α - and β -subunits individually superpose well with the N-terminal and C-terminal domains of LmFH with an RMSD of 1.8 Å and 1.4 Å, respectively (**Fig. 1.9A, B**), structural superposition of the biological assemblies of the two proteins showed gross differences with an RMSD of 4.95 Å (**Fig. 1.9C**). A visual inspection revealed that LmFH structure is more compact compared to MjFH and measuring the distance between conserved residues located at the extreme termini of both proteins showed that the α -subunits and N-terminal domains of MjFH and LmFH are equidistant, while the β subunits in MjFH are spaced farther apart compared to C-terminal domains of LmFH (**Fig. 1.9D**). In LmFH, the CTD has been shown to be mobile and its mobility changes with substrate/inhibitor binding. The B factor values for the swiveling β -subunit, as expected is higher than that of the α -subunit of the protein (**Fig. 1.9E**). A major proportion of residues at the NTD/NTD or α/α dimer interface of class-I FH are found to be conserved (Feliciano et al., 2016) and an interface analysis reveals that LmFH has a larger number of interactions between the domains of the protein possibly accounting for its compact structure. While the NTD/NTD interface in LmFH is stabilized by 42 hydrogen bonds with an overall interface interaction area of 3712 Å², the α/α subunit interface in MjFH is stabilized by 34 hydrogen bonds with an interface interaction area of 2521 Å². Similarly, 19 hydrogen bonds stabilize NTD-CTD domains of LmFH with an interface area of 1503 Å² while only 5 hydrogen bonds are observed between the α and β subunits of MjFH with an interface area of 1275 Å². 5 residues are involved in van der Waals interactions between the CTDs of LmFH but the β subunits of MjFH do not form any contacts (**Fig. 1.9F**). To understand if there are any collective functional motions in the protein and compare between the LmFH holo-protein and MjFH apo-protein, normal mode analysis (NMA) was carried out on iMODS server (López-Blanco et al., 2014). This reveals that the β -subunit in MjFH is indeed dynamic capable of inward shift by “swinging-in” while LmFH C-terminal domain can displace outward by “swinging-out” (**Fig. 1.9G**).

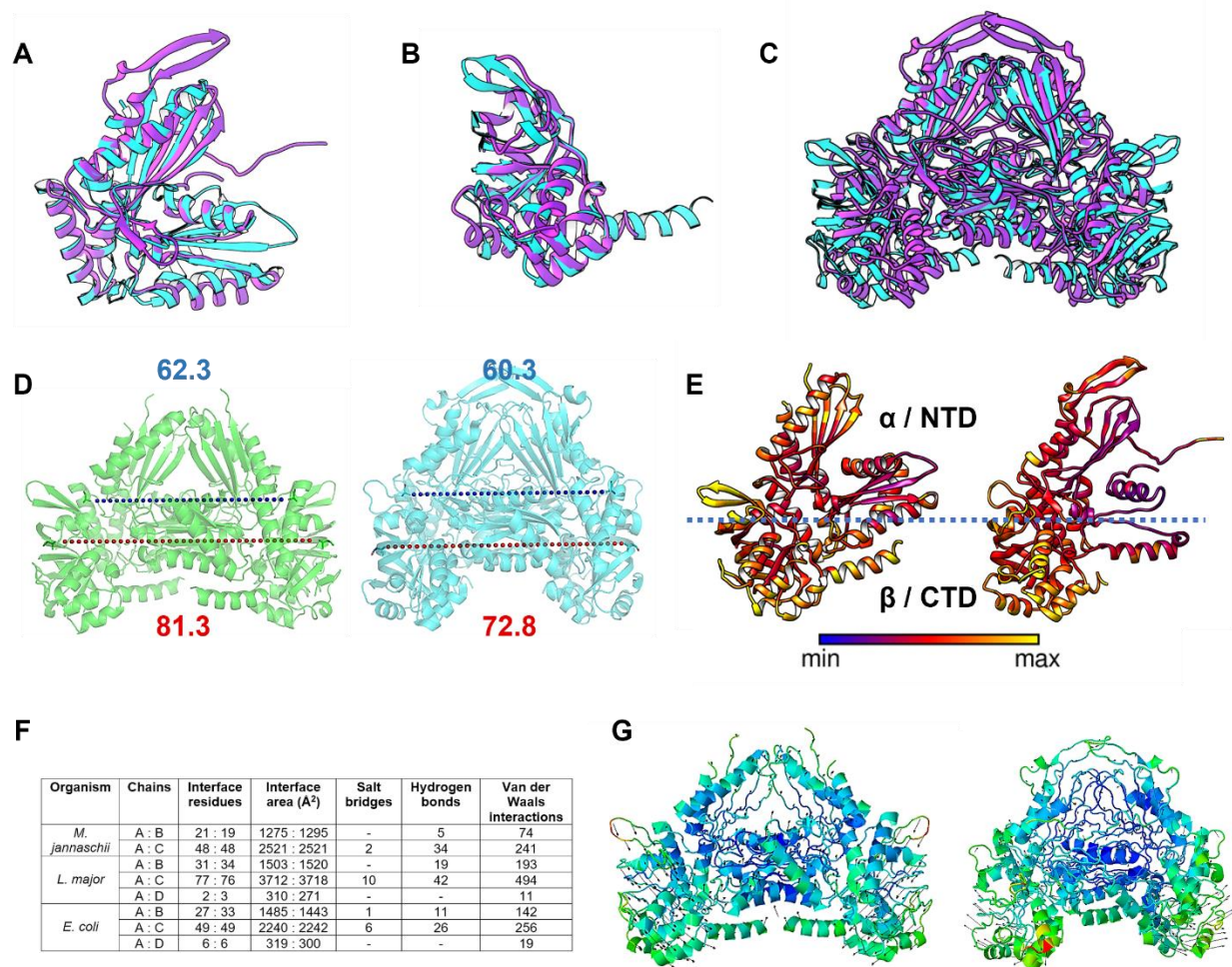


Fig. 1.9. Differences between apo-MjFH and holo-LmFH (5l2r) structures. Superposition of LmFH NTD on MjFH α (A) and LmFH CTD on MjFH β (B) show good alignment with RMSD values of 1.8 Å and 1.4 Å, respectively. C. Overlay of LmFH functional complex on MjFH shows gross structural differences and a lack of clear alignment, with an RMSD of 4.95 Å. D. Distance between conserved residues A52 and A125 located at the extreme ends of MjFH α (green) and LmFH NTD (blue) dimers show similar spacing, while distance between conserved residues V160 and V546 of MjFH β and LmFH CTD show that β -subunits of MjFH are more distantly spaced compared to CTD of LmFH (Residue numbering from LmFH). E. B-factor analysis reveals that MjFH β and LmFH CTD have higher B-factor values compared to the overall average of the protein (we should indicate α and β and NTD and CTD on the figure for clarity). F. Comparison of subunit/domain interfaces between MjFH and LmFH. Chains A and C refer to α -subunit of MjFH and NTD of LmFH, and chains B and D refer to β -subunit of MjFH and CTD of LmFH. LmFH reveals a larger area of interface interactions compared to MjFH. G. Normal mode analysis. β -subunit of MjFH (left panel) is the dynamic segment capable of an inward shift through a “swinging-in” movement while in LmFH the CTD (right panel) displaces outward through a “swinging-out” motion. Red regions in the protein are prone to larger scale movements followed by green and then blue. The direction of movement is displayed as arrow marks on the respective regions.

1.3.2.3. Differences in active site architecture between Mj and LmFH.

Transposing the Fe-S cluster from LmFH onto MjFH reveals that three cysteine residues, Cys60, Cys182 and Cys269 in MjFH are well positioned to bind to the cluster (**Fig. 1.10A**). LmFH active site has 9 residues hydrogen-bonding with the substrate at the active site. 6 of these residues are positioned similarly in MjFH apo-protein active site, while three other residues (Gln61 and Arg102 from the α -subunit and Arg⁸⁴ from β -subunit) adopt a different side chain rotameric conformation going to show that the active site is largely pre-organized even in the absence of cluster and substrate (**Fig. 1.10B**).

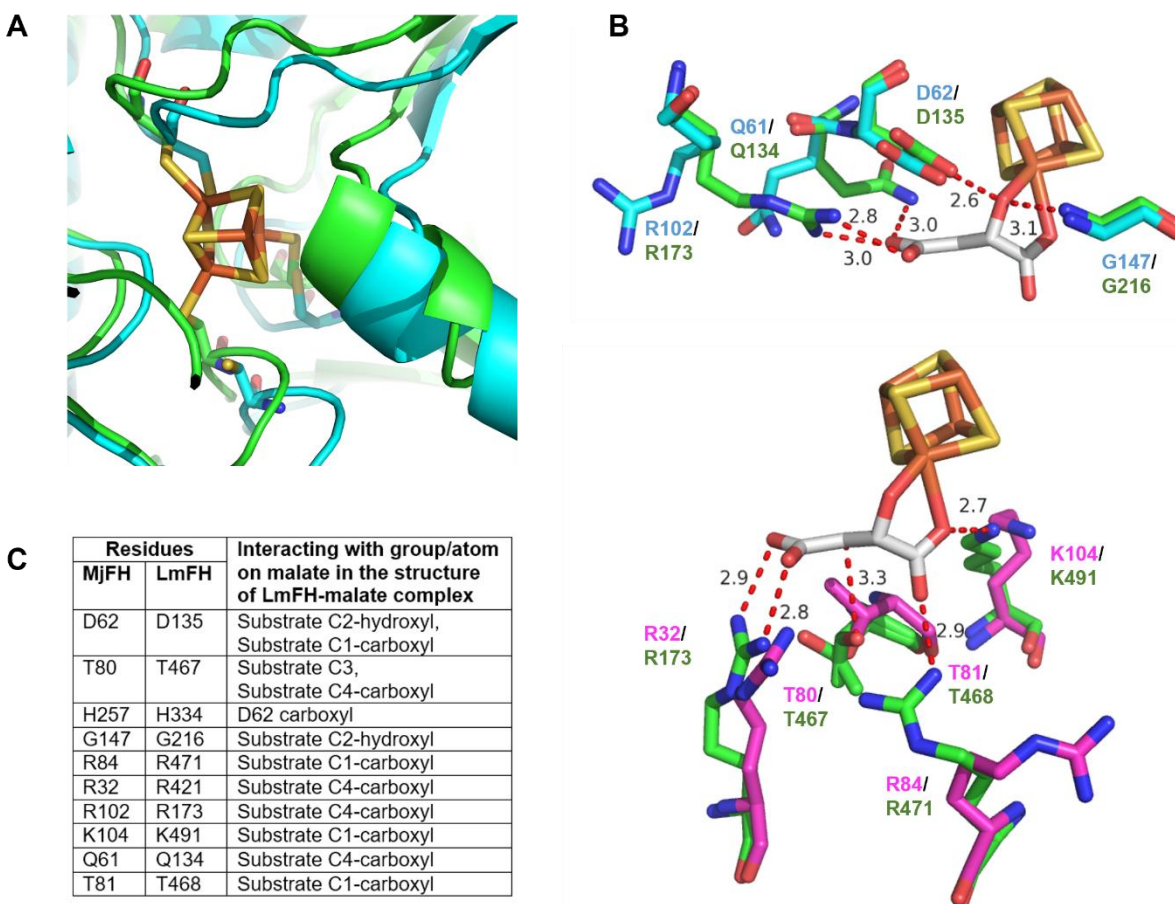


Fig. 1.10. Active site comparison between apo-MjFH and holo-LmFH (5I2r). **A.** Overlay of Fe-S cluster from LmFH onto MjFH shows that the active site pocket with three cysteine residues are well poised to accommodate the cluster. **B.** Active site residues in MjFH compared to LmFH. Top panel shows the superposition of active site residues from MjFH α (blue) on LmFH NTD (green) and bottom panel shows the superposition of active site residues from MjFH β (magenta) on LmFH CTD (green). Malate bound to the Fe-S cluster is highlighted in grey. Residues Gln61 and Arg102 from the α -subunit and Arg84 from β -subunit of MjFH show different side chain rotameric conformation. **C.** Numbering of the active site residues of MjFH and LmFH. Note that the contacts with the atoms on the substrate are from the LmFH structure.

1.3.2.4. Binding of cluster and substrate compacts quaternary fold

Inter-CTD distance measured for both apo-MjFH and holo-LmFH shows that these values are larger for MjFH compared to LmFH (**Fig. 1.9D**) as can also be visually inferred from the superposition of the two structures (**Fig. 1.9C**). To identify the molecular determinants of this difference, atomistic molecular dynamics simulation was carried out on both apo- and holo- forms of MjFH and LmFH (*All modeling and MD simulation studies were carried out by Sudharshan Behera from Prof. Balasubramanian Sundaram lab at CPMU, JNCASR*). For this, the missing polypeptide segments/residues/atoms of missing regions of the protein for both MjFH and LmFH structures were first modelled. 63 residues at the N-terminal end, the 10-residue linker between the NTD and CTD, and 112 atoms from the 42 residues that are missing in the LmFH structure were modelled before simulation. Similarly, one and two residues from the N- and C-termini, respectively along with few other atoms that are missing in the MjFH structure were also modelled. Force field parameters for the Fe₄S₄ cluster and the bound cysteine residues were taken from a previous report (Chang & Kim, 2009). As multiple attempts at obtaining a stable contracted form of holo-MjFH did not yield satisfactory results, ligand was removed from the LmFH structure and simulated. The Rg and inter-CTD values reach and converge close to that of apo-MjFH by around 300 ns time point beyond which the values saturated (**Fig. 1.11A, B**). A representative structure for the expanded form of apo-LmFH obtained from the simulation, between 300 to 400 ns time frames when superposed on apo-MjFH shows the two structures superpose well post simulation, confirming that class-I FH do indeed form a stable expanded/relaxed structure in the absence of Fe-S cluster and substrate (**Fig. 1.11C**).

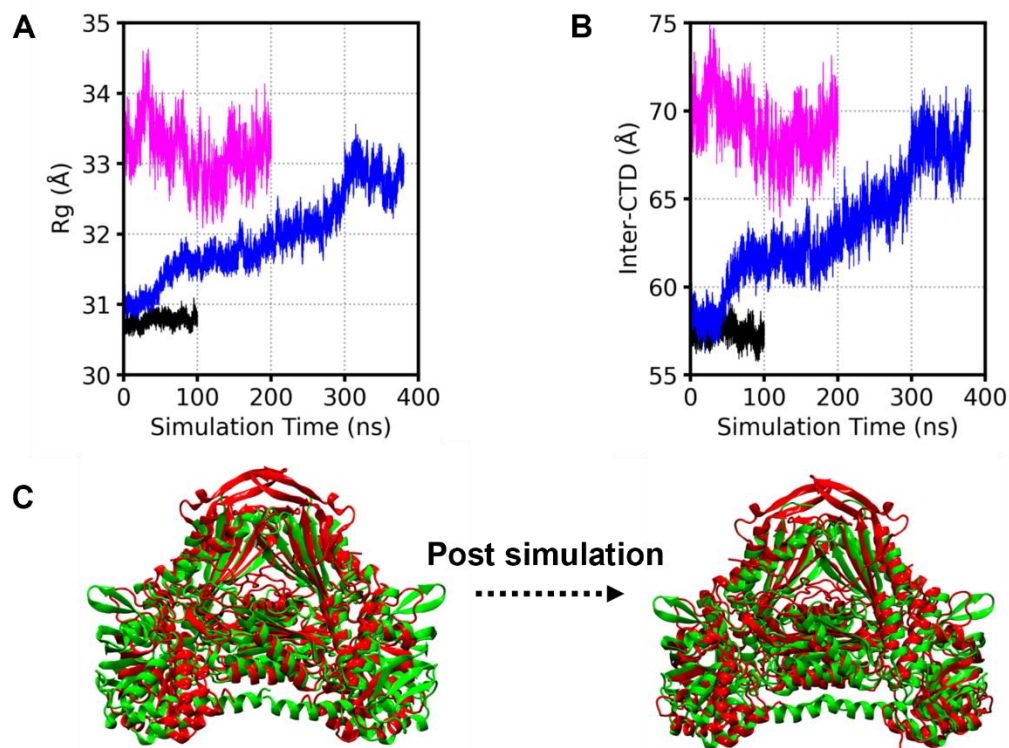


Fig. 1.11. MD simulation reveals class-I FH adopts two different states based on presence or absence of a bound Fe-S cluster and substrate. A. Rg and B. inter-CTD distance as a function of simulation time for holo-LmFH (black), apo-LmFH (blue), apo-MjFH (magenta). There is no change in the Rg and inter-CTD distance values in the simulation of holo-LmFH. The apo-LmFH expands, and both the Rg and inter-CTD distance values become close to the ones in apo-MjFH. C. Superposition of the crystal structure of apo-MjFH (green cartoon) on crystal structure of holo-LmFH (red cartoon, 5l2r) before and after 300 to 400 ns simulation of LmFH structure. The two structures superpose well post simulation of apo-LmFH structure, affirming that the enzymes adapt two different states in the presence or absence of a Fe-S cluster and bound substrate. Images generated by Sudharshan Behera from Prof. Balasubramanian Sundaram lab at CPMU, JNCASR

1.3.3. Mutational analysis of the active site residues.

To understand the catalytic mechanism of class-I FH, residues were selected for site-directed mutagenesis in MjFH, which according to the *S*-malate bound LmFH structure and more recently solved inhibitor – thiomalate bound structure interact with the substrate. A set of 10 residues are located at the active site of which 9 residues are in hydrogen bonding distance with *S*-malate (Gln61, Asp62, Arg102, Gly147 from α subunit and Arg32, Thr80, Thr81, Arg84, Lys104 from β subunit) (**Fig. 1.10B**). The residue/s closest to C3 carbon of *S*-malate is Thr80 at 3.34 Å and to C2 hydroxyl group are Asp62 and Gly147 at 2.64 Å and 3.1 Å distance, respectively. Another residue, His257 from the α -subunit, although not interacting with the substrate, is hydrogen-bonded to

Asp62 at a distance of 2.81 Å and could stabilize the charge state of this residue. All other residues form H-bonds with the carboxyl groups of *S*-malate. Sequence alignment also shows that Gly147 is part of a motif ‘KGXGS’ that is highly conserved in class-I FH (**Fig. 1.12**) and considering their proximity to the substrate, Lys144 and Ser148 were also chosen for mutation. To examine their function, all these 12 residues were replaced by site-directed mutagenesis. Mutant enzymes were generated and purified using protocols similar to that for the WT protein. Steady state kinetics were conducted under conditions similar to that used for WT and the kinetic parameters along with fold change in catalytic efficiency relative to WT are summarized in **Table. 1.4**. Mutants with measurable activities showed a small drop in K_M value, between 1.5 to 3 folds for both substrates, which is close to that observed in LmFH mutants. All inactive mutants showed a characteristic CD spectrum in the near-UV/visible region confirming that Fe-S cluster assembly is intact in these mutants (**Fig. 1.13**).

	140	150
Methanocaldococcus_janschii	E I I A F P	K G A G S E N M S A L
Leishmania_major	E F L F I A	K G G G S A N K A Y L
E_coli	K F L C I A	K G G G S A N K T Y L
Plasmodium_falciparum	E L I F I A	K G G G S A N K T F L
Trypanosoma_brucei	E F F F V A	K G G G S A N K A F L
Pyrococcus_furiosus	R I S I L I	K G G G S E N C S A L
Pelotomaculum_thermopropionicum	R I I V A P	K G G G S E N M S A I
MPOB	K F L F V A	K G G G S A N K N Y L
Burkholderia_xenovorans	D V Q V A A	K G G G S E N K S K F
Archaeoglobus_fulgidus	R M V V M P	K G A G S E N V S A L
Methanohalophilus_halophilus	R I T A V P	K G A G S E N M S I L
Shigella_sonnei	K F L C I A	K G G G S A N K T Y L
Strigomonas_culicis	K F M F V A	K G G G S A N K T Y L
Entamoeba_histolytica	K F L F C V	K G G G S A N K S Y L
Perkinsus_marinus	K F H V I A	K G G G S A N K F Q L
Blastocystis_sp	K F M F M A	K G G G S A N K S Y L
Hymenolepis_microstoma	N F L F M A	K G G G S A N K S F L
Thecamonas_trahens	N F M F M A	K G G G S A N K T F L
Echinococcus_granulosus	S F L F M A	K G G G S A N K S F L
Angomonas_deanei	D F L F I A	K G G G S A N K A Y L
Salmonella_typhimurium	T I T V M P	K G G G S E N M G T F
Helicobacter_hepaticus	H L K V C P	K G F G S E N K S V L
Thermococcus_sp	K I A I L P	K G G G S E N C S A L
Methanonatronarchaeum_thermophilum	K L T V L I	K G G G S E N V A R Q
Thermoproteus_tenax	Q F T Y V P	K G G G S E L P G K A

Fig. 1.12. Motif KGXGS is identical and conserved across all class-I FH.

Table. 1.4. Kinetic parameters of MjFH WT and mutant enzymes

MjFH mutants	Corresponding residues in LmFH	Malate					Fumarate				
		V_{max} ($\mu\text{mol}^{-1} \text{ minute}^{-1} \text{ mg}$)	K_M (μM)	K_I (μM)	k_{cat}/K_M ($\text{s}^{-1} \text{M}^{-1}$)	Fold change in catalytic efficiency	V_{max} ($\mu\text{mol}^{-1} \text{ minute}^{-1} \text{ mg}$)	K_M (μM)	K_I (μM)	k_{cat}/K_M ($\text{s}^{-1} \text{M}^{-1}$)	Fold change in catalytic efficiency
WT	-	882.7 \pm 54.9	1554 \pm 169.4	11539 \pm 1466	5.1 $\times 10^5$	-	4309 \pm 384.5	2785 \pm 397.4	10320 \pm 1548	1.4 $\times 10^6$	-
D62V	D135	15.7 \pm 2.1	1030 \pm 237.5	13750 \pm 4377	1.37 $\times 10^4$	37.2	45.32 \pm 4.9	807.6 \pm 153.9	8839 \pm 1951	5 $\times 10^4$	28
T80V	T467	4.7 \pm 0.8	1632 \pm 388.5	6453 \pm 1840	2.6 $\times 10^3$	196.1	11.3 \pm 1.5	1713 \pm 324.4	4151 \pm 858	5.97 $\times 10^3$	234.5
T80C	T467					Inactive					
H257N	H334	11.8 \pm 0.9	589.9 \pm 85.9	12247 \pm 2157	1.8 $\times 10^4$	28.3	118 \pm 13.1	3343 \pm 558.8	11451 \pm 2401	3.19 $\times 10^4$	43.8
D62V+H257N	D135+H334					Inactive					
G147A	G216					Inactive					
K144L	K213	6.5 \pm 0.3	446.2 \pm 50.5	35077 \pm 7308	1.3 $\times 10^4$	39.2	17.2 \pm 0.9	556.7 \pm 72.4	24571 \pm 4843	2.8 $\times 10^4$	50
S148A	S217	333 \pm 29.5	680.2 \pm 121.5	21790 \pm 5099	4.4 $\times 10^5$	1.1	1286 \pm 275.5	1755 \pm 517	4605 \pm 1531	6.63 $\times 10^5$	2.2
R84K	R471	34.5 \pm 2.8	749.4 \pm 109.9	10986 \pm 1958	4.1 $\times 10^4$	12.4	108.6 \pm 6.5	1063 \pm 124.7	19872 \pm 3426	9.2 $\times 10^4$	14.2
R32K*	R421	S.A		0.4625 \pm 0.06			S.A		1.3615 \pm 0.23		
R102K*	R173	1mM Malate		1.115 \pm 0.1			500 μM fumarate		0.2985 \pm 0.02		
K104L*	K491			0.825 \pm 0.001					0.937 \pm 0.03		
Q61L	Q134					Inactive					
T81V	T468					Inactive					
R32L	R421					Inactive					
R84L	R471					Inactive					
R102L	R173					Inactive					

* Since these mutants were weakly activity, only specific activity values are reported.

Footnote: K_I refers to inhibitory constant arising from substrate inhibition.

Representative time course curves for the mutants as a comparison with WT curves are provided in Appendix B.

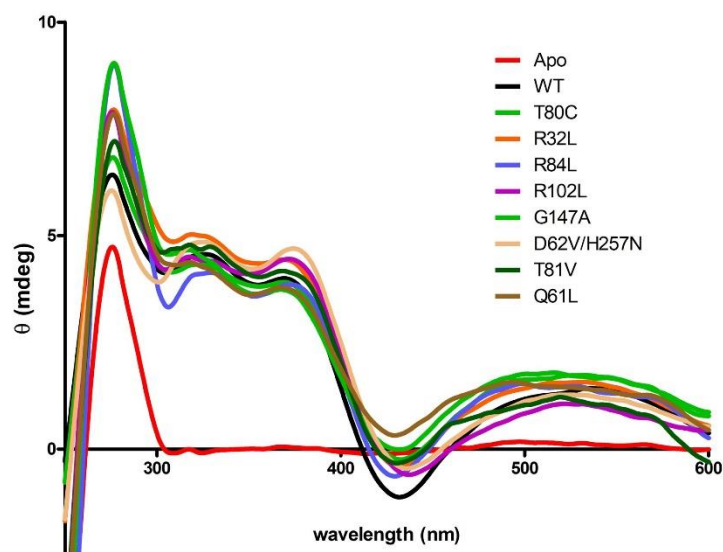


Fig. 1.13. Circular dichroism spectra of apo- and Fe-S reconstituted wild type and mutants of MjFH.

1.3.3.1. Catalytic acid and base mutants retain activity.

Unlike LmFH, where mutation of residues corresponding to threonine 80 and aspartate 62, the proposed catalytic base and acid, respectively led to a dramatic drop (2500-5500-fold) in catalytic efficiency, MjFH T80V and MjFH D62V retained significant activity. Mutant T80V of MjFH showed 196- and 235-fold drop in catalytic efficiency for malate and fumarate, respectively while D62V showed a 37- and 28-fold reduction for the two substrates. Mutation of threonine to cysteine however, led to T80C being inactive. A similar trend has been observed in adenylosuccinate lyase,

where mutating the serine residue believed to be the catalytic base to alanine led to 1000-fold drop in specific activity while mutating it to cysteine abolished activity. It is suspected that the highly reactive thiol group of cysteine disrupts the active site by geometric and electrostatic perturbations (Bulusu et al., 2009). A close inspection of the active site of substrate bound LmFH structure revealed that the threonine residue has occupancy in two different states; one where it is in direction of the substrate C3 carbon and other where it points towards a structural water molecule (**Fig. 1.14**). The structural water molecule is conserved in several solved structures of LmFH bound to different substrate analogues. The water molecule is in hydrogen bonding distance from one of the arginine residues (Arg32) and at 4.4 Å from the C3 carbon of the substrate. In the absence of the catalytic base threonine residue, water molecule in its protonated/deprotonated state mediated by the neighboring Arg³² residue can potentially take over the role of catalytic base, explaining why threonine mutant retained activity. Mutation of the histidine residue in contact with the predicted catalytic acid to asparagine in MjFH (H257N) also led to a 28- and 44-fold reduction in catalytic efficiency for both the substrates. A double mutant of D62V and H257N of MjFH was inactive suggesting the requirement of at least one of these residues at the active site. In the absence of the catalytic acid aspartate, its hydrogen bonded partner could possibly replace its role.

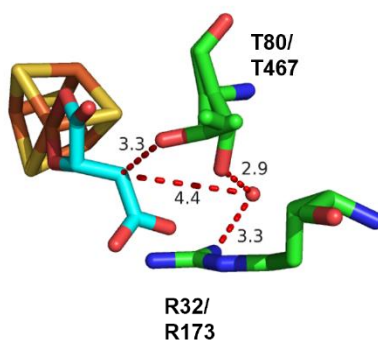


Fig. 1.14. A conserved water molecule could potentially take over the role of catalytic base in the absence of the threonine residue.

1.3.3.2. Mutation of residues involved in electrostatic interactions with the substrate led to loss of activity.

Leaving aside the catalytic acid and base residues, mutation of all other residues at the active site that are in close contact with the carboxyl groups of the substrate to a residue with non-polar side chain led to inactive mutants except for the mutant MjFH K104L that exhibited feeble activity

precluding estimation of K_M and V_{max} values. Two arginine residues and one glutamine (Arg32, Arg102 and Gln61) clamping the substrate from one side (**Side A, Fig. 1.15**) and one lysine, arginine, and threonine residue (Lys104, Arg84 and Thr81) from the other (**Side B, Fig. 1.15**), charge-polarize the substrate with stronger hydrogen bonding network mediated by 4 strong bonds formed by two arginine residues on Side A compared to 2 bonds formed by one arginine and one lysine residue on Side B. As the charged side chains of these residues could essentially have a role in catalysis by polarizing the substrate and priming it for reaction, the three arginine residues were mutated to lysine and the mutants probed for activity. Both the arginine residues on side A, which form stronger hydrogen bonds with substrate, on mutating to lysine (R32K and R102K) retained feeble activity precluding estimation of kinetic parameters, while mutating the arginine on side B to lysine (R84K) led to a 10-fold drop in catalytic efficiency for both substrates. Mutating the charged residues on side A was more detrimental compared to side B.

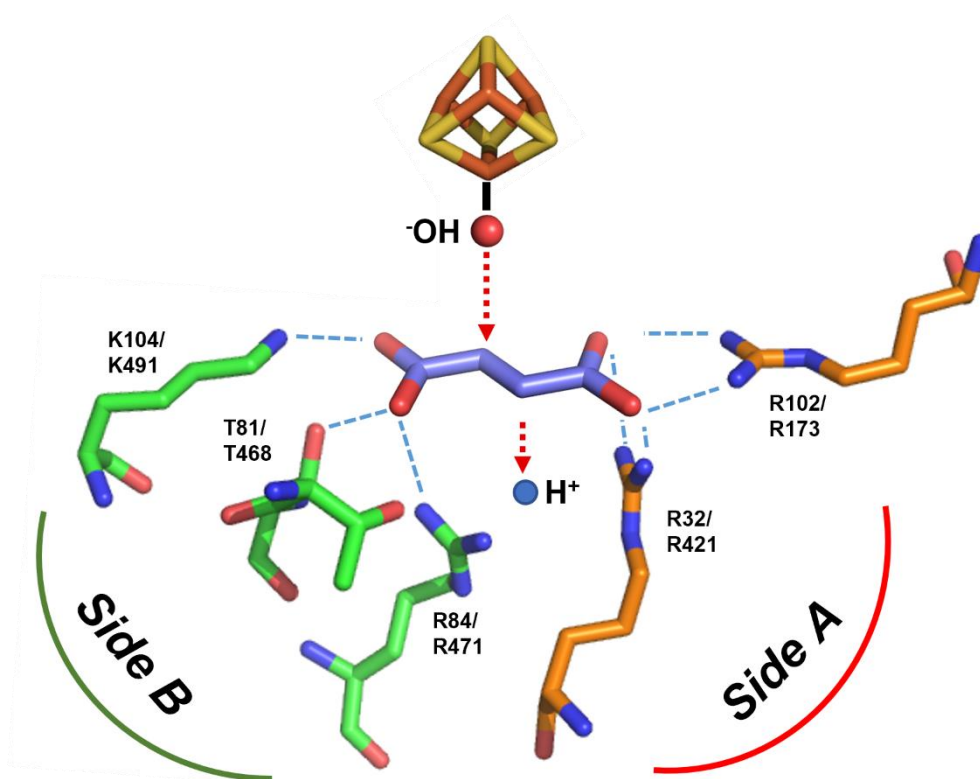


Fig.1.15. Substrate is charge-polarized at the active site with more charge-interactions on one side compared to other. Residues Arg102 and Arg32 on side A form 4 strong hydrogen bonds with carboxyl group of bound substrate compared to one bond each formed by Lys104 and Arg84 on side B.

1.3.3.3. The conserved motif – KGXGS has a role in maintaining the environment around the Fe-S cluster

While mutant K144L from the KGXGS motif showed a drop in catalytic efficiency of 39- and 50-fold for malate and fumarate as substrates, S148A showed no drop in catalytic efficiency for malate and a 2-fold drop for fumarate as substrate. MjFH G147A mutant from the motif was inactive. The backbone amide NH of G147 is at 3.1 Å from the C2 hydroxyl group of malate in the malate-bound structure of LmFH and mutating it to alanine could lead to the side chain occupying the active site space, resulting in short contacts that occlude substrate binding (**Fig. 1.16**). The K144L mutant where K144 is part of the KGXGS motif had the highest drop in K_M values by 3.5 and 5 folds, respectively for malate and fumarate suggesting a possible role in substrate binding.

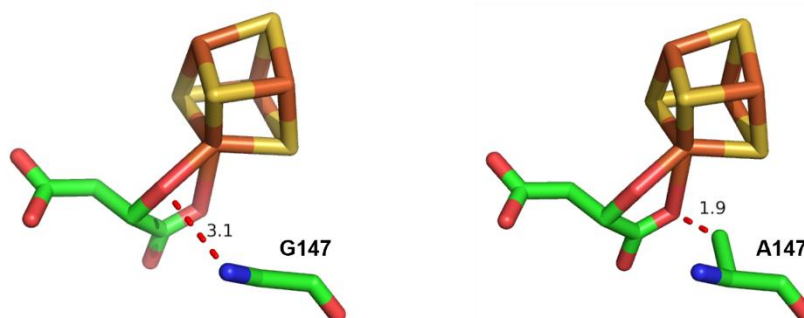


Fig. 1.16. Mutation of Gly147 to Ala will lead to the side chain methyl group coming in close contact with the carboxylate group of the bound substrate, thus possibly occluding the binding of the substrate in a catalytically active fashion.

1.3.4. QM studies gives insights into the mechanism of catalysis in class-I FH

Information derived from the kinetic properties of WT and active site mutants of MjFH along with the structure revealing the differential localization of charges on either side of the substrate, clearly suggests that electrostatics at the active site play a vital role in catalysis. In this scenario, the substrate possibly gets charge polarized during the catalytic cycle aiding easy addition/removal of a water molecule. Density function theory (DFT) calculations can effectively help model complex catalytic reaction pathways (Lee, 2021), which would provide valuable information on the importance of the active site residues in MjFH. This apart, natural population analysis (NPA) helps understand the electron distribution in compounds of high ionic character (Reed et al., 1998). By resorting to both these calculations, the magnitude of the role of substrate charge polarization in catalysis mediated by class-I FH can be distinctly established. Although in aqueous medium, the thermodynamics for malate to fumarate conversion has been worked out (Bearne & Wolfenden,

1995), the reaction mechanism has not been fully explored. Reaction mechanism in aqueous medium and at the active site of class-I FH WT and catalytic acid/base mutants was probed using the X-ray crystal structure of Fe-S cluster bound LmFH-malate complex (5l2r). The same model was used to understand substrate charge polarization using NPA calculations (*All calculations were carried out by Soumik Das from Dr. Garima Jindal's laboratory at the Department of Organic Chemistry, Indian Institute of Science*).

1.3.4.1. Fumarate hydration in aqueous solution proceeds through the carbanion intermediate

Interconversion between fumarate and malate can proceed through three different mechanisms (**Fig. 1.17**); pathway w1 involving the carbanion, pathway w2 involving a carbocation, and pathway w3 which involves a concerted TS (transition state). In pathway w1, the first step of malate to fumarate conversion involves the proton abstraction by a water molecule via TS(A-B)_{w1} . The second step involves the removal of a water molecule via TS(B-C)_{w1} . In pathway w2, the first step involves a water removal as shown in TS(A-B)_{w2} whereas, the second step involves proton abstraction via TS(B-C)_{w2} as opposed to pathway w1. For pathway w3, the conversion happens through a single TS of water removal. Study revealed that the carbanion pathway, *i.e.*, pathway w1 proceeds with the lowest energy as seen from the relative free energies provided in parentheses in **Fig. 1.17**. It presented an activation free energy barrier of 41.7 kcal mol⁻¹ which is in accordance with the experimental value of 36 kcal mol⁻¹ (Bearne & Wolfenden, 1995).

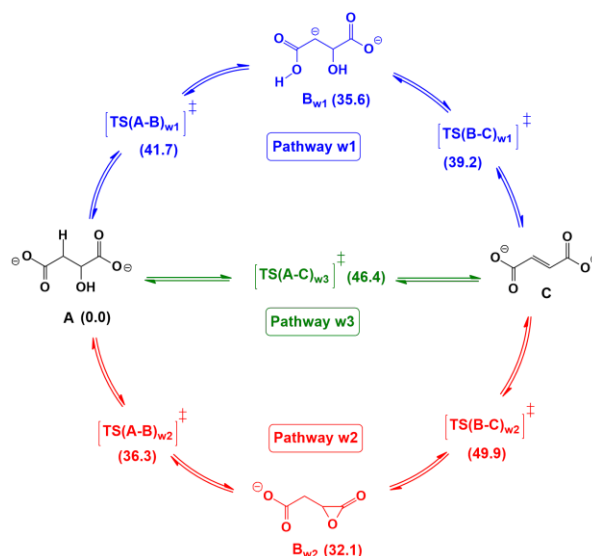


Fig. 1.17. Three pathways of interconversion between malate and fumarate in aqueous medium. Values in parentheses are the relative electronic energies in kcal mol⁻¹. The energy shown for TS(A-C)_{w3} involves

an external water molecule, concerted TS with no external water molecule was found to be even higher in energy.

1.3.4.2. Reaction mechanism at the enzyme active site

Akin to the aqueous medium, at the active site of the enzyme as well, three possible mechanistic pathways of interconversion were probed. Studies carried out with a cluster of 86 atoms (QM1), with only five essential amino acid residues (Asp135, His334, Thr467, Thr468 & Arg471) which are believed to play an active role in the catalytic cycle showed an energy barrier of 21.2 kcal mol⁻¹ which is higher than the experimentally known value of 13.8 kcal mol⁻¹ (Feliciano & Drennan, 2019). Using a larger cluster of 207 atoms (QM2), consisting of five additional amino-acid residues, which interact non-covalently with the bound substrate via hydrogen bonding (Gln134, Arg173, Gly216, Arg421 and Lys491) along with a crystal water molecule reduced the free energy to 16.7 kcal mol⁻¹, signifying the importance of non-covalent interaction present at the active site in catalysis. Carbanion pathway turned out to be the most favorable pathway at the enzymatic active site as well.

The catalytic base T467 and catalytic acid D135 at the active site in LmFH, as seen by the H-bonding contacts in the crystal structure, are suitably placed to abstract a proton from C3 and donate a proton to C2 hydroxyl group of malate (**Fig. 1.5**). In TS(A-B)wt, T467 abstracts a proton from the substrate leading to a carbanion intermediate and simultaneously donates its proton to the neighboring R471 residue (**Fig. 1.18**). In TS(B-C)wt, OH⁻ dissociates from the carbanion and at the same time accepts a proton from D135, thereby generating the final fumarate product with the release of a water molecule.

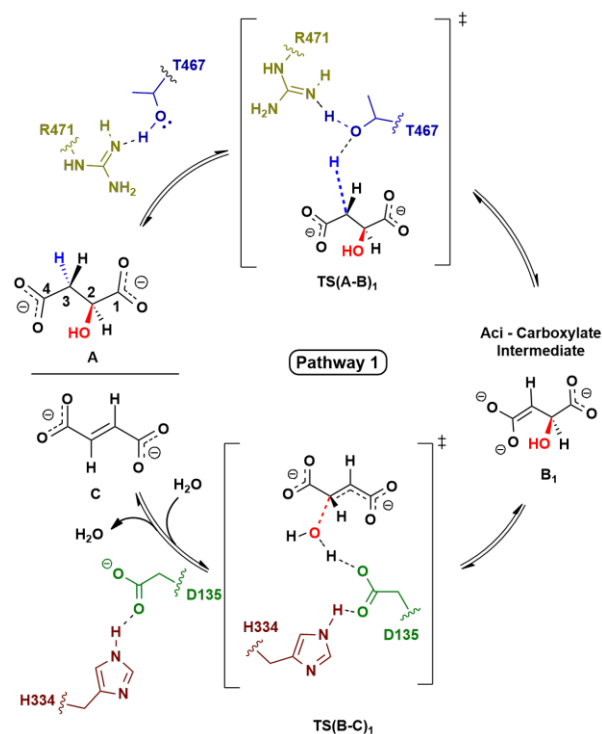


Fig. 1.18. Stepwise mechanism of reversible (de)hydration in the protein environment proceeding through the carbanion pathway.

1.3.4.3. Alternative mechanism in the absence of catalytic residues

Evaluation of effect of mutation of catalytic residues on the reaction mechanism was carried out using mutants T467A (M1) and D135A (M2) in QM2 model. In M1, since Thr467 is mutated to alanine which cannot relay the proton transfer, the obvious choice for an alternate base was the crystal water which can assist the proton transfer from C3 of malate to R471 as shown in TS(A-B)_{M1} (**Fig. 1.19**). Following this, Asp135 plays a similar role as in WT. In the second mutant D135A (M2), where the catalytic acid Asp135 was mutated to alanine, an alternative possibility is that His334 residue exists in its protonated form (H334-H⁺). It can then transfer a proton through a water molecule to the departing OH⁻ eventually eliminating a water molecule from the carbanion via TS(B-C)_{M2} (**Fig. 1.19**). The results obtained from both M1 and M2 using the QM2 model are in line with those obtained from mutational studies of MjFH. The activation free energy barriers for M1 and M2 were 26.4 kcal mol⁻¹ and 23.2 kcal mol⁻¹, respectively (**Fig. 1.21**). A drop in free energy barriers even in the absence of catalytic residues further strengthened the observation that activation energy barrier could be reduced substantially by other residues at the active site apart from the catalytic acid and the base residues.

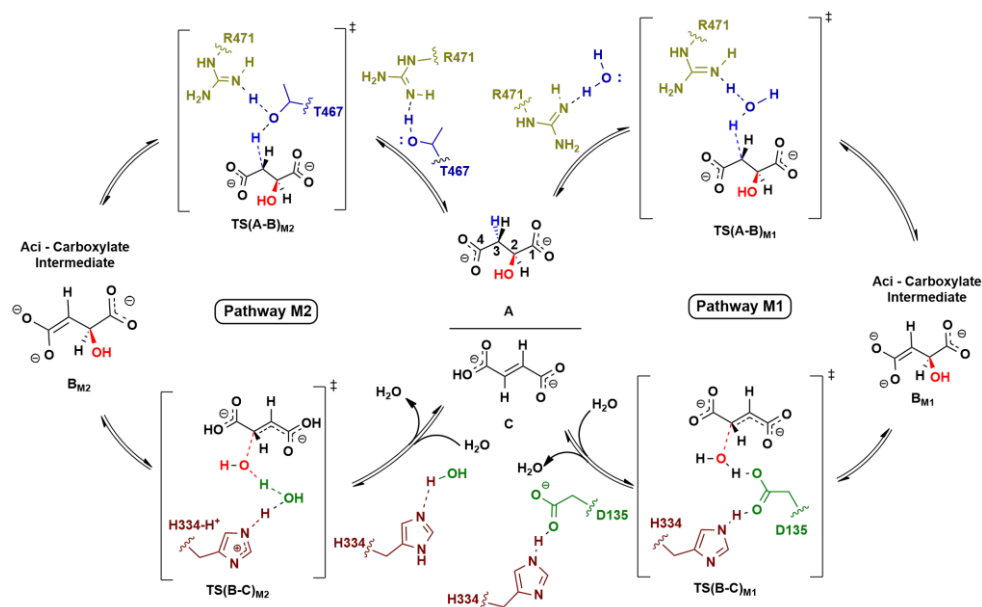


Fig. 1.19. Plausible mechanistic pathway of reversible (de)hydration in T467A mutant (pathway M1) and D135A mutant (pathway M2).

1.3.5. Natural population analysis reveals that the enzyme is indeed charge polarized at the active site

Results from NPA calculation on fumarate bound enzyme-substrate complex in the larger model, QM2 for WT and various mutants are highlighted in **Table 1.5**. From the NPA calculation, it was observed that the charge separation is maximum for WT with K491A and Q134A mutants being the only exceptions. As the active site residues were mutated, the relative charge separation reduced making the fumarate less nucleophilic, which can in turn cause an increase in activation energies as already observed in the case of T467A and D135A mutants. Importantly, the active site arginine residues play a significant role in charge separation. As the three arginine residues are removed sequentially from the active site, the charge separation not only reduces but the polarity completely gets reversed. Nucleophilicity of the substrate totally diminishes as the arginine residues are mutated, which in turn possibly demolishes the enzyme activity. From the good correlation obtained from NPA, it is imperative that non-covalent interactions present at the active site of FH play an essential part in making the enzyme more efficient by polarizing the substrate.

Table 1.5. NPA charges on bound fumarate for different mutants of LmFH

Mutant	NPA Charge on C3	NPA Charge on C2	Relative charge on C2 w.r.t. C3
WT	- 0.306	- 0.212	+ 0.094
R471A	-0.286	-0.222	+0.064
R421A	- 0.265	- 0.231	+ 0.034
R173A	-0.273	-0.212	+0.061
R421A+R173A	- 0.237	- 0.272	- 0.035
R421A+R173A+R471A	- 0.240	- 0.294	- 0.054
K491A	-0.314	-0.198	+0.116
D135A	-0.302	-0.234	+0.068
D135A+H334A	-0.275	-0.234	+0.041
Q134A	-0.318	-0.214	+0.104
T467A	-0.264	-0.228	+0.036
T468A	-0.282	-0.201	+0.081
G216A	-0.284	-0.223	+0.061
Fumarate	-0.285	-0.285	0.000
Fumarate + [4Fe-4S] cluster	-0.245	-0.286	-0.041

1.4. Summary

In this study, through mutational analysis and QM calculations, we propose a mechanism for the rate enhancement elicited by class-I FH. We also see features distinct to MjFH.

1.4.1. MjFH has two binding sites for substrates

The two subunits of MjFH remain tightly associated even in the absence of ligand or Fe-S cluster. α -subunit of the protein is independently capable of binding the Fe-S cluster; however, it requires the association of the β -subunit for activity. This was confirmed through *in vitro* activity measurements. The kinetic parameters obtained for the enzyme are in concordance with previous reports on class-I FH. However, we observed substrate inhibition in this enzyme, which suggests the presence of an alternative binding site for the substrate. This is the first enzyme of this class to exhibit such a regulation with the only other known regulation reported in class-I FH being cooperativity observed in *Trypanosoma cruzi* (de Pádua et al., 2017). Interestingly, malonate a known inhibitor of class-II FH was bound to two different regions of the tunnel in LmFH, one at the entrance and other at the center which runs across the entire breadth of the protein without connecting to the active site (Feliciano et al., 2016). Two of the residues in the tunnel which bind to malonate in LmFH, Gln195 and Glu267 (LmFH numbering) are conserved while Asp197 is present either as a glutamate or histidine in other class-I FH. Residues Lys213 and Ser217 (part of KGXGS motif) are close to the tunnel with Lys213 at around 10 Å from Gln195 and Glu267. An exhaustive search for side chain rotamers of Lys213 followed by energy minimization in LmFH structure reveals that the lysine residue comes in close contact of 3.8 Å from Glu267 (**Fig. 1.20**). Mutating the lysine and serine residue in MjFH results in large changes in K_i suggesting that the alternative substrate-binding site in MjFH might be located within the tunnel.

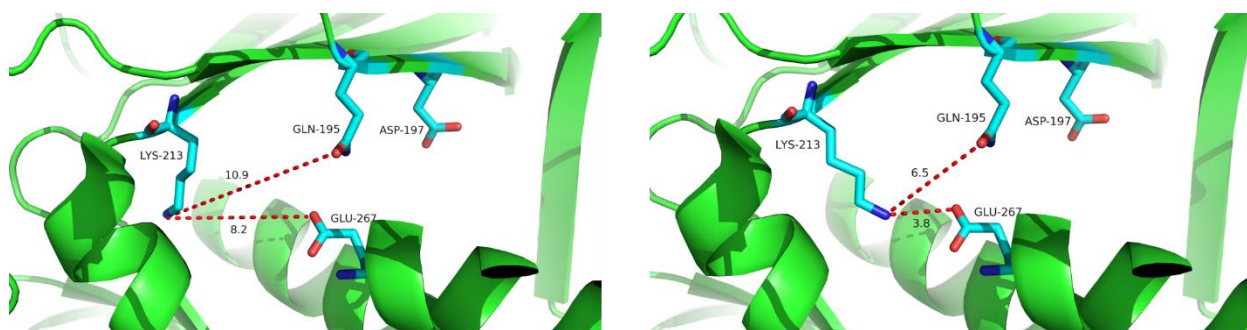


Fig. 1.20. Contacts of Lys213 of the conserved KGXGS motif with Gln195 and Glu267 that interact with malonate bound at the tunnel in LmFH (5l2r). Panel on left shows interatomic distance between Lys²¹³ residue, and Gln¹⁹⁵ and Glu²⁶⁷ present in the tunnel. Search for possible side chain rotamers of Lys²¹³ in

the same structure followed by energy minimization reveals that Lys²¹³ can come close to Gln¹⁹⁵ and Glu²⁶⁷. The tunnel in LmFH structure is independent of the active site. The numbers above the dashed lines are distances in Å.

1.4.2. Binding of Fe-S cluster and substrate compacts FH quaternary structure

Structure of MjFH solved is the first structure of a two-subunit class-I FH. At the time of solving the MjFH apo-protein structure, LmFH structure was not available and SAD phasing was carried out to solve the structure of the $\alpha\beta$ complex. Structure superposition of MjFH over LmFH shows global difference between the two proteins. The two structures are indicative of two different states of the protein, a slightly relaxed MjFH apo-protein state and a more compact LmFH holo-protein state. The primary difference is in the β -subunit and the CTD of the proteins which are spaced differently. We suspected that the subunit/domain possibly moves in or out based on the presence or absence of Fe-S cluster and bound substrate. β -subunit having a “swiveling $\beta/\beta/\alpha$ ” fold, a mobile segment with high B-factor was in support of the assumption. MD simulations carried out with LmFH structure shows that the enzyme expands in 200 ns upon removal of the Fe-S cluster and substrate, going to show that the observed domain motion in apo- and holo-enzymes may be a consequence of Fe-S cluster and substrate binding. Fe-S cluster bound but substrate and inhibitor free structure of LmFH showed increased mobility of the C-terminus domain as compared to the substrate/inhibitor bound structure (Feliciano & Drennan, 2019). Since the pocket for entry of substrate to the active site lies between the α - and β -subunit interfaces, it is possible that the β -subunit mobility may regulate substrate entry/product exit into the active site by opening/closing during the catalytic cycles. Although, in apo-MjFH the cysteine residues that ligate the Fe-S cluster are well poised to accommodate the cluster, active site residues which bind the substrate are present in the loops and some of these residues (Gln⁶¹ and Arg¹⁰² from the α -subunit and Arg⁸⁴ from β -subunit) have a different side chain confirmation (**Fig. 1.10B**). These residues must reorient upon cluster and substrate binding.

1.4.3. Insights into catalytic mechanism of Class-I FH from site-directed mutants of MjFH

Recently the first attempt at understanding the mechanism of catalysis in class-I FH has been carried out in LmFH through mutational studies of active site residues. The Thr467 (LmFH numbering) residue closest to the C3 carbon of bound substrate along with its hydrogen-bonded partners consisting of Arg421 and Arg471 has been predicted to act as catalytic base while Asp135

along with its hydrogen-bonded partner His334 performs the role of catalytic acid (Feliciano & Drennan, 2019). Mutation of the predicted catalytic acid aspartate and catalytic base threonine to alanine respectively, in LmFH led to 3,357- and 3916-fold drop in catalytic efficiency for malate as substrate, and 2538- and 5500-fold drop in catalytic efficiency for fumarate as substrate. These values turned out to be much smaller upon mutating the predicted catalytic residues to valine in MjFH. Reduction in activity of the MjFH catalytic acid and base mutants (D62V and T80V) respectively, corresponded to only 37- and 196-fold for malate as substrate, and 28- and 234-fold for fumarate as substrate. It must be noted that the catalytic rate enhancement offered by the enzyme is of the order of 3×10^{15} and a 5000 fold drop in activity of the enzyme would still make it a phenomenal catalyst. This suggests that the rate determining step in the fumarase reaction in the case of MjFH is not proton abstraction. We must therefore look elsewhere for the molecular origin on the large transition state stabilization first noted by Bearne and Wolfenden (Bearne & Wolfenden, 1995). Mutation of all the residues that form a hydrogen-bonded network with the carboxylate groups of the substrate led to complete loss of activity in MjFH. It must be noted that the LmFH mutants R173A, R421A and R471A corresponding to R102L, R32L and R84L MjFH mutants exhibited extremely weak activity. Our inability to detect this weak activity could be due to the differences in oxygen levels across the two experimental setups: 0.1 ppm versus 6 ppm. As a consensus, the catalytic residues on mutation retained some activity, while mutating the residues forming hydrogen-bonded network with the carboxylate groups of the substrate had little or abolished activity suggesting the importance of binding residues in catalysis.

1.4.4. QM calculations support the dominant role of substrate charge polarization in catalysis

With the current studies on the catalytic mechanism in class-I FH still in its nascent state, our understanding on its catalysis is majorly derived from studies carried out in class-II FH. Though the biochemical equilibrium of the reaction catalyzed by FH is in the direction of fumarate to malate ($K_{eq}=4$), the problem has been viewed in the light of dehydration rather than hydration. A serine residue has been identified as the catalytic base in class-II FH, where it is believed to form an oxyanion hole stabilized by interaction with positively charged neighbors. Previous studies however, have revealed that the rate limiting step does not involve a carbon-hydrogen bond cleavage, clearly suggesting that the reaction may not be driven largely through catalytic base

mediated proton abstraction (Bearne & Wolfenden, 1997). Both class-I and class-II FH catalyze the unidirectional hydration of acetylenedicarboxylate to oxaloacetate with good efficiencies (Flint, 1994; Teipel et al., 1968). This reaction, similar to the reaction catalyzed by acetylene hydratase does not require a catalytic base and the reaction is catalyzed by a water molecule held in position bound to a tungsten cofactor which is rendered nucleophilic or electrophilic by an aspartic acid residue thus allowing it to attack the triple bond (Seiffert et al., 2007). In recent years, quantum chemical studies have proven successful to understand mechanism of enzymatic reactions. DFT calculations on both electrophilic and nucleophilic pathways for water attack on triple bond of acetylene have shown very high energy barriers suggesting an alternative mechanism in acetylene hydratases. Quantum chemical studies on class I-FH to understand the mechanism that this enzyme adopts to catalyze the hydration/dehydration reaction of fumarate was carried out by Soumik Das from Dr. Garima Jindal's laboratory at the Department of Organic Chemistry, Indian Institute of Science, Bangalore.

From DFT calculations it was observed that the carbanion pathway is the most favorable route of interconversion between malate and fumarate in aqueous medium as well as in the enzymatic active site. The uncatalyzed reaction in aqueous medium has a significant energy barrier (41.7 kcal mol⁻¹), which drastically goes down by 25.1 kcal mol⁻¹ in the protein active site environment (**Fig. 1.21**). A significant decrease in free energy barrier is observed from that of aqueous medium in the catalytic acid and the base mutants (M1 and M2), going to show that non-covalently interacting residues around the substrate have an important role in modification of the electronic structure of the substrate.

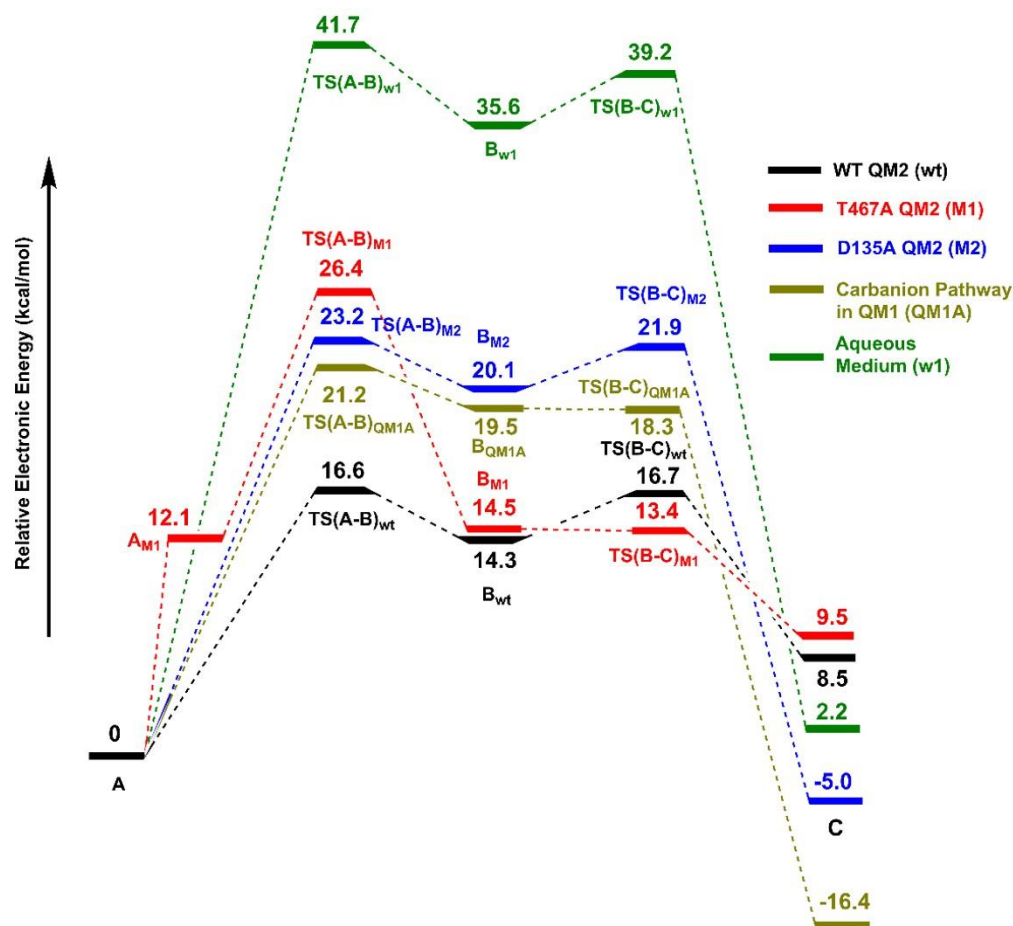


Fig. 1.21. Electronic energy profiles obtained from different model systems

A closer look at the active site of class-I FH reveals that the environment is polarized with more positively charged interaction at one end of the substrate compared to the other (**Fig. 1.15**). We hypothesized that the active site environment polarizes the double bond of fumarate so that the nucleophilic attack of the hydroxyl group becomes favorable. To understand if the double bond of fumarate is really primed for attack of the water molecule, Natural Population Analysis (NPA) calculation on fumarate bound enzyme-substrate complex in the larger model, QM2 for WT and various mutants was carried out by Soumik at Dr. Jindal's laboratory at OC, IISc (**Table 1.5**). From the NPA calculation it was observed that a charge separation does occur with the relative charge on C2 being highly positive in comparison to C3 in WT LmFH. Mutation of residues that bind to the substrate, decreased the polarity on the substrate going to show the importance of the active site in efficiently polarizing the substrate.

1.4.5. Proposed mechanism of catalysis in class-I FH

With the understanding that the base mediated proton abstraction is not the rate limiting step of the reaction and reduction in free energy barriers is not achieved solely by the catalytic acid and base pair, the transition state analogues have been used to explain the series of covalent changes that take place during the fumarase reaction of an enzyme bound substrate (Porter & Bright, 1980). Rate enhancements in enzymes take place through reducing the activation barrier of the reaction by binding strongly to the altered substrate in its transition state compared to the substrate in its ground state (Bearne & Wolfenden, 1995). FH stabilizes this altered substrate in its transition state by a reduction in free energy by at least 30 kcal mol⁻¹ with a release in enthalpy of -24 kcal mol⁻¹ and a gain in entropy of 19 kcal mol⁻¹ (Bearne & Wolfenden, 1995). It is observed that the C3-carbanion intermediate binds FH very strongly, mimicking the transition state (Porter & Bright, 1980). The large favorable enthalpy gain in the fumarase reaction clearly arises from the strong network of electrostatic hydrogen bonds formed by the substrate at the preorganized active site. In the absence of substrate, the charged active site residues are likely to be heavily hydrated. Binding of the dianionic substrate will be accompanied by a release of the hydration shells around both the positive side chain of enzymes and negative charge of carboxylate groups of substrate, contributing to large gain in entropy.

Looking at the active site of class-II FH as well, reveals a pattern similar to class-I FH, where the number of hydrogen-bonds with the substrate is much larger on one end compared to the other (**Fig. 1.22**). We hypothesize that polarization of the substrate at the active site mediated by non-covalent interactions could potentially form the basis of catalysis where binding of the substrate and catalysis become inseparable events as predicted by the transition state theory. Upon substrate binding to the active site, the surrounding residues alter the substrate for nucleophilic attack and provide electrostatic stabilization to the negative charge on the carbanionic intermediate form of the substrate approaching the transition state (**Fig. 1.23**). Bearne and Wolfenden had long predicted that the electrostatic stabilization could explain the large gain in entropy (19 kcal/mol) observed for binding the altered substrate in its transition state relative to the ground state (Bearne & Wolfenden, 1995). We believe that the affinity of FH active site for the binding of the transition state far exceeds the affinity for the substrate, thus catalyzing the reaction.

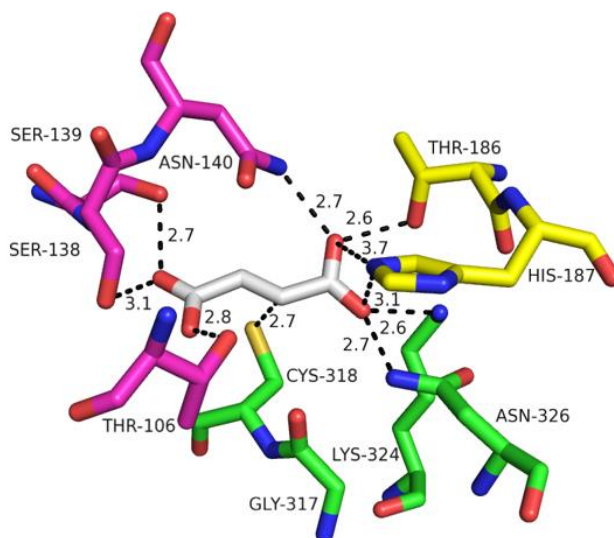


Fig. 1.22. Active site architecture of class-II FH taken from *M. tuberculosis* FH reveals a larger number of H-bonds at one end of the substrate compared to other. Residues from three subunits contribute to catalysis in class-II FH, labelled as magenta, green and yellow. We see that while two serine and a threonine residue are in H-bonding distance from the carboxylate group of bound fumarate (grey) from one end, stronger interactions with lysine, histidine and asparagine are observed at the other end.

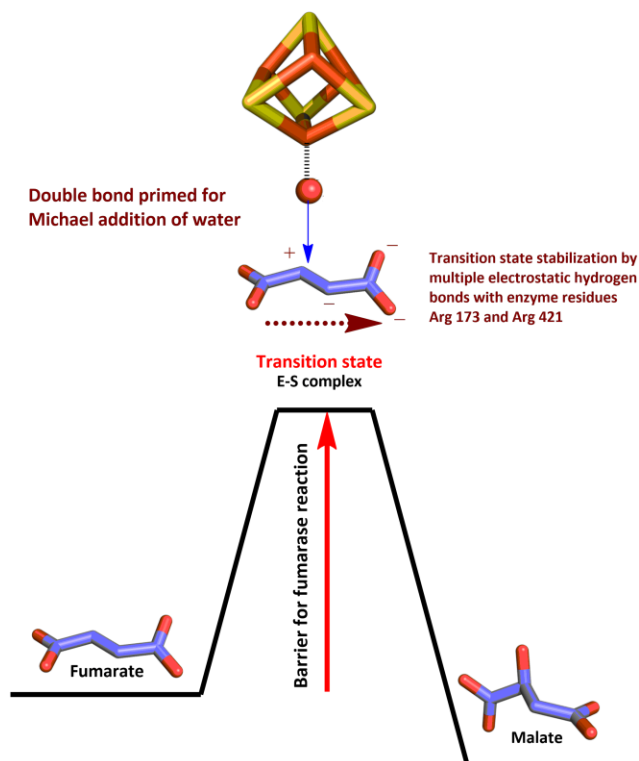


Fig. 1.23. Mechanism of enzyme catalysis in FH. Upon binding to the active site of the enzyme, the substrate is raised to its transition state mediated by the positively charged residues on both sides of the molecule; strongly polarizing it. The transition state complex is well primed for attack of water molecule held in place by the Fe-S cluster following which the product is released.

Fumarate hydratase is an example of a true “Pauling enzyme” (Pauling, 1946). The Fe-S cluster may have been retained as a part of enzyme evolution and only plays a non-redox role in holding the water molecule in place for catalysis similar to that observed in the tungsten containing enzyme, acetylene hydratase (Seiffert et al., 2007). Pre-organized polar environments of active sites have been invoked to drive enzymatic catalysis through electrostatic stabilization of the substrate (Warshel et al., 2006). Recent studies on ketosteroid isomerase, which is an exceptional enzyme that stabilizes the substrate in the transition state by 21 kcal/mol have shown the role of electrostatic field at the active site on enzyme catalysis. Using vibrational stark effect spectroscopy, it was observed that the electrostatic field at the active site contributes significantly to reduction in free energy barrier and subsequently its catalytic rate. Of the 11.5 kcal/mol activation barrier, reduction of 7.3 kcal/mol which corresponds to 70% of the enzyme catalytic rate compared to uncatalyzed reaction was attributed to the electrostatic field alone. Positioning of the catalytic base, an aspartate residue contributed to the rest of the catalytic rate enhancement (Fried & Bagchi, 2014). Probing enzymes for electric field effects as a criterion for rate enhancements can provide useful insights on some of the exceptional enzymes in nature that carry out reactions at enormous rates which otherwise would not have been possible. Although, catalysis in the view of electrostatic transition state stabilization has received good support over the years, theoretical studies backed by reproducible experimental observations are needed to a large extent to address the larger problem of the origin of enzyme catalysis.

1.4.6. Conclusion

The extraordinary proficiency of fumarate hydratases in catalyzing the stereospecific addition of water to the planar, symmetrical, dianionic substrate, fumarate is in large measure due to the constellation of positively charged residues that line the pre-organized active site, polarizing the substrate electron distribution such that in the transition state the olefinic double bond distorts to the aci-carboxylate form. Dramatic electrostatic stabilization of transition states in enzyme catalysis is provided by the example of ketosteroid isomerase, where the activation barrier is reduced by an estimated 21 kcal mol⁻¹ (Fried & Bagchi, 2014). Fumarate hydratases may provide an even more extreme example of transition state stabilization, envisaged decades ago by Polanyi, Haldane and Pauling and elaborated in later years by the seminal work of Jencks (Haldane, 2003; Jencks, 1987; Pauling, 1946; Robinson & Holbrook, 1972). The studies described above also provide firm experimental and theoretical support for the molecular mechanisms which enable

fumarate hydratase to bear the energetic burden of catalysis clearly enunciated in the work of Bearne and Wolfenden (Bearne & Wolfenden, 1995).

1.5. Future directions

Future directions regarding this part of the study will be to solve the structure of MjFH protein with the Fe-S cluster and substrate bound to help understand tertiary fold and local active site rearrangements better. This will also help distinguish the single-subunit and two-subunit type class-I FH. A tunnel that runs through the entire breadth of the protein without connecting the active site has been observed in both MjFH and LmFH structures. The tunnel in LmFH has been found to be occupied by ligands such as glycerol and the dicarboxylic acid molecule – malonate. Malonate is a weak inhibitor of class-I FH and the binding site in the tunnel for malonate have also been found to be conserved. As MjFH displays substrate inhibition, it will be interesting to see if this tunnel has an alternative binding site for the substrate. Mutation of the residues that bind malonate derived from LmFH structure followed by observing levels of substrate inhibition must throw light in this direction. Alternatively, structure of MjFH can be solved in the presence of ligands to observe if the tunnel indeed binds the substrate analogues. Lastly, class-I and class-II FH have been shown to have promiscuous activity on several substrate analogues. One of the substrate analogues is mesaconate, having a methyl group on C2 carbon of fumarate and gets catalyzed by class-I FH by a similar hydration reaction as fumarate to form citramalate. Kinetics was carried out on mesaconate using MjFH and the kinetic parameters were obtained by fitting the data to substrate inhibition equation similar to malate and fumarate (**Fig. 1.24**). Comparison of kinetic parameters with that of fumarate reveals that MjFH has close to equal activity and affinity for mesaconate. Activity on mesaconate has been tested in only three other class-I FH and in all these cases, it is observed that the enzyme retains a high activity on mesaconate, and a good affinity in comparison to fumarate (Jayaraman et al., 2018; Kronen et al., 2015; Kronen & Berg, 2015). Mesaconate is part of the methylaspartate pathway of glutamate fermentation operating in many anaerobic bacteria, all of which have class-I FH. Interestingly class-II FH is inactive on mesaconate. In fact, this is not the only difference between these two classes of enzymes. Thiomalate has been found to be a class-I specific FH inhibitor, not inhibiting the human FH. Since class II FH has been studied in greater depth compared to class-I counterpart, a lot has been understood regarding some of its inactive substrate analogues and inhibitors (details in chapter 2).

Additional substrate screening in class-I FH will give us an idea on differences between the two enzymes. This is especially important as some of the unicellular eukaryotic pathogens which are known to cause disease in humans such as *Plasmodium*, *Cryptosporidium*, *Toxoplasma*, *Trypanosoma*, *Leishmania*, and a few more solely possess class-I enzymes. The fact that humans have only class-II FH makes it even more relevant to design class-I FH specific inhibitors. With the available knowledge of substrates and inhibitors of class-I FH, it will be interesting to design more analogues with thio substitutions on C2 and C3 carbons to obtain more potent inhibitors.

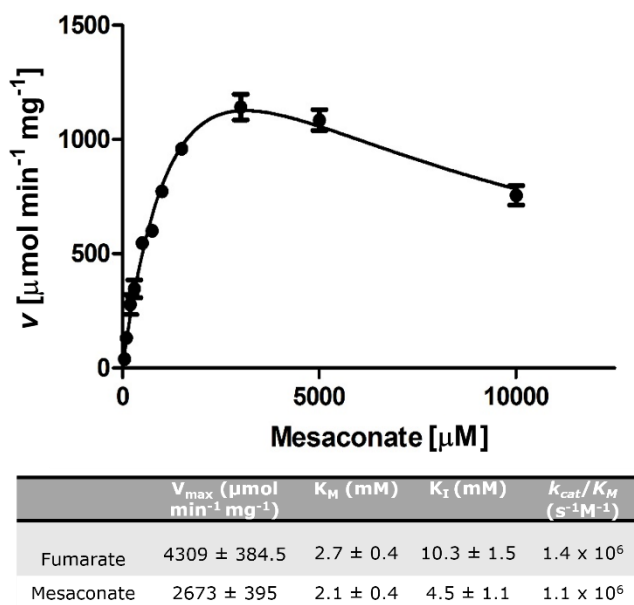


Fig. 1.24. MjFH enzyme activity on mesaconate. The graph represents initial velocity as a function of mesaconate concentration. The data were fit to substrate inhibition equation. Provided in the table below the graph are the kinetic parameters for the enzyme on mesaconate and fumarate. MjFH has comparable activity and affinity for both substrates.

Chapter 2

Understanding stereospsecificity in class-I FH and L-tartrate dehydratase

Class-I fumarate hydratase (FH) has high sequence identity with another enzyme, L-tartrate dehydratase (L-TTD); so much so that the two have been misannotated in some of the sequence databases. The two enzyme sequences are hard to distinguish apart and share identical active site residues. Despite possessing active site identity, they catalyze reactions on opposite stereoisomers. This is the first example of two enzymes with active site possessing identical residues, having specificity for opposite stereoisomers. Efforts made at understanding the substrate specificity in these two enzymes are discussed in this chapter. A phylogenetic tree has been constructed and the conserved motifs ‘KGXGS’ in class-I FH and ‘AGGGC’ in L-TTD have been identified to be a distinguishing feature between the two enzymes. KGXGS/AGGGC motif swap has been attempted to observe any switch in enzyme specificity. Mutations that may reduce the free energy of binding for the opposite isomer, L-tartrate to class-I FH have been identified and the results pertaining to these mutants are discussed. Screening for activity on various substrate analogues has revealed that L-TTD and MjFH have specificity for substrates of opposite chirality, disclosing an enigmatic nature of substrate specificity in this class of enzymes. A close observation of the active site residues of class-I FH offers leads on how the two enzymes could potentially catalyze reactions on opposite stereoisomers. As to how this observation can be established/validated is provided in the future directions of the study towards the end of the chapter.

2.1 Introduction

2.1.1. Types of substrate specificity offered by enzymes

Specificity of an enzyme is defined as the ability of its binding site to bind specific ligands. Fewer the number of ligands, greater its specificity. Electrostatic and hydrophobic interactions constitute a strong force in determining enzyme-ligand specificity (Srinivasan, 2021b). Apart from the specificity offered by the enzyme in binding its substrate, precise orientation as well as affinity for transition state enables catalysis (Srinivasan, 2021a). In general, a key feature of enzymes is their ability to catalyze highly specific reactions. However, a large number of enzymes are exceptions to this universally accepted enzyme specificity. More often than not, enzymes are capable of catalyzing reactions on molecules other than their physiological substrates. The physiologically irrelevant promiscuous activity is hypothesized to be the point of origin for evolution of new enzymes (Khersonsky & Tawfik, 2010). Cross reactivity of enzymes usually occurs on similar substrates through an identical reaction mechanism. Substrate specificity offered by enzymes are broadly divided into four groups: absolute specificity, group specificity, linkage specificity and stereospecificity.

Absolutely specific enzymes exclusively catalyze reaction only on a single substrate. Glucokinase, which catalyzes phosphorylation of glucose to glucose-6-phosphate is a classic example of the same. Group specificity occurs when an enzyme is capable of catalyzing reaction on substrates with specific functional group. Hexokinase unlike glucokinase has group specificity and can catalyze phosphorylation of multiple hexoses. Linkage specificity or bond specificity of enzymes help enzymes catalyze reactions on a particular bond type. Peptidase is an example of an enzyme having linkage specificity as they catalyze reactions on peptide bonds. Substrates having a chiral center can exist as optical isomers and stereospecific enzymes catalyze reactions on only specific optical isomers called stereoisomers. As an example, amylase a glucosidase can specifically break alpha-glycosidic linkages in glycogen and starch, but not beta-glycosidic linkages observed in cellulose.

2.1.2. How do enzymes deal with stereoisomers?

First discovered by Louis Pasteur, chirality is an inherent property of organic molecules (Pasteur, 1864) that possess a tetrahedral sp^3 hybridized carbon with four different substituents. A molecule

with n chiral centers will have 2^n diastereoisomers. Every major category of biological molecules including nucleic acids, proteins, lipids, polysaccharides, and steroids are of only one chirality. Enzymes, therefore, are exclusively selective catalysts that catalyze reactions on only one stereoisomer and to catalyze the reaction on the opposite enantiomer there is altogether a different enzyme. In case of NAD-dependent dehydrogenases, two enzymes with different folds but having a mirror inversion of the catalytic residues catalyze reaction on the L- and the D-isomers (Lamzin et al., 1994). In case of glycosyl hydrolases, nature has evolved different families of enzymes with totally different 3D folds with an inversion of catalytic base with respect to the substrate, leading to recognition and catalysis on the anomeric carbon of opposite chiralities (McCarter & Stephen Withers, 1994). Perhaps lipases are the only known exceptions where the enzyme can catalyze reactions on both enantiomers though having a clear preference for one over the other. Chirality of the active site in lipases gives the enzyme preference for one enantiomer without discriminating the other (Lamzin et al., 1995). There are a varied set of ways in which enzymes have evolved to work around stereoisomers. The most common being evolution of different enzymes to carry out reactions on opposite stereoisomers with both enzymes having different active site architectures. Stereospecificity of an enzyme arises from the three-dimensional structure of the active site that snugly fits in the substrate in alignment with the catalytic residues (Hedstrom, 2010).

2.1.3. Significance of enzyme stereospecificity

Different enantiomers of a biologically active molecule generally have different physiological properties, in fact with some having toxic side effects. More than half of the small molecule drugs clinically used today are chiral (Z. Shen et al., 2016). Compared to the traditional organic synthesis, enzyme catalyzed reactions yield selective products with a specific configuration (Wombacher et al., 2006). Importance of asymmetric catalysis in pharmacology is perhaps best exemplified by the use of thalidomide for insomnia and morning sickness in the 1960s. While the *R*-enantiomer was therapeutically active, *S*-enantiomer was teratogenic, causing delivery of newborn babies with shortened flipper-like limbs (Vargesson, 2015). Despite remarkable progress in the last couple of decades in the field of asymmetric synthesis, complete control of stereogenic centers in catalyst-controlled reactions have remained a challenge (Xu et al., 2019). The broad application of enzymes in obtaining stereospecific products has driven rapid development of chiral compounds and opportunities to obtain therapeutically active drugs (Mu et al., 2020).

2.1.4. Methodologies adapted to switch stereospecificity

Biocatalysts are receiving more attention in chemistry for a myriad of applications relating to the catalysts' ability to enable reactions to proceed at high efficiency with greater yield while following an eco-friendly process. Because of the importance of stereoselectivity in pharmacology, structural and mechanistic basis underlying the phenomena mediated by the enzyme has been a subject of intense study (Chan et al., 2019). In this interest, various tools and methods have been used to manipulate stereoselective properties of enzymes such as, de novo enzyme engineering, computational rational designing, and in vitro evolution.

In an earlier publication, attempts at de novo enzyme design to catalyze a Diels-Alder reaction, which forms two carbon-carbon bonds and four stereocenters has been reported (Siegel et al., 2010). The strategy was aimed at stabilizing the transition state of the substrates through hydrogen bonding. Computational strategies adapted to sample side chain rotamers, catalytic sites, enzyme-substrate hydrogen-bond symmetry, atomic clashes, and transition state binding energies yielded 84 predictions of enzyme constructs. However, a low hit of only 2 enzyme designs obtained experimentally showed Diels-alderase activity, indicating challenges associated with pure de novo design to sample large conformational flexibilities between enzyme and substrate. Rational enzyme engineering can also be carried out through modern computational methodologies and algorithms that can aid modulation of enzyme catalytic properties including stereospecificity. Molecular docking and MD simulation approach provide insights into the atomic interactions between the enzyme and the substrate (Chan et al., 2018). Moreover, allosteric effects of mutations at a site distant from the active site of the enzyme can also be predicted through MD simulation (Reetz, 2011). Many in silico tools have also been developed that carry out relative binding energy evaluations, such as Poisson-Boltzmann surface area continuum solvation (Genheden & Ryde, 2015), free energy perturbation (FEP) (Jorgensen & Thomas, 2008) and thermodynamic integration (TI) (Ruiter & Oostenbrink, 2016). Most rational designing revolves around known enzyme active sites, where mutations of key polar residues into hydrophobic ones modulate the active site pocket size and restrain substrate orientation to achieve stereospecificity. Experimentally, directed evolution is the most promising strategy to improve stereoselectivity. The methodology involves recursive cycles of mutagenesis followed by protein expression and

screening for the desired property (Li & Reetz, 2016). Through this technique, enzymes have been evolved to catalyze reactions that no known enzymes in nature can (Hammer et al., 2017).

2.1.5. Class-I FH and L-TTD, a unique case of opposite stereospecificity

Class-I FH shares a high similarity with L-tartrate dehydratase (L-TTD) (45%) which belongs to the hydrolyase family of enzymes. L-TTD catalyzes the stereospecific conversion of L-tartrate to oxaloacetate with its activity being observed only in prokaryotes and remains to be identified in archaea and eukaryotes. L-TTD has limited studies with regard to structure and catalysis with no structure of the enzyme solved to date and biochemical characterization carried out in only three organisms; *Pseudomonas putida* (Kelly & Scopes, 1986), *E. coli* (Reaney, Begg, et al., 2009) and an L-tartrate fermenting anaerobic bacteria (Schink, 1984). The structure of L-TTD has been predicted from primary sequence using constraints derived from sequence co-evolution and modelling (Ovchinnikov et al., 2015). This structure does not have a bound 4Fe-4S cluster, and the three cysteine residues known to coordinate the cluster are not very well poised to accommodate it, making it difficult to consider the structure with confidence. L-TTD and class-I FH share such high similarity that the two have been misannotated in the RCSB and UniProt database. The RCSB database in which the X-ray structure of the β -subunit of MjFH has been deposited shows the component of the protein to be a putative L-tartrate dehydratase β -subunit and in the UniProt database, under the protein tartrate dehydratase β -subunit, structure of MjFH β -subunit shows up as an entry. Class-I FH has been shown to catalyze reaction on D-tartrate to form oxaloacetate going to show that the two enzymes have opposite stereospecificities (Kronen & Berg, 2015; van Vugt-Lussenburg et al., 2009, 2013). A multiple sequence alignment of class-I FH and L-TTD revealed that the two enzymes share an active site with identical residues, with all the residues involved in binding the substrate in case of FH being conserved in both these proteins (**Fig. 1.1**). Surprisingly, two enzymes with active site possessing identical residues catalyzing reactions on opposite stereoisomers have not been identified, making class-I FH and L-TTD a notably unique set of stereospecific enzymes.

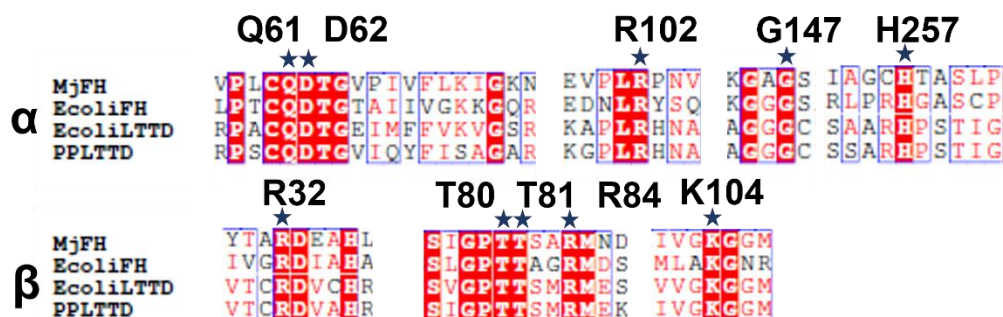


Fig. 2.1. Multiple sequence alignment of class-I FH and L-TTD highlighting the active site residues involved in substrate binding and catalysis. Active site residues deciphered from the structure of MjFH (chapter 1) have been highlighted over the alignment with a blue star mark. Representative sequences from *Methanocaldococcus jannaschii* and *Escherichia coli* for class-I FH and *Pseudomonas putida* and *Escherichia coli* for L-TTD were considered for the alignment.

2.1.6. Objectives

L-tartrate dehydratase has not been kinetically characterized to date and the aim of this study was to i) characterize the enzyme kinetically and structurally from *E. coli*, ii) understand if a motif distinguishes the two enzymes to help solve the issue of misannotation, iii) carry out a wider substrate screen to understand similarities and differences between the two enzymes and finally iv) understand specificity offering residues in the two enzymes in order to switch stereospecificity.

2.2. Experimental procedures

2.2.1. Chemicals, strains, and molecular biology reagents

Restriction enzymes, Phusion polymerase, and T4 DNA ligase were obtained from New England Biolabs, USA. Ni-NTA conjugated agarose beads were obtained from Thermo Fisher Scientific Inc. Malic dehydrogenase from the porcine heart was commercially procured from Sigma-Aldrich (product number M2634), USA. Primers were custom synthesized from Sigma Aldrich, Bengaluru. Media components for growing *E. coli* cultures were from Himedia, Mumbai, India.

2.2.2. Sequence alignment and phylogenetic tree construction

Sequence of class-I FH and L-TTD were obtained from NCBI BLAST-P tool using *Methanocaldococcus jannaschii* FH (MjFH), *E. coli* FH (*fumA*) and *E. coli* L-TTD protein sequences as queries. Clustal Omega (Sievers et al., 2011) was used for multiple sequence alignment and output was processed using ESPRIPT (Robert & Gouet, 2014) for visualization. For the construction of phylogenetic tree, Excel was used to concatenate α and β protein sequences of two-subunit class-I FH and L-TTD. The duplicate sequences were removed and clustered using CD-HIT online suite at a cut-off of 85% sequence identity to remove redundancies. The sequences were then aligned in MEGA X (Kumar et al., 2016) using MUSCLE algorithm. The evolutionary history was inferred by using the Maximum Likelihood method and Le Gascuel model. The tree with the highest log likelihood (-88755.07) is shown. 500 bootstrap replicates (Felsenstein, 1985) were performed and branches corresponding to partitions reproduced in less than 60% bootstrap replicates are collapsed. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.0890)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 2.92% sites). This analysis involved 258 amino acid sequences. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous amino acids were allowed at any position (partial deletion option). There was a total of 428 positions in the final dataset.

2.2.3. Cloning, protein expression and purification

Genes for the two subunits, L-TTD α (orfz1) and L-TTD β (orfz2) were PCR amplified from the genomic DNA of the *E. coli* strain DH5 α strain using the primers listed in **Table. 2.1** and cloned into two tandem multiple cloning sites (MCS) of the pET-Duet vector using the enzymes *EcoRI* and *HindIII* for L-TTD α such that the (His)₆ tag is at its N-terminus and enzymes *NdeI* and *XhoI* for L-TTD β . Cloning was confirmed by insert release and sequencing. AGGGC motif along with adjacent residues (YMAGGGCTLP) from the above clone was replaced with KGXGS motif (FPKGAGSENM) from MjFH using the primers listed in **Table. 2.1** and confirmed through sequencing.

E. coli FH (*fumA*) was also PCR amplified from genomic DNA of *E. coli* strain DH5 α using primers listed in **Table. 2.1** and cloned into the MCS of pET-Duet vector using the enzymes *BamHI* and *Sall* such that the (His)₆ tag is at its N-terminus. Parent plasmid containing *fumA* was used as a template to sub-clone the N-terminal domain (NTD) and C-terminal domain (CTD) of the proteins separately into another pET-Duet vector using the primers listed in **Table. 2.1**. NTD of *fumA* was cloned between *BamHI* and *Sall* and CTD was cloned between *NdeI* and *KpnI*. All the above clones were confirmed through insert release and sequencing.

Table. 2.1. Oligonucleotide sequences used for cloning *E. coli* WT L-TTD α and β genes and *fumA* in pET-Duet vector and generating site-directed mutants

Primer name	Primer sequence (5' to 3')
TTDA_US	CCGGAATTCGATGAGCGAAAAGTAATAAGCAACAGGCACTG
TTDA_DS	CCCAAGCTTTCATAACGCGCTCCGGGTGTGAGACAG
TTDB_US	CTAATTCATATGAAAAAGATCCTGACAACCCCGATCAAAGCTG
TTDB_DS	CCGCTCGAGTTATTTGATGTAATGGACGTGCTCGCAGATCTC
FumA_FP	GATGTAGGATCCATCAAACAAACCCTTTCATTATCAGGCTC
FumA_RP	GCAATAGTCGACTTATTTACACAGCGGGTGCATTGTG
NTD_FP	GATGTAGGATCCATCAAACAAACCCTTTCATTATCAGGC
NTD_RP	GCAATAGTCGACTTAACGGTTGATCTTCGCTTTGATATTACGGTC
CTD_FP	GAACCTCATATGCGCGTTGACCTTAACCGTCCGATGAAAGAG
CTD_RP	GATGTAGGTACCTTATTTACACAGCGGGTGCATTGTGTGAG
TTD_KGXGS_FP	TTTCCAAAAGGGGCAGGAAGCGAAAACATGGGCCGCTCGAAAAGTGTT AATGCCGTCAG
TTD_KGXGS_RP	CATGTTTTCGCTTCCTGCCCTTTTGAAAAACTTCGATTTCCGCATCG TCATTGTGCG
MjFH_R84K_FP	GCCCAACAACATCTGCAAAAATGAATGATGTTGAAG
MjFH_R84K_RP	CTTCAACATCATTCAATTTTGCAGATGTTGTTGGGC
MjFH_R32K_FP	GGCAAAATATACTGCGAAAGATGAAGCACATTTAAAAATTATTG
MjFH_R32K_RP	CAATAATTTTAAATGTGCTTCATCTTTCGCAGTGTATATTTGCC

MjFH_G145K_A1 46R_S148P_FP	GATAATTGCATTTCCAAAAAACGTGGACCGGAAAACATGAGTGCTT TAAAG
MjFH_G145K_A1 46R_S148P_RP	CTTTAAAGCACTCATGTTTTCCGGTCCACGTTTTTTGGAAATGCAAT TATC
MjFH_P79R_FP	GGGTTTGTGTTTCTATAGGCCGTACAACATCTGCAAAAATG
MjFH_P79R_RP	CATTTTTGCAGATGTTGTACGGCCTATAGAAACACAAACCC
MjFH_R102K_FP	CGGAAGAGGTTTCCTTTAAAACCTAATGTAGTTCATCC
MjFH_R102K_RP	GGATGAACTACATTAGGTTTTAAAGGAACCTCTTCCG

In order to express the L-TTD protein, pET-Duet containing both the genes for *E. coli* L-TTD α and β subunits, with N-terminus of α subunit hexa-histidine tagged (pETduet_LTTD $\alpha\beta$) was used for transformation and over-expression in BL21(DE3)–RIL *E. coli* cells containing an additional plasmid (RIL) expressing tRNAs for the rare codons of arginine, isoleucine, and leucine. To over-express L-TTD, BL21 (DE3) RIL *E. coli* cells transformed with the recombinant plasmid – pETduet_LTTD $\alpha\beta$ were grown at 37 °C to an O.D of 0.45, induced with 0.3 mM IPTG and grown for further 8 hours at 37 °C. The cells were pelleted and lysed with lysis buffer (50mM Tris HCl, pH 7.4, 5% glycerol, 1mM DTT and 1mM PMSF) using French pressure cell press (Thermo IEC Inc., USA). The lysate was clarified by centrifuging at 30,000 x g for 15 minutes. Supernatant was carefully collected without disturbing the pellet and a 1 mL slurry of Ni-NTA beads (Novagen) pre-equilibrated with lysis buffer was added for the binding of His-tagged protein. The binding was done at 4 °C for 3 hours and thereafter, the beads were washed with 50 mL of lysis buffer to remove unbound proteins and Ni-NTA bound protein was eluted with different concentrations of imidazole. Fractions containing pure protein were pooled, dialysed in dialysis buffer containing 50mM Tris HCl, pH 7.4, 5% glycerol and 1mM DTT and stored at -80 °C. The purified protein was analyzed by SDS-PAGE and protein concentration was determined by the Bradford method with bovine serum albumin (BSA) as the standard.

Expression of *E. coli fumA*, NTD and CTD proteins were checked by transformation of the respective constructs into BL21 (DE3) RIL *E. coli* competent cells. The transformed culture was grown at 37 °C to an O.D of 0.45, induced with 0.5 mM IPTG and grown further for 6 hours at 37 °C. The cells were pelleted, and an aliquot of the lysate was run on SDS-PAGE to check for protein expression. Protein purification through Ni-NTA affinity chromatography was carried out for *fumA* similar to the purification protocol used for L-TTD.

2.2.4. Reconstitution of Fe-S cluster

The protein obtained after purification and dialysis is the apoprotein lacking the iron-sulfur cluster. Protocol similar to that followed for reconstitution of MjFH in Chapter 1 was followed for reconstitution of Fe-S cluster into L-TTD. Reconstitution was carried out in an anaerobic chamber to prevent oxidation of the cluster. Purified protein was placed inside the chamber and 50x Molar excess of DTT was added and stirred for 30 minutes followed by addition of 10x Molar excess of each $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ and Na_2S . The protein solution was stirred for 3 hours to obtain the holoprotein with fully reconstituted iron-sulfur cluster.

2.2.5. Enzyme activity

All reactions for enzyme activity measurement unless specified otherwise, were conducted at 37 °C using a water-circulated cell holder fitted to Hitachi U2010 spectrophotometer. Fumarate to malate conversion was monitored as a decrease in absorbance at 240 nm (molar extinction coefficient (ϵ) of fumarate is $2.4 \text{ mM}^{-1}\text{cm}^{-1}$ at 240 nm) and as an increase in absorbance for malate to fumarate conversion. The assay mixture contained 500 μM fumarate or 1 mM malate in 50 mM Tris-HCl, pH 7.4. Reaction was initiated by the addition of enzyme and monitored for around one minute to obtain reliable slopes. Activity on tartrate was monitored using a coupled enzyme, porcine malate dehydrogenase, which was procured commercially. The reaction was carried out in a final volume of 500 μL in 50 mM Tris-HCl, pH 7.4, containing 250 μM NADH, 2.5 μg MDH with varying concentrations of L-tartrate. Conversion of L-tartrate to oxaloacetate was coupled to NAD^+ formation by the coupling enzyme which converts oxaloacetate to L-malate utilizing a molecule of NADH. Reaction was initiated by the addition of 1.44 μg *E. coli* L-TTD and monitored at 340 nm (molar extinction coefficient(ϵ) of NADH is $6.22 \text{ mM}^{-1}\text{cm}^{-1}$ at 340 nm) for around one minute to obtain reliable slopes.

To confirm the formation of malate from fumarate catalyzed by L-TTD, LC-ESI MS was carried out. LC-MS analysis was performed using HILIC chromatography (ZIC-pHILIC column; Sequant, 150 mm x 2.1 mm, 5 μm particle size with guard column 20 mm x 2.1 mm, 5 μm particle size) using an Ultimate 3000 UHPLC system (Dionex) coupled to Q Exactive Orbitrap HF mass spectrometer (ThermoFisher) operating in the negative ion mode with HESI ion source. The mobile phase used was, solvent A - 10mM ammonium carbonate, pH 8.4 and solvent B was

acetonitrile. Gradient used was, 80% B to 40% B in 20 min, 40% to 20% in 15 minutes, 20% B to 80% B in 1 min, re-equilibration with 80% B for 10 min at 0.1ml/min flow rate. The column temperature was 45 °C, injection volume of 2 µL of the sample. The concentration of malate and fumarate when used as standards was 5 µM. A reaction mix consisting of 5 µM fumarate was incubated with 9.8 µM enzyme overnight in 5 mM Tris-HCl, pH 7.4. Thereafter, the sample was passed through a spin column, having a membrane with a 10 kDa molecular mass cutoff to remove the enzyme from the assay mixture and 2 µL of the sample was subjected to LC-MS analysis. Standards were also incubated overnight in 5 mM Tris-HCl, pH 7.4, similar to the test sample. The spectrometer was operated in negative polarity with the following settings: resolution 30000, AGC 3×10^6 , m/z range 75-500, sheath gas 30, auxiliary gas 15, probe temperature 100 °C S-lens RF level 50, and capillary temperature 320 °C, spray voltage 3.5 kV (negative mode), maximum ion injection time 200 ms, microscans 3.

2.2.6. Kinetics of inhibition by substrate analogs

Inhibition kinetics of MjFH and L-TTD by substrate analogues meso-tartrate, L-tartrate and L-malate was examined by fixing the substrate concentration and varying the concentrations of the inhibitors. The data was fit to the equation for competitive inhibition $v = V_{\max}[S]/(K_m(1+[I]/K_i)+[S])$ and the K_i was estimated using nonlinear regression analysis. Graphpad Prism software was used for data fitting and analysis.

2.2.7. Generation of L-TTD null *E. coli* strains and validation

L-TTD null strains were generated in BW25113 strain of *E. coli* obtained from Coli Genetic Stock Centre (CGSC) (Baba et al., 2006) and in Δ *fumACB* fumarase knockout strain of *E. coli* that was previously generated at our lab (Jayaraman et al., 2018), using standard protocols (Datsenko & Wanner, 2000). The *orfz1* and *orfz2* tandem genes coding for the L-TTD α and β subunits were knocked out together. In order to knockout the genes, primers TTDABKO_US and TTDKOAB_DS (**Table. 2.2**) were used to amplify kanamycin cassette using plasmid pKD13 as the template such that the PCR product carries the homologous regions corresponding to the 5' flank of *orfz1* and 3' flank of *orfz2* genes. The PCR product was then subjected to *DpnI* digestion and transformed to BW25113/ Δ *fumACB* strains through electroporation and selected on LB medium containing kanamycin. Kanamycin resistant colonies were screened by PCR using

TTDKO_P3 and TTDKO_P6 primers (**Table. 2.2**) to validate the insertion of antibiotic cassette in the right locus. The kanamycin cassette at the *orfz1/orfz2* gene locus was removed by expressing flippase in the strain resulting in generation of marker free $\Delta L\text{-TTD}/\Delta\text{fumACB}_{\Delta L\text{-TTD}}$ strain. The above knockout strains were grown in LB medium overnight, and an aliquot of both was washed three times with autoclaved M9 medium to remove traces of LB medium. The washed cells were resuspended in M9 medium, and an aliquot of the suspension was streaked on minimal medium plate containing the appropriate carbon source (20 mM of malate, fumarate, oxaloacetate, or L-tartrate) and antibiotics. The plate was incubated at 37 °C for 110 hours and growth phenotype was scored.

Table. 2.2. Oligonucleotide sequences used for knockout of L-TTD *orfz1* and *orfz2* genes in WT BW25113 *E. coli* and ΔfumACB strain

Primer name	Primer sequence (5' to 3')
TTDABKO_US	TAGTTTGTGGCGGTGAATTTTTCCAGACAAATACAAAACTGGAGTTGCC GTGTAGGCTGGAGCTGCTC
TTDKOAB_DS	CATCGTGTTATGCCTGCGAAAAAGGAGAGGGGTGAGCCTCTCCGGGAGG GCTGTCAAACATGAGAATTAATCC
TTDKO_P3	CGCTAAAACCTCCCGCTGTGAACGATTTCTACCAG
TTDKO_P6	GCAGTGAATTTAAAGCCTGGCTGAGCGAGC

2.2.8 Computational design for switching stereospecificity in class I FH to catalyze L-tartrate dehydratase activity

Iterative Protein Redesign and Optimization (IPRO) requires a model of the protein complex structure and predetermined residues in the structure that are mutable called the design position and a set of constraints that the protein variants have to reach relative to wild-type (Grisewood et al., 2017). The constraints include a stronger interaction energy with the new substrate and a weaker interaction with the native substrate. After selecting the design position which involves residues from and around the active site of *Leishmania major* class I FH structure (512r), IPRO algorithm was used to search for mutations that led to improved binding of FH with L-tartrate simultaneously discouraging L-malate binding. Fitness of each predicted protein variant was assessed by its interaction energy as an alternative for binding energy (i.e., ΔG). Each IPRO iteration begins with the first step where local backbone perturbation nearby the selected design position takes place followed by identification of optimal set of amino acid for new backbone conformation using a mixed-integer linear program (MILP). In the third step, a local rigid-body

docking takes place to reorient ligand within the binding site followed by energy minimization of the entire enzyme complex in the fourth step. The fifth and final step evaluates the set constraints, and the results of the iteration are retained or discarded based on Metropolis criterion. Multiple IPRO trajectories were simulated to improve altered substrate specificity for L-tartrate and in each case the interaction energies were calculated. Variations were prioritized based on extent of energy differences and five variants were selected to build a focused library for experimental testing. *All the computational work utilizing IPRO to suggest energy guided mutants that could potentially bind L-tartrate was performed by Ratul Chowdhury at Prof. Costas D. Maranas laboratory, Chemical and Biological Systems Optimization Lab, Penn. State University.*

2.3. Results and discussion

2.3.1 Multiple sequence alignment

Alignment of class-I FH and L-TTD sequences from organisms in which they have been biochemically characterized was carried out after fusing the α - and β -subunits of two-subunit FH and L-TTD. The multiple sequence alignment of the single-subunit FH, two-subunit FH and L-TTD reveals good sequence conservation and a domain boundary in single-subunit type FH (**Fig. 2.2**). Much of the differences seen between these proteins arises from the N-terminus of the N-terminal domain which has an extension in the single-subunit FH that is not seen in two-subunit type protein. The C-terminus stretch of the C-terminal domain also lacks good conservation. The alignment reveals that the motif ‘KGXGS’ of class-I FH located between residues 144 and 148 in MjFH is very well conserved and is present as an ‘AGGGC’ motif in L-TTD (**Fig. 2.2**). From *Leishmania major* FH (LmFH) structure, 9 residues in the active site of the protein are known to form hydrogen bonds with the substrate L-malate. Corresponding residues in MjFH – Gln⁶¹, Asp⁶², Arg¹⁰² and Gly¹⁴⁷ from α -subunit and Arg³², Thr⁸⁰, Thr⁸¹, Arg⁸⁴ and Lys¹⁰⁴ (Arg³¹⁷, Thr³⁶⁵, Thr³⁶⁶, Arg³⁶⁹ and Lys³⁸⁹ in the fused sequence) from the β -subunit are well conserved across all class-I FH and L-TTD enzymes. Overall class-I FH shares 45% similarity with L-TTD making it hard to discern the two with only sequence information.

```

Methanocaldococcus_janschii_FH      1  .....MSNKPFHYQAPFPPLKK...DDTEYYLLTS...EHVSVSEFEGQEILKV
E_coli_FumA                          1  MSLCDQCEIGCRRRVGKIDEDASAVNADPFHSAIFQPTDPHHHQTFAKVEGSEKYVEEVEVFGRAQLKV
Leishmania_major_FH                  1  .....MSNKPFHYQSPFPPLKK...NDTEYYLLTS...EHVSVSEFEGQEILKV
Shigella_sonnei_FumA                 1  .....MAEFVYQDPFPLGK...DETKYRLVTK...DHVSVGNFDGKEILKI
Pseudomonas_putida_FumA              1  .....MTVI
Mpcb                                  1  .....MSDNKRPVSTEL...AMTVI
Burkholderia_xenovorans_FH           1  .....MSRI
Pyrococcus_furiosus_FH               1  .....MRE
Salmonella_typhimurium_FH            1  .....M
Helicobacter_hepaticus_FH            1  .....M
Archaeoglobus_fulgidus_FH            1  .....M
Pelotomaculum_thermopropionicum_FH  1  .....M...TIRTI
E_coli_TTD                            1  .....MMSE...SNKQQ
Shigella_sonnei_TTD                  1  .....MMSE...SNKQQ
Pseudomonas_putida_TTD                1  .....MET...VDKEA

Methanocaldococcus_janschii_FH      1  ..MKISDVVVEIFREAIYLPEDVKNAL.EAYKKESESISKNTLKAITENNKIAEETQVPIAGDTGVPIT
E_coli_FumA                          43  APEALTLARQAFHDAFSLRPAHQQQVADILRDFEASENDRYVALQFLRNSDIAAKGVLPICGDDTGTAI
Leishmania_major_FH                  71  NPEALTLARHAFSDVHHFRKDHLEGWRRATEDEASDNDRYVATLLKNACTAAGRVLPICGDDTGTAI
Shigella_sonnei_FumA                 43  APEALTLARQAFHDAFSLRPAHQQQVADILRDFEASENDRYVALQFLRNSDIAAKGVLPICGDDTGTAI
Pseudomonas_putida_FumA              5   KQDDLIQSVADALQFISYYHPVFIQAMH.EAYLREESPAARDSTIAQLINSRMCATGHRPICGDDTGI
Mpcb                                  41  DPEALALLANAMRDVVSFMLRPAHLEKVAAILSDPESSANDRGVAVALLRNAEVAAGKGVLPICGDDTGT
Burkholderia_xenovorans_FH           18  KQEDLIQSIADSLQYISYHPLOYIQALG.RAYELQSPAAKDATAQILNTRMCAEGKRPICGDDTGI
Pyrococcus_furiosus_FH                6   .MIKVEDIVEALKLAVINLPEQVALALK.NAYEREDVEIARYNLEIMLRVAKTSKKEKVPICGDDTGT
Salmonella_typhimurium_FH             1  NTEIITIPITEKLVKACYELDDNLMCSFR.KAYDKEESKIGKETIKLIDNNGEYAKKEQLACGDDTGT
Helicobacter_hepaticus_FH            5  RYEDIKSAVAKLAINACCIOPTPIKRAFS.AAKGSEKALGQNILOTLIENGGKIAESNMMPICGDDTGM
Archaeoglobus_fulgidus_FH            2  EFDVAVNAVVDALKRAHTELPDQVLEAIK.RAYGEENEIARKNLEAILKNIEAAKKGVPICGDDTGI
Pelotomaculum_thermopropionicum_FH  7  NTSVITENVARLCQEAANFLGHDVMEAFK.KAQEEISILTGKGILODLIVNAEIAACEEQAPICGDDTGY
E_coli_TTD                            10  AVNKLTEIVANFTAMISTRMPDQVVDKLG.LKDAETSSMGKIIYHTMFDNMMQKAIDLNRPICGDDTGEI
Shigella_sonnei_TTD                  10  AVNKLTEIVANFTAMISTRMPDQVVDKLG.LKDAETSSMGKIIYHTMFDNMMQKAIDLNRPICGDDTGEI
Pseudomonas_putida_TTD                9  AVESLTDVLAKEFTAYIGKRLPKQVKEKTA.KLRAAETNPLAVYDSMANQDYADKLNRPICGDDTGI

```

Methanocaldococcus_janschii_FH	68	VFKIKGNINSSEI.MKIIIEEIKGVKKATEE.VP	LRNVVHPL...TRENFKNVGLNSPFIINIEFDE
E_coli_FumA	113	IVGKKGQRV.WTGGG..DEAALARGVYNTYTE..DN	LRVSNQAPLDMYKEVNTGTNLPFA...QIDLAVD
Leishmania_major_FH	141	VLGKRGELC.WTGGG..DEKYLKGIWNAYRY..HN	LRVSNQAPLDMYKEVNTGTNLPFA...QIDLAVD
Shigella_sonnei_FumA	113	IVGKKGQRV.WTGGG..DEAALARGVYNTYTE..DN	LRVSNQAPLDMYKEVNTGTNLPFA...QIDLAVD
Pseudomonas_putida_FumA	74	VFKIRVGMQDVRWDGATMSVDDMNEGVRRAAYLPEVNV	LRASILADP.AGARKNTKDNTPA...VIHYSIVP
Mpob	111	IVGKKGQRV.WTGGG..DEKYLKGIWNAYRY..HN	LRVSNQAPLDMYKEVNTGTNLPFA...QIDLAVD
Burkholderia_xenovorans_FH	87	VFKVGMQDVRWDGATMGVTDMEGVRRAAYLPEVNV	LRASIVSPP.EGGRKTKDNTPA...VIHYSIVP
Pyrococcus_furiosus_FH	68	FFVETDGTWNVKEI...YETIAEAVKRRATDE..TP	LRNNAICLL...SGKVVG...NVPELHIEP..
Salmonella_typhimurium_FH	75	VIMEIGQHVCEWPKP...LKDQVNVQGRQGYEN..GY	LRKSMVADP.L.ERINTNDNTPA...LIHTEIVD
Helicobacter_hepaticus_FH	74	VFVEIGQDVHITGGY..LEDAINEGKKEGYTN..GY	LRKSVVEEP.LYERINTNDNTPA...VIHTRIK
Archaeoglobus_fulgidus_FH	71	FFVEVGRRELCLD...FDLKEAIVEGRRKATAE..TP	LRNNAVHPI...TRENSGDNTGLGIPQINIVLVK
Pelotomaculum_thermopropionicum_FH	76	IFLELQDQVRIE.GGD..LYEAINEGVRKGYTE..GY	LRKSIWVGH.L.ERKNTGDNTPA...VIHTRIK
E_coli_TTD	79	FFVKVGSRRFLLG...ELQSIKQAVEEATVK..AP	LRHNAVEIF...DEVNTGKNTSGGVWVWTDIIP
Shigella_sonnei_TTD	79	FFVKVGSRRFLLG...ELQSIKQAVEEATVK..AP	LRHNAVEIF...DEVNTGKNTSGGVWVWTDIIP
Pseudomonas_putida_TTD	78	YFUSAGARFPLLG...EMEGILENATKEATIK..GP	LRHNAVETTF...IEKNTGNTSGKIPWLDWEIIP
Methanocaldococcus_janschii_FH	132	SLDREIEITIAEIKGAGSINMSA...KMIKPSDGIETK	KNFVLEBTANAGKPCPPILVVGIGIGGTA.
E_coli_FumA	175	..GDEYKFLC...KGGSSNKTYLYQETKALLTPG...	KLKNYLVEKMRTLGTAACTPPYHIAFVIGGTA
Leishmania_major_FH	203	..GSDYEFLLF...KGGSSNKAYLYQETKALLPK...	SRAFIEEKKTLGTAACTPPYHIAFVIGGTA
Shigella_sonnei_FumA	175	..GDEYKFLC...KGGSSNKTYLYQETKALLTPG...	KLKNYLVEKMRTLGTAACTPPYHIAFVIGGTA
Pseudomonas_putida_FumA	140	..GDKVVDVVA...KGGSSNKSKM...AMLNPSDSI...	VDWVLKVTMTMGAGKCPPLMGIGIGGTA
Mpob	173	..GMEYKFLF...KGGSSNKTYLYQETKALLTP...	SLEKFLVEKMRTLGTAACTPPYHIAFVIGGTA
Burkholderia_xenovorans_FH	153	..GNTVDVQVA...KGGSSNKSKF...AMLNPSDSI...	VDWVLKVTMTMGAGKCPPLMGIGIGGTA
Pyrococcus_furiosus_FH	122	..GKRNRTIS...KGGSSNCSAL...FPITPGDWFEG	KKKIIDHTRSCGGKCPPIIVGIGIASFA.
Salmonella_typhimurium_FH	136	..GDRVTITV...KGGSSNMGTF...KTLPPGDGIDG	KDFVLEWRRVGGKCPPIIVGIGIGGTM.
Helicobacter_hepaticus_FH	136	..GDKLHLKVC...KGGSSNKSVAL...KMLVPTADG	IEGVKKVFEAVKNAAGPNACTPPMVIIVGIGGTM.
Archaeoglobus_fulgidus_FH	133	..GDELRMV...KGGSSNVSAL...KMLLPTQVD	KIKDFVLEWKNAGGKCPPIIVGIGIGGTM.
Pelotomaculum_thermopropionicum_FH	137	..GDKLRTIV...KGGSSNMSAI...KMLRPTADG	VEGVKKFVIDWKAAGPNACTPPMVIIVGIGGTF.
E_coli_TTD	141	DND.DAETEV...KGGSSLPGRS...KVLMPSEGYEG	VVKFVFENISTIAVNACTPPVIVGIGIATSV.
Shigella_sonnei_TTD	141	DND.DAETEV...KGGSSLPGRS...KVLMPSEGYEG	VVKFVFENISTIAVNACTPPVIVGIGIATSV.
Pseudomonas_putida_TTD	140	DAD.YAIVD...KGGSSLPGSA...KVLMPGQGYEG	VTEVVFVDTSRGVNACTPPLVIGVGVSTV.
Methanocaldococcus_janschii_FH	196	DVAIKLARKKALRKRGERH...RDKEIANLEKELEKINS	LGICAMGLGCDITALDVFTEIAGCGHT
E_coli_FumA	239	ETNLKTVHLASAKRYDELPTEGNEHGQAFRDVELEKELL	LIEACNLGICAQ.FGCKYFAHDIRVIRLFRPHG
Leishmania_major_FH	267	EMTKTVHLASAKRYDLSLPTTGDKRYGAFRDVEPEWKIV	MEVAQKSGCAQ.FGCKYFAHDIRVIRLFRPHG
Shigella_sonnei_FumA	239	ETNLKTVHLASAKRYDELPTEGNEHGQAFRDVELEKELL	LIEACNLGICAQ.FGCKYFAHDIRVIRLFRPHG
Pseudomonas_putida_FumA	199	EKAAMARVLEMLSEVDELIEL..KARGQNRLEELKEL	LIEACNLGICAQ.FGCGLTITVLVDKIMDYPRHA
Mpob	237	EGCCLKTVLAASTGYLDALDPTGNKGQAFRDLEMEDKL	LKAAKLGICAQ.FGCKHFVLDVRRVRLFRPHG
Burkholderia_xenovorans_FH	212	EKAAMARVLEMLSDPDIQDVTARGGKDWIEELRVEL	HEKVNALGICAQ.FGCGLATVLDVRRVRLFRPHG
Pyrococcus_furiosus_FH	184	ERAMVMAKESLFRKIGERH...E.KLGVLEELIELEINFLD	GPMLGGKTTALDVKVHMAEPRHA
Salmonella_typhimurium_FH	198	HRATTLAKKSLFRKIGERH...AKPLYAQLEALLEALNNTG	GPMLGGKTTALDVKVHMAEPRHA
Helicobacter_hepaticus_FH	198	HRCSSWMAKSLFRKIGERH...AKPLYAQLEALLEALNNTG	GPMLGGKTTALDVKVHMAEPRHA
Archaeoglobus_fulgidus_FH	194	EKAAMARVLEMLSDPDIQDVTARGGKDWIEELRVEL	HEKVNALGICAQ.FGCGLATVLDVRRVRLFRPHG
Pelotomaculum_thermopropionicum_FH	199	DGAAMARVLEMLSDPDIQDVTARGGKDWIEELRVEL	HEKVNALGICAQ.FGCGLATVLDVRRVRLFRPHG
E_coli_TTD	204	EKAAMARVLEMLSDPDIQDVTARGGKDWIEELRVEL	HEKVNALGICAQ.FGCGLATVLDVRRVRLFRPHG
Shigella_sonnei_TTD	204	EKAAMARVLEMLSDPDIQDVTARGGKDWIEELRVEL	HEKVNALGICAQ.FGCGLATVLDVRRVRLFRPHG
Pseudomonas_putida_TTD	203	ETAAVLSRKAALRPLGSRH...PNPAAELRLRLEGLNRLG	GPGLTGSNSVMGVNHESSARHP
Methanocaldococcus_janschii_FH	259	ASLFPVGTICQWARRRAIKRITKLDACKLMEY...FNKLT	K...KDV
E_coli_FumA	308	ASCFVGMGVSADRNRIKAKINRQGIWI...EKLEHNP	PKYIPEELRKAEGEAVRVDLNRMKIEILAQ
Leishmania_major_FH	336	ASCFVGLAVSADRQIILAHINKSGIYI...EQLEQNP	QAQYLPDIPVHLSTTSVKVDLNRDKVQQ
Shigella_sonnei_FumA	308	ASCFVGMGVSADRNRIKAKINRQGIWI...EKLEHNP	PKYIPEELRKAEGEAVRVDLNRMKIEILAQ
Pseudomonas_putida_FumA	267	ASLFPVCMIPNCAARRAHAFVLDGTGPALEAPSLDAY...	PEIWEAGP.SARRVNLDDITP...EEV
Mpob	306	ASLFPVCMIPNCAARRAHAFVLDGTGPALEAPSLDAY...	PEIWEAGP.SARRVNLDDITP...EEV
Burkholderia_xenovorans_FH	280	ASLFPVCMIPNCAARRAHAFVLDGTGPALEAPSLDAY...	PEIWEAGP.SARRVNLDDITP...EEV
Pyrococcus_furiosus_FH	245	ASLFPVCMIPNCAARRAHAFVLDGTGPALEAPSLDAY...	PEIWEAGP.SARRVNLDDITP...EEV
Salmonella_typhimurium_FH	261	TALFPVAVNINCAARRAHAFVLDGTGPALEAPSLDAY...	PEIWEAGP.SARRVNLDDITP...EEV
Helicobacter_hepaticus_FH	261	TALFPVAVNINCAARRAHAFVLDGTGPALEAPSLDAY...	PEIWEAGP.SARRVNLDDITP...EEV
Archaeoglobus_fulgidus_FH	251	ASLFPVAVNINCAARRAHAFVLDGTGPALEAPSLDAY...	PEIWEAGP.SARRVNLDDITP...EEV
Pelotomaculum_thermopropionicum_FH	262	ASLFPVAVNINCAARRAHAFVLDGTGPALEAPSLDAY...	PEIWEAGP.SARRVNLDDITP...EEV
E_coli_TTD	267	STIGVAVSTGCAARRRGTLLVHADLTLENLS...HTRSALM	KKILITTPKA...EDL
Shigella_sonnei_TTD	267	STIGVAVSTGCAARRRGTLLVHADLTLENLS...HTRSALM	KKILITTPKA...EDL
Pseudomonas_putida_TTD	266	STIGVAVSTGCAARRRGTLLVHADLTLENLS...HEGVVLM	KKILITTPSD...EDL
Methanocaldococcus_janschii_FH	299	KKLVGDIVVYNGKITYTARDEARLKIEMKSNKELP	FDLNEISITVHAGPMMKK.VNDSWVCSGPTTTS
E_coli_FumA	375	SOYPVSTRVSLNGTIVIGRDIAHAKIKERMONGEGLP	QYIKDHPVYAGPAPKTP...GYASGSGPTTA
Leishmania_major_FH	403	SOYPVSTRVSLNGTIVIGRDIAHAKIKERMONGEGLP	QYIKDHPVYAGPAPKTP...GYASGSGPTTA
Shigella_sonnei_FumA	375	SOYPVSTRVSLNGTIVIGRDIAHAKIKERMONGEGLP	QYIKDHPVYAGPAPKTP...GYASGSGPTTA
Pseudomonas_putida_FumA	328	ASWKPGETILLNGKMLTGRDARHAKRMVEMLNREGE	ELPVDLKGFRVYVGPDPVVG...DEVVGPACTTA
Mpob	373	SKYPVTQLSLTGTIVIGRDIAHAKIKERIDKGEGLP	QYVVDHAIYVAGPAPKTP...GMASGSGPTTA
Burkholderia_xenovorans_FH	342	AAWTPGQTLKLSKMLTGRDARHAKRMVEMLNREGE	ELPVDLKGFRVYVGPDPVVG...DEVVGPACTTA
Pyrococcus_furiosus_FH	290	RDLIGDRVLLSGLTYTARDEARLKIEMKSNKELP	FDLNEISITVHAGPMMKK.VNDSWVCSGPTTTS
Salmonella_typhimurium_FH	294	KTLKVGDIILFTGAYTCRDAARHAKKLEVALSNKES	PSVDFQNTQIVYAGPAPKTP...GLPIVSGPTTA
Helicobacter_hepaticus_FH	298	KSLKAGESVLLSGLTYTARDEARHAKKLEVALSNKES	PSVDFQNTQIVYAGPAPKTP...GLPIVSGPTTA
Archaeoglobus_fulgidus_FH	287	LKLVGQDVVYTCGTYTARDEARHAKKLEVALSNKES	PSVDFQNTQIVYAGPAPKTP...GLPIVSGPTTA
Pelotomaculum_thermopropionicum_FH	298	EKLIRVGRVLLSGLTYTARDEARHAKKLEVALSNKES	PSVDFQNTQIVYAGPAPKTP...GLPIVSGPTTA
E_coli_TTD	318	QDIRVGDVYVLTGTYTARDEARHAKKLEVALSNKES	PSVDFQNTQIVYAGPAPKTP...GLPIVSGPTTA
Shigella_sonnei_TTD	318	QDIRVGDVYVLTGTYTARDEARHAKKLEVALSNKES	PSVDFQNTQIVYAGPAPKTP...GLPIVSGPTTA
Pseudomonas_putida_TTD	317	ASLNVGDIVVYTCGTYTARDEARHAKKLEVALSNKES	PSVDFQNTQIVYAGPAPKTP...GLPIVSGPTTA
Methanocaldococcus_janschii_FH	368	ARMNDVEEPIKLTNISAIVGKIGMCKEQLRTFEDYGV	VYIAPAGCAALL.ANSVKKRVNDVYVLDLGLM
E_coli_FumA	442	GRMDSYVDLQQAQGSIMILAKNRSRQVTDACKKHGG	FYLGSIAGCAAVLAAQSTIKSLECVL.YEPLGLM
Leishmania_major_FH	470	GRMDSYVDLQQAQGSIMILAKNRSRQVTDACKKHGG	FYLGSIAGCAAVLAAQSTIKSLECVL.YEPLGLM
Shigella_sonnei_FumA	442	GRMDSYVDLQQAQGSIMILAKNRSRQVTDACKKHGG	FYLGSIAGCAAVLAAQSTIKSLECVL.YEPLGLM
Pseudomonas_putida_FumA	395	TRMDKFTROILEQTLGIMIGKSERGPTAIEAIDKNK	AVYLLMAVAGCAAYLV.AQATIRKSKVLA.YEPLGLM
Mpob	440	GRMDSYVDLQQAQGSIMILAKNRSRQVTDACKKHGG	FYLGSIAGCAAVLAAQSTIKSLECVL.YEPLGLM
Burkholderia_xenovorans_FH	409	TRMDKFTROILEQTLGIMIGKSERGPTAIEAIDKNK	AVYLLMAVAGCAAYLV.AQATIRKSKVLA.YEPLGLM
Pyrococcus_furiosus_FH	350	ARMKAYLD.FLFRVGRVIGKAMNHEKFKG...RAVYFS	YSPAGSVA.SKTRIVKRVYVW.YEPLGLM
Salmonella_typhimurium_FH	361	ARMKAYLD.FLFRVGRVIGKAMNHEKFKG...RAVYFS	YSPAGSVA.SKTRIVKRVYVW.YEPLGLM
Helicobacter_hepaticus_FH	365	GRMDKFTYPTIE.QGATHEIKGYSRSEVVIDSMVKH	GVVYLVAVAGCAAAI.LTKHIVKVEVA.YEPLGLM
Archaeoglobus_fulgidus_FH	355	ARMNPTTPKILEKVECMGIIKGMSEVVEAMRGK.AA	YFAFTAGAGALA.AMSIKKVKGVV.WEPLGLM
Pelotomaculum_thermopropionicum_FH	365	YRMDPVPAPKLYA.LGLKATIGKRRSPVVIKAMKQY	KGVYLLAATAGCAALL.ARCVTKDAKVLA.YEPLGLM
E_coli_TTD	384	MRMPSFEREFTEQTKVLLVGGKMGPLTEGCKQFK	ALHVFPAAGCAVLA.ATQVEELERH.WEPLGLM
Shigella_sonnei_TTD	384	MRMPSFEREFTEQTKVLLVGGKMGPLTEGCKQFK	ALHVFPAAGCAVLA.ATQVEELERH.WEPLGLM
Pseudomonas_putida_TTD	384	MRMPSFEREFTEQTKVLLVGGKMGPLTEGCKQFK	ALHVFPAAGCAVLA.ATQVEELERH.WEPLGLM

```

Methanocaldococcus_janschii_FH 437 P E A V W E L E V N N F G P I I V A M D S H C N S I Y E E V N K K V Y E K L N E L I G L . . . . .
E_coli_FumA 511 . E A I W K I E V E D F . P A F I L V D D R G N D F F Q Q I Q L T Q C T R C V K . . . . .
Leishmania_major_FH 539 . E A V W K I E V E D F . P A F I V V D D R G N D M Y S K T L A . . . . .
Shigella_sonnei_FumA 511 . E A I W K I E V E D F . P A F I L V D D R G N D F F Q Q I Q L T Q C T R C V K . . . . .
Pseudomonas_putida_FumA 463 . E A I Y E F E V K D M . P V T V A V D S N G E S V H I T G P A L W Q S K I A E S L A V E V K . . . . .
Mpob 509 . E A I W K I E V E D F . P A F I L V D D R G N D F F Q Q L S C . . . . .
Burkholderia_xenovorans_FH 477 . E A I Y E F D V Q D M . P V T V A V D S N G T S V H Q T G P K E W Q A R I G K I P V A T V . . . . .
Pyrococcus_furiosus_FH 413 T D A V W E L E V R D M . P L L V S I D S R G R S L Y R K N . . . . .
Salmonella_typhimurium_FH 428 . E S I K K L Y F D H Y . R V I V G I D S E G N T L Q K Q E V P K Y A K . . . . .
Helicobacter_hepaticus_FH 432 . E A V A R L T I E N F . P A I V A I D A S G N N Y Y E V G Q A P Y K K I . . . . .
Archaeoglobus_fulgidus_FH 422 P E A V W L L E V E R F G P C I V A I D A H C N S L Y R R . . . . .
Pelotomaculum_thermopropionicum_FH 432 . E A I H E L V E D L . P V I V V N D T L G D L Y D E G I K I Y G V K . . . . .
E_coli_TTD 452 P E S L W V C R V K E F G P L I V S I D T H C N N I I A E N K K L F A E R R D P I V E E I C E H V H Y I K
Shigella_sonnei_TTD 452 P E S L W V C R V K E F G P L I V S I D T H C N N I I A E N K K L F A E R R D P I V E E I C E H V H Y I K
Pseudomonas_putida_TTD 452 P E T L W V N R V R E F G P L I L S I D T R G N N I F E Q N K A R F N E R K G A V I E K I N S Q V R F I K

```

Fig. 2.2 Multiple sequence alignment of biochemically characterized class I FH and L-TTD. KGXGS motif present as an AGGGC motif in L-TTD is highlighted in a box in the alignment. Linker that connects the NTD and CTD of single-subunit FH has been underlined. Single subunit FH from *E. coli*, *S. sonnei* and *Pseudomonas putida* are annotated as *fumA* in the alignment. Single subunit FH from *Syntrophic propionate-oxidizing bacteria* is abbreviated as *Mpob* in the alignment.

2.3.2. Phylogenetic analysis

Phylogenetic tree drawn with a large dataset of 258 sequences of class-I FH and L-TTD (**Fig. 2.3**) reveals a conserved motif in these two proteins as was observed in an earlier report (Reaney, Bungard, et al., 2009). All experimentally studied class-I FH and the sequences that group along with them show the presence of a conserved ‘KGXGS’ motif. Considering the high identity that the two proteins share, a number of sequences in the database are annotated as FH/L-TTD. Sequences annotated as L-TTD or as FH/L-TTD but grouping with characterized fumarases in the phylogenetic tree had KGXGS motif. These are mostly archaeal sequences, and no L-tartrate dehydratase activity has been reported in archaea. In addition, Archaea do not possess two genes coding independently for FH and L-TTD and hence with fair confidence these sequences should be two-subunit fumarases. Single subunit bacilli sequences have a KGGGC motif instead of a KGXGS or AGGGC motif. These sequences fall in the same group as *Bacillus stearothermophilus*, which has been studied as a single subunit FH (Reaney, Bungard, et al., 2009). It was also noticed that *B. stearothermophilus* FH shares more similarity with L-TTD sequences than other fumarases which is reflected in the tree. Both archaeal class-I two-subunit FH and bacterial L-TTD originate from the same ancestral node suggesting a possible, horizontal gene transfer event between the two, similar to what was observed using fewer L-TTD sequences in an earlier report (Kronen & Berg, 2015). Bacilli FH could be an intermediate in this horizontal gene transfer (HGT) process from archeal FH to bacterial L-TTD considering its origin from the same ancestral node as the two proteins. This is also substantiated by the fact that Bacilli FH from 5 different genus (*paenibacillus*, *brevibacillus*, *lysini bacillus*, *fictibacillus* and *geobacillus*) shared more similarity with L-TTD than

FH and have a KGGGC motif, sharing a similarity with both proteins. Single-subunit class-I FH divide into two separate clades with one clade closer to the two-subunit type containing proteobacteria and firmicutes and the second clade grouping separately containing protozoa and actinobacteria.



Fig. 2.3. Phylogenetic tree (cladogram) for fumarate hydratase Class-I and L-tartrate dehydratase protein sequences. Single-subunit class-I FH and concatenated sequences of two-subunit class-I FH/L-TTD, representing all the taxa in which these proteins are present were used for the analysis. The tree was constructed using the Maximum Likelihood method and Le Gascuel model performing 500 bootstrap replicates (Felsenstein, 1985). Highlighted in blue are the single-subunit class-I fumarases, red – archaeal two-subunit class-I fumarases, green – bacterial L-tartrate dehydratases, yellow – single-subunit Bacilli class-I fumarases and black – sequences annotated as either or both FH/L-TTD having KXGS motif.

2.3.3. Cloning, expression, and purification

E. coli L-TTD clone was made in pET-Duet vector that permits co-expression of genes coding for subunits TTD α and TTD β from independent promoters (**Fig. 2.4A**). WT clone was made and validated by insert release (**Fig. 2.4B**) and sequencing. The clone generated was used for protein expression in BL21 (DE3)-RIL strain. Upon induction with 300 μ M IPTG, there was significant overexpression of the desired protein subunits (**Fig. 2.4C**). WT protein was obtained to good purity upon Ni-NTA chromatography as can be seen from the imidazole elutes analyzed on SDS-PAGE (**Fig. 2.4C**). Swapping the AGGGC motif from L-TTD with KGGGS motif from MjFH lead to the lack of expression of soluble α -subunit with only expression of the β -subunit seen (**Fig. 2.4C**). Single-subunit FH was aligned with two-subunit FH to decipher the boundary in-between N- and C-terminal domains in single-subunit proteins (**Fig. 2.4D**). A segment of 29 residues corresponding to the loop that connects these two domains was deleted in *E. coli FumA* and the two domains cloned separately. Clones of *fumA*, NTD and CTD were generated in pET-Duet vector (**Fig. 2.4E**) and validated through insert release (**Fig. 2.4F**). All the three proteins were found to be expressed in significant amounts upon induction with 500 μ M IPTG (**Fig. 2.4G**). *FumA* was obtained without contamination, at good purity upon Ni-NTA chromatography as can be seen from the imidazole elutes analyzed on SDS-PAGE (**Fig. 2.4G**).

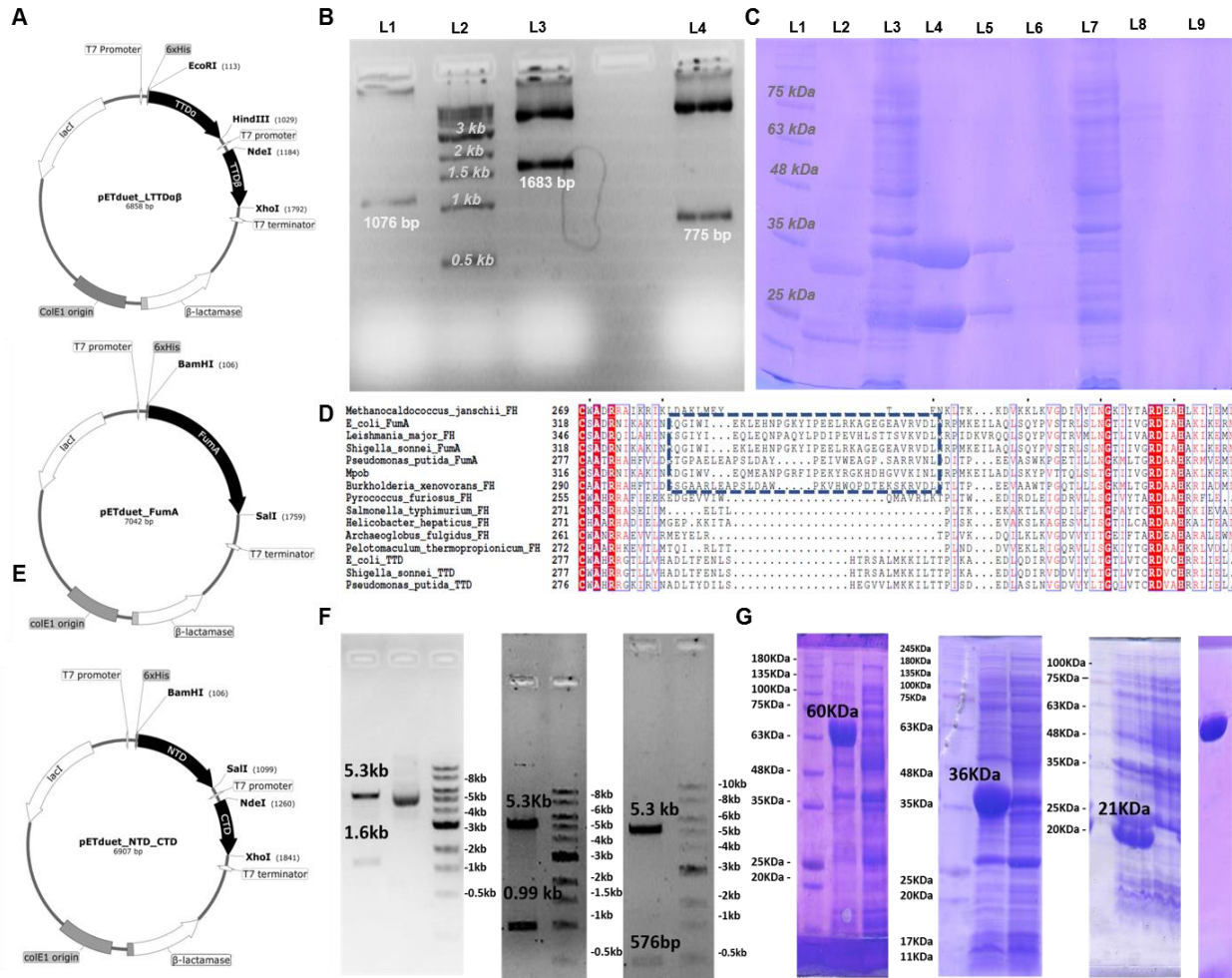
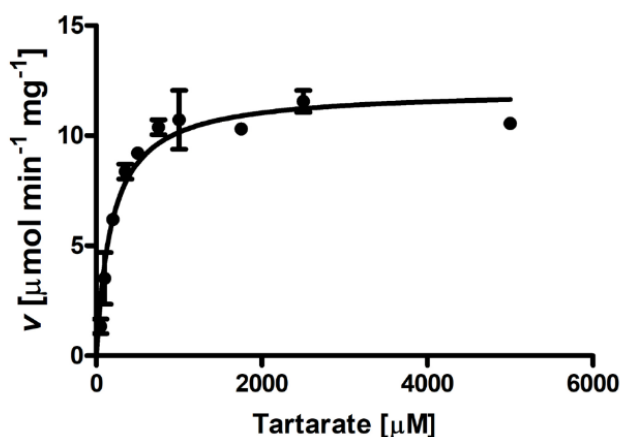


Fig. 2.4. Cloning, expression, and purification of *E. coli* L-TTD and *fuma*. **A.** Vector map showing the restriction enzymes sites used for cloning the two subunits of L-TTD. Similar constructs with only the individual subunits were also made. **B.** Cloning of both the subunits was validated by releasing the insert using appropriate restriction enzymes. L1, L3 and L4 represent L-TTD α , L-TTD $\alpha\beta$ and L-TTD β clones which were digested with *EcoRI* and *XhoI* to release the insert containing the respective genes of size 1076, 1683 and 775 base pairs. DNA ladder was loaded in L2. **C.** SDS-PAGE showing the expression and purification of L-TTD WT and KGXGS motif swapped mutant. L1, protein molecular weight marker; L2, purified MjFH $\alpha\beta$ protein as control; L3, lysate from WT L-TTD expressing (DE3)-RIL cells; L4, L5 and L6, 250 μ M imidazole eluates; L7, lysate from mutant KGXGS motif containing L-TTD expressing (DE3)-RIL cells; L8 and L9, 250 μ M imidazole eluates. **D.** Multiple sequence alignment of class-I FH. Highlighted in the box is the linker region that connects the N- and C-terminal domains of single-subunit class I FH. Single subunit FH from *E. coli* and *Pseudomonas putida* are annotated as *fumA* in the alignment. Single subunit FH from Syntrophic propionate-oxidizing bacteria is abbreviated as Mpob in the alignment. **E.** Vector map showing the restriction enzymes sites used for cloning *E. coli* *fumA* and N-terminal and C-terminal domains of the protein. Similar constructs with only the individual subunits were also made. **F.** Plasmids containing genes for *fumA* and its subdomains was validated by releasing the insert using appropriate restriction enzymes. FL-*fumA*, *fumA* NTD and *fumA* CTD clones were digested with *BamHI* and *XhoI* to release the insert containing the respective genes of size 1647, 987 and 576 base pairs, respectively. **G.** SDS-PAGE showing the over-expression of *fumA*, *fumA* NTD and *fumA* CTD proteins in

BL21(DE3)-RIL cell lysate along with the lysates of uninduced cells as control; Purified 500 μ M imidazole eluate of FumA can be seen in the final gel image.

2.3.4. Kinetic characterization of L-TTD

Reconstituted L-TTD $\alpha\beta$ complex was tested for activity on L-tartrate employing a coupled enzyme assay, in which oxaloacetate obtained from L-tartrate conversion was utilized by the coupling enzyme malate dehydrogenase, converting NADH to NAD⁺, resulting in a drop in absorbance at 340 nm that was measured using spectrophotometry. Kinetic parameters for the enzyme were derived from a v versus [L-tartrate] plot followed by fitting the data to Michaelis-Menten equation. V_{\max} value of $12 \pm 0.4 \mu\text{mol min}^{-1} \text{mg}^{-1}$ and K_m value of $190 \pm 30 \mu\text{M}$ was obtained from the fit (**Fig. 2.5**). Catalytic efficiency (k_{cat}/K_m) of *E. coli* L-TTD turned out to be $5.88 \times 10^4 \text{ s}^{-1} \text{M}^{-1}$, which is almost the same as *E. coli* FumB on D-tartrate ($1.2 \times 10^4 \text{ s}^{-1} \text{M}^{-1}$) and 10-fold higher than the earlier reported catalytic efficiency of *E. coli* FumA on D-tartrate ($3 \times 10^3 \text{ s}^{-1} \text{M}^{-1}$) (van Vugt-Lussenburg et al., 2013).



Enzyme	V_{\max} ($\mu\text{mol min}^{-1} \text{mg}^{-1}$)	K_M (mM)	k_{cat}/K_M ($\text{s}^{-1} \text{M}^{-1}$)
L-TTD	12 ± 0.4	0.19 ± 0.3	5.88×10^4

Fig. 2.5. Plot of initial rate vs L-tartrate concentration using reconstituted L-TTD. Kinetic parameters obtained from the Michaelis-Menten plot are represented in the table.

2.3.5. Substrate specificity in C-I FH and L-TTD

Tartrate is different from malate by only the presence of an extra hydroxyl group on the C-3 carbon (**Fig. 2.6A**). Tartrate has two chiral centers which gives rise to four stereoisomers, R,R tartrate (L-tartrate), S,S tartrate (D-tartrate) and R,S tartrate/S,R tartrate (meso-tartrate) (**Fig. 2.6B**). Multiple

reports have revealed that class-I FH also have a D-tartrate dehydratase activity and are inactive on L-tartrate (Kronen & Berg, 2015; van Vugt-Lussenburg et al., 2009, 2013). Meso-tartrate has been shown to be an inhibitor of class-I FH (Genda et al., 2006) and isn't shown to be a substrate for L-TTD. Promiscuous activity of L-TTD on fumarate or L- and D- malate has not been probed into earlier. To test if the two enzymes, class-I FH and L-TTD have opposite stereospecificity and understand how different the two are with regard to substrate specificity, activity of MjFH and *E. coli* L-TTD on the above-mentioned substrate analogues were tested.

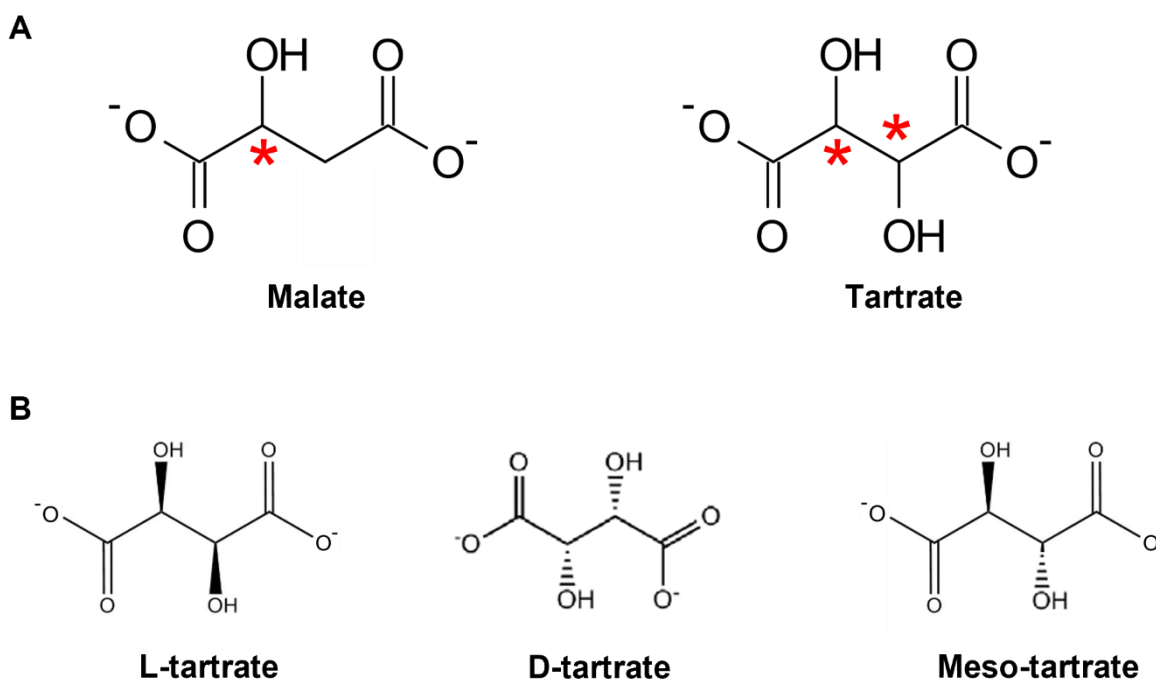


Fig. 2.6. Stereoisomers of tartrate. A. Tartrate is different from malate by only the presence of an extra hydroxyl group on C3 carbon and two chiral centers (Red Asterix) which gives rise to the diastereomers, B. L-tartrate (*R,R* tartrate), D-tartrate (*S,S* tartrate) and meso-tartrate (*S,R* or *R,S* tartrate).

2.3.5.1. Class-I FH substrate specificity

MjFH was found to be active on D-tartrate as has been reported in the literature previously (Kronen & Berg, 2015; van Vugt-Lussenburg et al., 2009, 2013) (**Fig. 2.7A**). Progress curves of the reaction were slightly noisy as the reactions were carried out at 37 °C, the optimum temperature of the coupling enzyme, mesophilic malate dehydrogenase, instead of the optimal temperature of 70 °C of MjFH. Nevertheless, activity of MjFH on D-tartrate further confirmed that class-I FH can utilize D-tartrate as substrate. Testing the activity of MjFH on 1 mM L-malate in the presence of 1 mM

L-tartrate or 1 mM D-malate in the assay mixture did not reduce the activity of the enzyme going to show that these enantiomers of the substrates do not even bind to the enzyme active site (**Fig. 2.7B**). However, L-tartrate has been shown to be an inhibitor of *E. coli fumA* and testing the activity of the enzyme on 500 μM L-malate in the presence of 250 μM and 500 μM L-tartrate respectively lead to a decrease in the slope of the progress curves confirming that *E. coli fumA* is indeed inhibited by L-tartrate (Flint, 1994) (**Fig. 2.7D**). Meso-tartrate turned out to be a strong competitive inhibitor of MjFH, inhibiting the enzyme with a K_i of $161 \pm 17.4 \mu\text{M}$ (**Fig. 2.7C**).

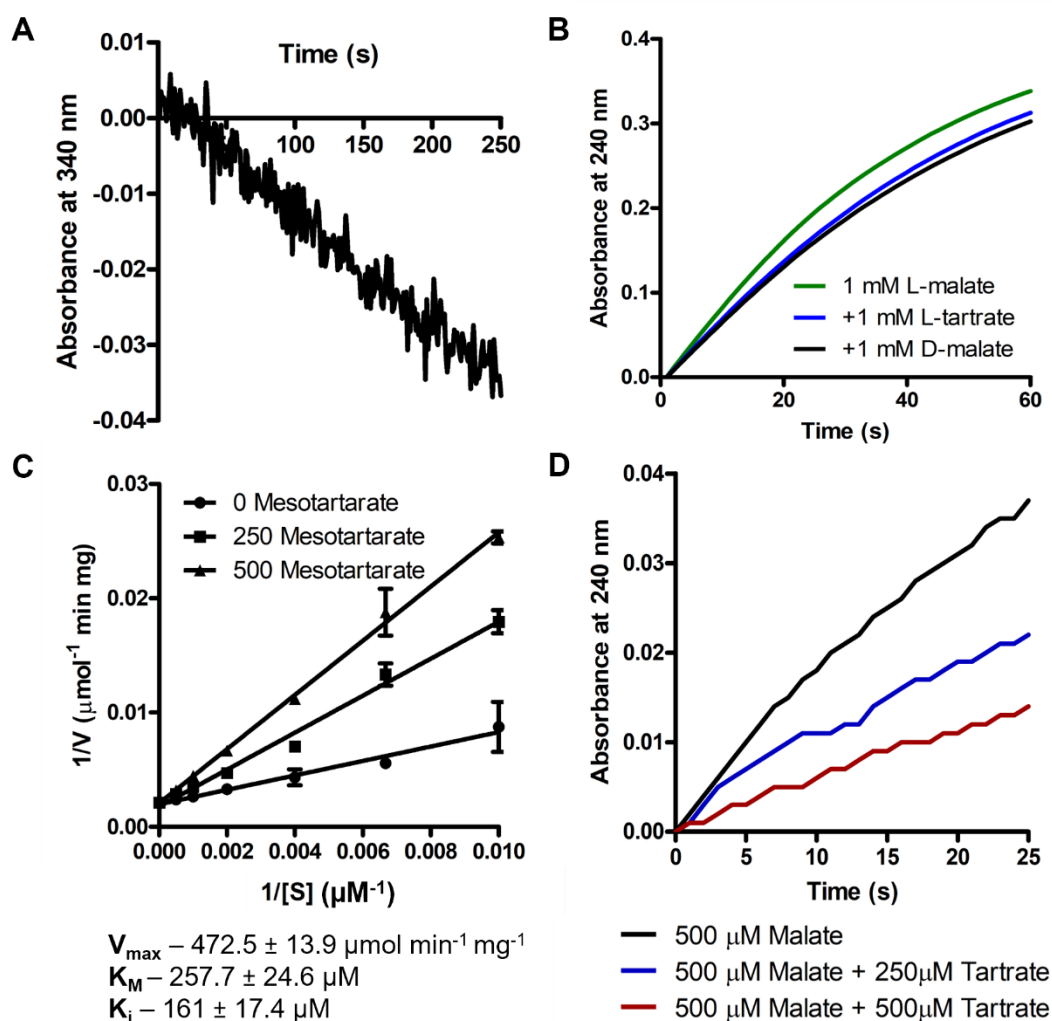


Fig. 2.7. Substrate specificity of class-I FH. **A.** Representative progress curve for the conversion of D-tartrate to oxaloacetate by WT MjFH. **B.** Representative progress curves for the conversion of 1mM L-malate to fumarate by WT MjFH (green line) in the presence of 1mM L-tartrate (blue) and 1 mM D-malate (black) going to show that MjFH is not inhibited by either L-tartrate or D-malate. **C.** Double reciprocal plots with meso-tartrate as inhibitor and L-malate as substrate. Concentration of L-malate was varied at different fixed concentrations of meso-tartrate (0, 250 and 500 μM). All experiments were repeated twice to confirm reproducibility. **D.** Representative progress curves for the conversion of 500 μM L-malate to

fumarate (black line) by E. coli fumA in the presence 250 μ M L-tartrate (blue line) and 500 μ M L-tartrate (brown line) showing that fumA is inhibited by L-tartrate.

2.3.5.2. L-TTD substrate specificity

Upon testing D-malate (at a concentration of 1 mM) as a substrate, L-TTD was found to be moderately active yielding reliable progress curves (**Fig. 2.8A**). Specific activity for 1 mM D-malate was $0.2 \mu\text{mol min}^{-1} \text{mg}^{-1}$ which was 50 folds lower than for its substrate L-tartrate. Testing the activity of the enzyme on fumarate did not give reliable progress curves and hence the specific activity could not be measured. To find out if L-TTD was capable of converting fumarate to D-malate, assay buffer containing 500 μ M fumarate was incubated with active reconstituted enzyme overnight followed by acquiring an absorbance spectrum from 200 and 300 nm. While addition of heat inactivated enzyme did not lead to a change in the spectrum compared to the fumarate only control, addition of active enzyme led to decrease in absorbance over the wavelength range of 200 to 300 nm. This clearly establishes that fumarate is converted to malate, albeit with a low specific activity (**Fig. 2.8B**). Further, to confirm that reduction in fumarate levels in the assay buffer was due to conversion to malate by L-TTD and not reduction to succinate, LC-ESI MS was carried out. Retention time of standards, malate, and fumarate on ZIC-pHILIC column were observed to be 20 – 30 minutes and 19 minutes respectively (**Fig. 2.9A, B**), yielding the expected mass for the two metabolites on MS (**Fig. 2.9C, D**). Fumarate incubated overnight with reconstituted L-TTD was found to have two peaks on the total ion chromatogram (TIC) (**Fig. 2.10A**), corresponding to the mass of malate and fumarate (**Fig. 2.10B, C**). Malate observed in MS has to be the D-isomer as D-malate is a substrate for L-TTD, while L-malate turned out to be a competitive inhibitor with a K_i of $260.9 \pm 28.8 \mu\text{M}$ (**Fig. 2.8D**). Appropriate control involving inhibition of the coupling enzyme (MDH) by L-malate at pH 7.4, the condition at which the coupled enzyme assay was carried out was included. Under these conditions, MDH was not inhibited by L-malate (Appendix C) going to show that the derived K_I value is a true reflection of L-TTD inhibition by L-malate. L-TTD was inactive on D-tartrate as evident from the progress curves for the reactions containing 1 mM L-tartrate alone and 1 mM L-tartrate with 1 mM D-tartrate, initiated by the addition of the enzyme (**Fig. 2.8C**). Similar to class-I FH, meso-tartrate turned out to be a strong competitive inhibitor of L-TTD, inhibiting the enzyme with a K_i of $148.6 \pm 14.5 \mu\text{M}$ (**Fig. 2.8E**).

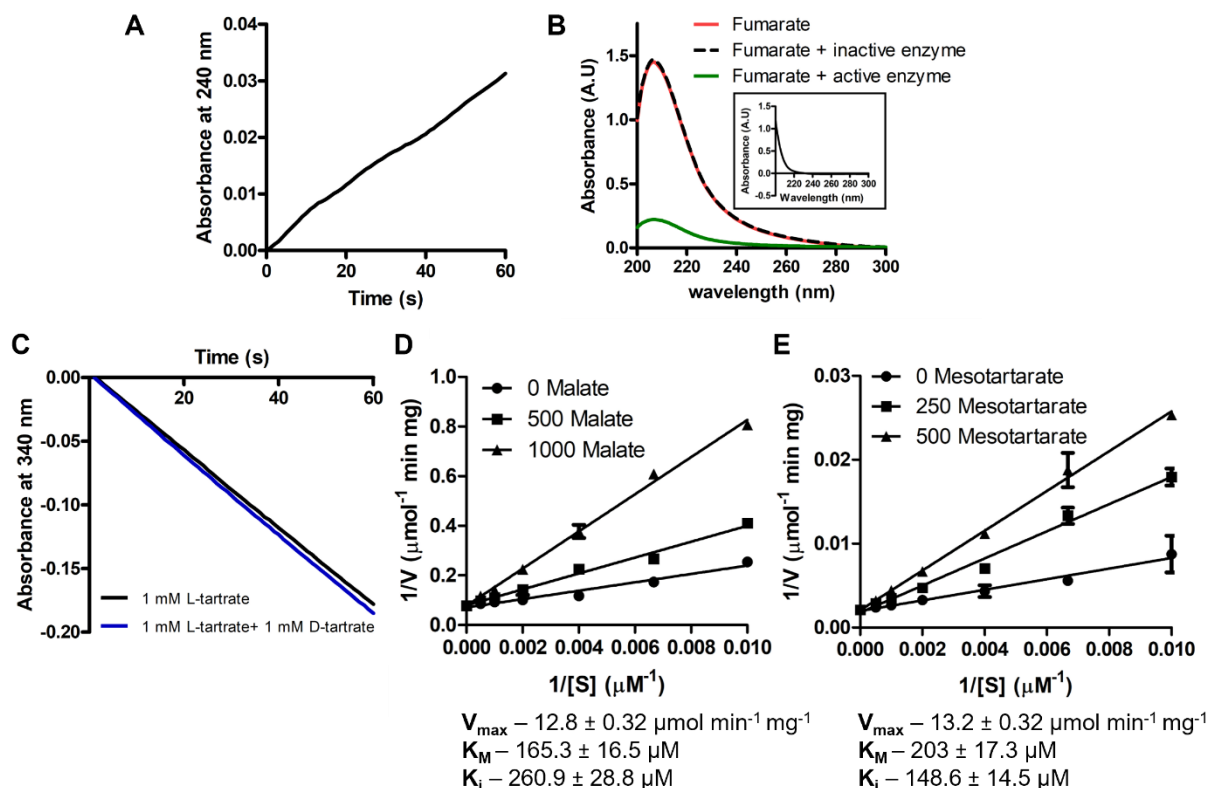


Fig. 2.8. Substrate specificity of L-TTD. **A.** Representative progress curve for the conversion of D-malate to fumarate by *E. coli* L-TTD. **B.** Spectral measurement in the wavelength range of 200 to 300 nm to monitor depletion of fumarate. Panel shows spectra of reaction mixture with fumarate but without enzyme (red line), fumarate with heat inactivated L-TTD (black dotted line) and fumarate with active reconstituted L-TTD (green line), all pre-incubated for 12 hours at room temperature. A decrease in absorbance over the range 200-300 nm with the active enzyme added shows that L-TTD converts fumarate to D-malate. Inset shows the buffer alone absorbance which was used as a blank for the test samples. **C.** Representative progress curves for the conversion of 1 mM L-tartrate to oxaloacetate by L-TTD (black line) in the presence of 1 mM D-tartrate (blue) show that *E. coli* L-TTD is not inhibited by D-tartrate. **D.** Double reciprocal plots with L-malate as inhibitor and L-tartrate as substrate. Concentration of L-tartrate was varied at different fixed concentrations of L-malate (0, 500 and 1000 μM). All experiments were repeated twice to confirm reproducibility. **E.** Double reciprocal plots with meso-tartrate as inhibitor and L-tartrate as substrate. Concentration of L-tartrate was varied at different fixed concentrations of meso-tartrate (0, 250 and 500 μM). All experiments were repeated twice to confirm reproducibility.

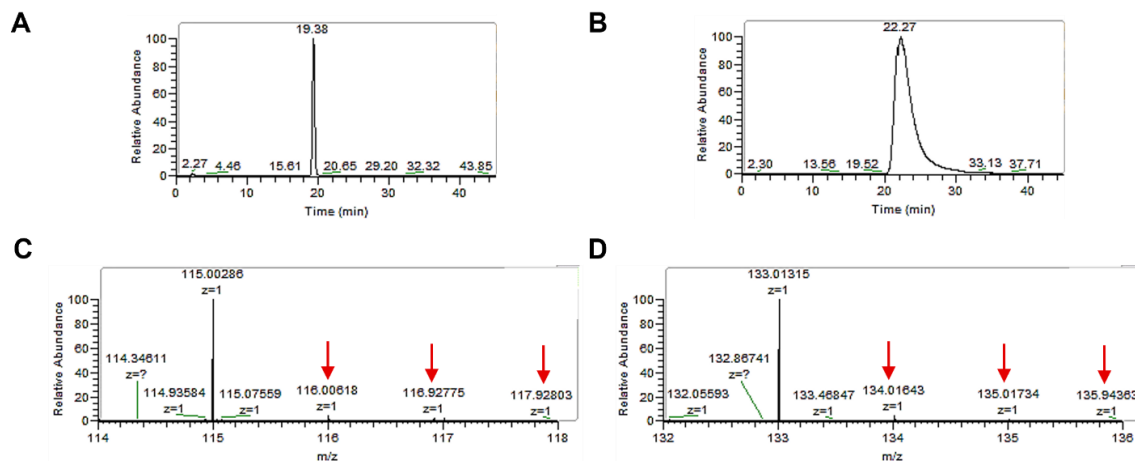


Fig. 2.9. Total ion chromatogram (TIC) and mass spectra of fumarate and malate standards. A, B. TIC of fumarate and malate with retention time (RT) of 19.38 min for fumarate and a broad peak of and 20 – 30 min for malate on ZIC-pHILIC column. C, D. Ion corresponding to the two metabolites was detected on MS. Observed mass of fumarate is 115.00289 and matches the expected mass of 115.00368 Da. Observed mass of malate is 133.01317 and matches the expected mass of 133.01425 Da. Isotopic distribution (indicated by red arrows) could be observed for both standards.

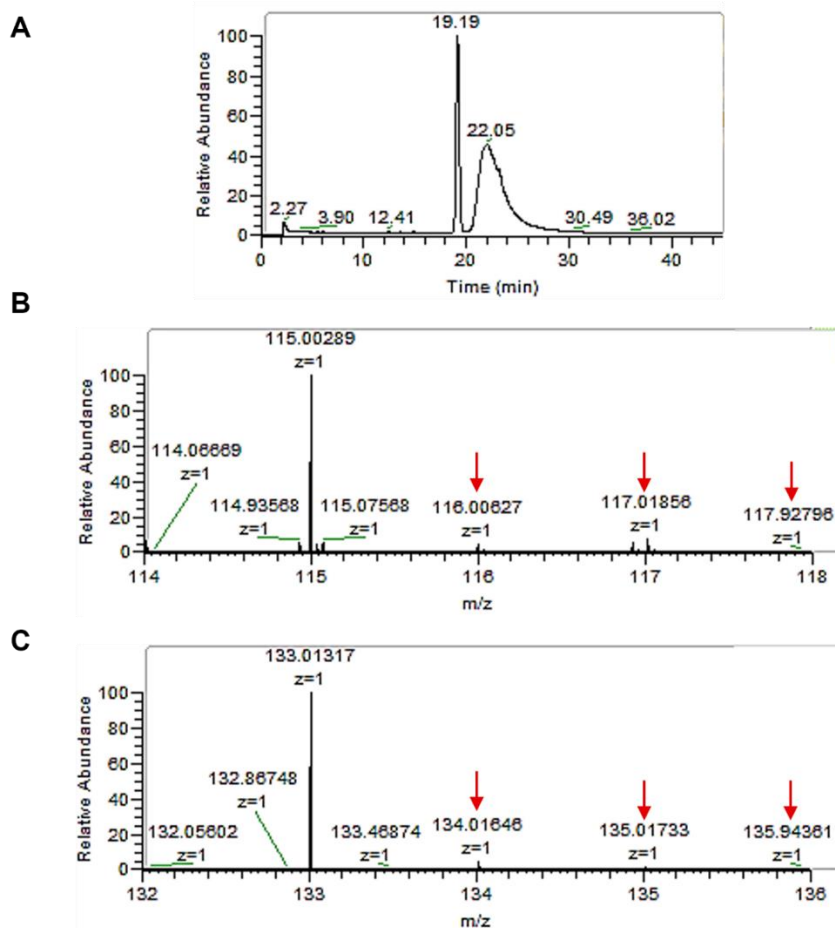


Fig. 2.10. Total ion chromatogram (TIC) and mass spectra of fumarate incubated with L-TTD. A. TIC shows two peaks, one at RT 19.19 min corresponding to fumarate and a broad peak between 20 – 30 min

for malate on ZIC-pHILIC column as was observed for the standards. **B.** Ion corresponding to fumarate on MS had an Observed mass of 115.00289 and matches the expected mass of 115.00368 Da. **C.** Ion corresponding to malate on MS had an observed mass of 133.01317 and matches the expected mass of 133.01425 Da. Isotopic distribution (indicated by red arrows) could be observed for both standards. Observation of malate in the MS of the sample confirmed L-TTD capability to convert fumarate to malate.

2.3.6. Motif swap to switch specificity in class-I FH and L-TTD

As KGXGS motif is the most obvious difference between the two groups of enzymes, class-I FH and L-TTD, the motif was inserted into *E. coli* L-TTD replacing the original AGGGC motif. The mutant enzyme unfortunately did not show any presence even in the insoluble fraction (**Fig. 2.4C**). Sequence alignment reveals that 2 residues in the motif and 4 residues flanking the motif are conserved differently in these two groups of enzymes. Using MjFH α subunit residue numbering, residues at position 140 and 142 which are non-polar in class-I FH are present as polar residues in L-TTD. Positions 143, 144, 148 and 150 have residues methionine, alanine, cysteine, and leucine conserved in L-TTD which are conserved differently with regard to residue property in class-I FH (**Fig. 2.11A**). These residues are present in a helix and a strand close to the substrate bound Fe-S cluster in the LmFH structure (**Fig. 2.11B**). Backbone of one of the glycy residues from the KGXGS motif (Gly147) in fact is in a hydrogen bonding distance to the substrate and mutants K144L and S148A in MjFH showed a drop in activity with decrease in K_M by 2-3 folds suggesting better accommodation of the substrate at the enzyme active site (details provided in chapter 1). As mutating these residues in L-TTD led to complete absence of expressed protein, attempts at swapping residues from L-TTD onto MjFH was attempted considering the archaeal enzyme is substantially stable. *In silico* mutating all the residues at one go in LmFH structure (512r) followed by energy minimization did not show short contacts and hence, mutations were sequentially incorporated into MjFH. 4 mutants were made, P143M + K144A (**MA**), K144A + S148C (**AC**), P143M + K144A + S148C (**MAC**) and I140D + F142Y + P143M + K144A + S148C + N150L (**DYMACL**). All the mutants expressed well and were obtained to good purity (**Fig. 2.11C**). While none of the mutants exhibited switch in stereospecificity, activity on malate although dramatically reduced was retained. Reliable progress curves were obtained for all mutants when testing for activity on 1 mM L-malate (**Fig. 2.11D**). There was a drop in specific activity between 35 to 700 folds for the mutants compared to WT MjFH (**Fig. 2.11E**). Since these mutants were active on L-malate, it allowed testing for activity on L-malate with L-tartrate in assay mix to understand if the mutants had acquired the anticipated property of L-tartrate binding. All the mutants displayed

similar activity both in presence and absence of L-tartrate in the assay buffer going to show that these mutants even failed to bind L-tartrate (**Fig. 2.11F**). The reduction in activity on malate must arise from a change in the conformation of the invariant Gly147, the backbone amino group of which is at a distance of 3.1 Å from the C2-OH group of bound malate in the LmFH structure. Mutation of Gly147 to alanine in MjFH led to an inactive mutant (chapter 1), going to show that the residue is critical for catalysis. The reduction in activity upon mutating neighboring residues as seen in this study, highlights the necessity for conservation of conformation in this segment for optimal catalysis. This can be confirmed with the availability of structure of the mutant enzyme.

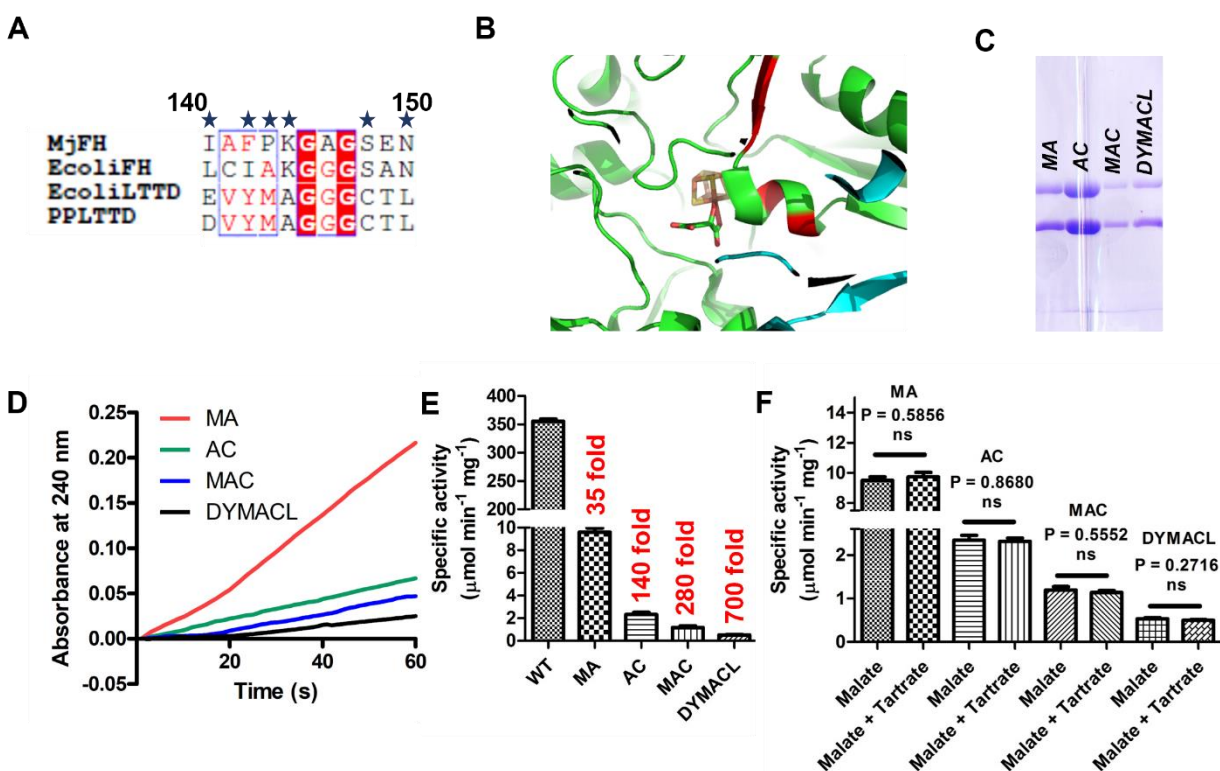


Fig. 2.11. Swapping the residues in the conserved motif from L-TTD onto MjFH does not lead to switch in stereospecificity. **A.** Sequence alignment between class-I FH and L-TTD shows that residues at positions 140, 142, 143, 144, 148 and 150 are conserved differently in class-I FH and L-TTD. **B.** Residues that correspond to the above 5 positions are located in a helix and a sheet close to the substrate and Fe-S cluster in the LmFH structure (5l2r) and are highlighted in red. **C.** SDS-PAGE showing the purified MjFH mutants MA in lane 1, AC in lane 2, MAC in lane 3 and DYMACL in lane 4. **D.** Representative progress curves for conversion of 1mM L-malate to fumarate by mutants MA (red line), AC (blue line), MAC (yellow line) and DYMACL (pink line). **E.** Bar graph for the specific activity of mutants MA, AC, MAC and DYMACL with 1 mM L-malate as substrate showing the fold change in activity in comparison to WT MjFH. **F.** Bar graph showing the comparison of specific activity for the mutants MA, AC, MAC and DYMACL with 1 mM L-malate as substrate in the presence/absence of 1 mM L-tartrate in the assay mixture. Statistical analysis was done using Student's unpaired t-test using Graphpad Prism.

2.3.7. Energy guided mutants to switch specificity in C-I FH

Iterative protein redesign and optimization (IPRO) suite of programs was used, which involves in silico mutagenesis followed by docking of the stereoisomer of interest and energy minimization to calculate binding energy. The procedure resulted in a series of mutant designs in which there was an improved binding of L-tartrate by class-I FH. WT LmFH enzyme was considered for the iterations due to the availability of its holoprotein structure (5l2r). Docking L-tartrate at the active site of the enzyme clearly showed a high ΔG value of +52.34 for L-tartrate binding indicating that the binding was not possible (**Fig. 2.12A**). However, a couple of active site residue mutations C252K + R471K resulted in significant decrease in binding energy of L-tartrate to LmFH (**Fig. 2.12B**). Further mutants were suggested where, in one of the mutant designs, a series of 9 mutations (C252K + R471K + R421K + G214K + G215R + S217P + P466R + C133R + R173K) led to a decrease in free energy of binding to a value of -107.4 kcal/mol (**Fig. 2.12C**). Two of the mutations C133R and C252K were omitted as these cysteine residues are the residues that ligate the Fe-S cluster and mutation of the same resulted in inactive mutants which failed to bind the Fe-S cluster (details in chapter 1). From the remaining 7 suggested mutations, a combination of 9 mutant designs were made in MjFH (**Fig. 2.12D**) and obtained to good purity (**Fig. 2.12E**). None of the mutants displayed activity on L-tartrate and since many of these mutants suggested were of the residues that bind the substrate at the active site, the mutants did not retain activity on L-malate also. As a result, it could not be deciphered if the mutants even bound L-tartrate.

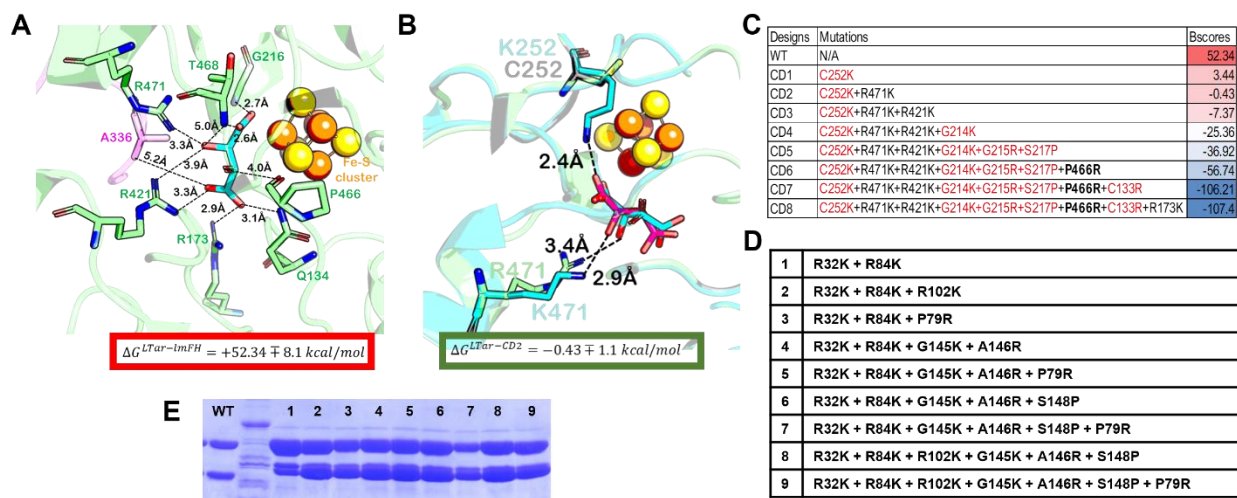


Fig. 2.12. L-tartrate docking to WT LmFH and in silico mutant design of LmFH. **A.** L-tartrate was docked into the active site of WT LmFH structure (5l2r), energy minimized, and binding energy calculated. A large positive ΔG value of +52.34 kcal/mol indicated a lack of feasible interaction. **B.** Mutant C252K

(C182K in MjFH) was suggested to improve L-tartrate binding at the active site from *in silico* analysis. A combination of just two mutants, R471K (R84K in MjFH) and C252K (C182K in MjFH) made binding of L-tartrate at LmFH active site feasible with a ΔG value of -0.43 kcal/mol. C. A list of LmFH mutants and their respective binding energies for the binding of L-tartrate at the active site. (Figures A, B & C were generated at Prof. Costas D. Maranas laboratory as a part of collaboration). D. Combination of mutants made in MjFH with the information derived from *in silico* design using LmFH structure. Cysteine mutants (C182K and C133R) were omitted from the list. E. SDS-PAGE showing the purified MjFH mutants loaded in lanes 1 to 9 in the order of the mutant list in panel D.

2.3.8. Examination of active site of class-I FH and predicted L-TTD structures

Although class-I FH apo- and holo-protein structures have become available in recent times, L-TTD structure has not been solved to date making it difficult to understand the difference between the two proteins with regard to the active site. Recently in 2015, structure of L-TTD was predicted by David Baker and coworkers along with 58 other large protein families in prokaryotes for which three-dimensional structural information was not available (Ovchinnikov et al., 2015) (**Fig. 2.13A**). The predicted structure was of an apoprotein, lacking bound Fe-S cluster/substrate and superposed well over LmFH structure with a rmsd of 5.2 Å. To gain insight into the difference in substrate specificity between the two enzymes, the active sites were examined after superposition of the two structures. It was observed that the active site residues in the predicted L-TTD structure are situated in loops making it hard to predict their location in space accurately. The residues corresponding to catalytic acid and the base, Asp73 α and Thr78 β , respectively were in hydrogen bonding distance from each other in the predicted L-TTD structure, making it hard to consider the structure with any confidence (**Fig. 2.13B**). L-TTD α and L-TTD β subunit structures have also been predicted separately and deposited in the AlphaFold database. These structures independently superposed well over LmFH structure with a rmsd of 1.99 Å and 1.57 Å for the α - and β -subunits, respectively (**Fig. 2.13C, D**). The functional form of the enzyme is a tetramer and the same was generated by superposing the L-TTD α and L-TTD β subunits over the LmFH structure (**Fig. 2.13E**). Superposing the AlphaFold predicted L-TTD structure over LmFH structure to observe the difference in active sites between the two proteins, we see that the three cysteine residues that ligate the Fe-S cluster in the two proteins superpose very well (**Fig. 2.13F**). Leaving apart a couple of arginine residues (R33 and R82) and a lysine residue (K102), all other residues that are in hydrogen bonding distance with the substrate superposed well between the two structures (**Fig. 2.13G**); not yielding any information on major differences between the two proteins. Docking L-tartrate and meso-tartrate, which are inhibitors of class-I FH into the LmFH active site, it was observed that the C2 hydroxyl group of L-tartrate points away from the Fe-S cluster. It must be

noted that the orientation of the carboxylate groups was retained as in the bound L-malate in the structure of LmFH-malate complex (**Fig. 2.14A, C**). Docking D-tartrate which is a substrate of class-I FH however, showed that the C2 hydroxyl group of D-tartrate came in hydrogen bonding distance with the Fe-S cluster and the catalytic acid Asp135, in a manner similar to that of L-malate in LmFH malate structure (**Fig. 2.14B**).

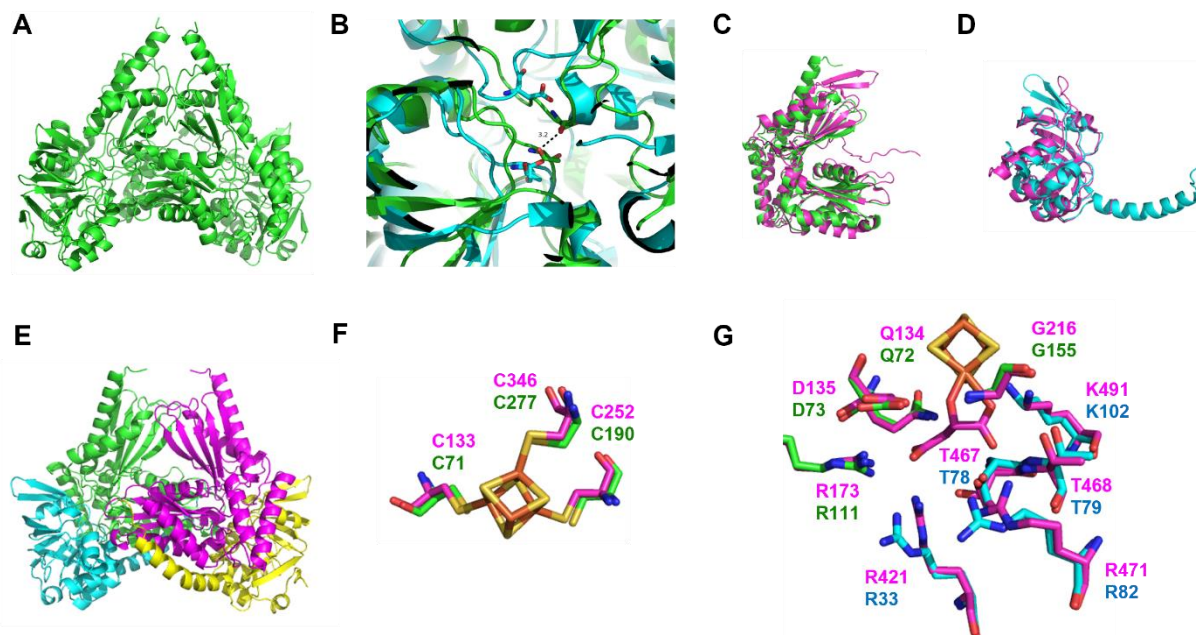


Fig. 2.13. Comparison of class-I FH with modelled structures of L-TTD to gain insights into difference in substrate specificity between the two enzymes. **A.** Predicted structure of L-TTD from the Rosetta structure prediction program utilizing residue-residue co-evolution information (Ovchinnikov et al., 2015). **B.** Superposition of L-TTD predicted structure (green) over LmFH structure (5l2r, blue) to understand differences at the active site of the two enzymes. The active site residues are in loops and the catalytic base threonine is in hydrogen bonding distance from the catalytic acid aspartate in L-TTD structure. Thus, the modelled structure could not be considered with confidence. **C.** Superposition of AlphaFold predicted L-TTD α structure highlighted in green over LmFH structure highlighted in magenta. The two structures align well with a rmsd of 1.99 Å. **D.** Superposition of AlphaFold predicted L-TTD β structure highlighted in cyan over LmFH structure highlighted in magenta. The two structures align well with a rmsd of 1.57 Å. **E.** L-TTD protein structure generated from AlphaFold predicted L-TTD α and L-TTD β subunit structures. **F.** Superposition of cysteine residues (C71, C190 and C277) that ligate Fe-S cluster from L-TTD predicted structure over cysteine residues (C133, C252 and C346) from LmFH structure. Fe-S cluster from LmFH structure is retained in the superposed snapshot. **G.** Superposition of all the active site residues that hydrogen bond with the substrate from LmFH structure with predicted L-TTD structure obtained from AlphaFold. LmFH residues are shown in magenta while L-TTD residues from α -subunit are in green and ones from β -subunit in cyan. Sidechains from a couple of arginine residues (R33 and R82) and a lysine (K102) residue in L-TTD point away from the substrate while most other residues superpose well.

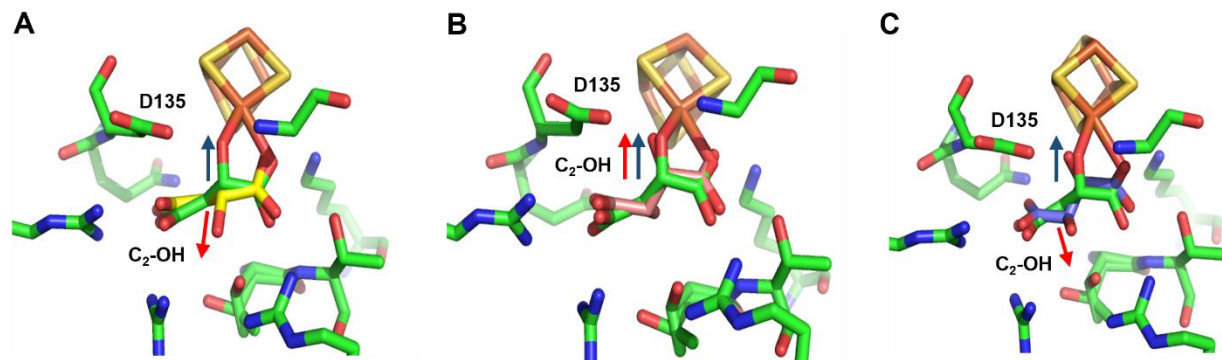


Fig. 2.14. *L-tartrate, D-tartrate and meso-tartrate docked active site in the structure of LmFH. A.* C2 hydroxyl group of *L-tartrate* (yellow) points away from Fe-S cluster and in the opposite direction compared to the bound *L-malate* (green) in the structure of the complex. **B.** *D-tartrate* (pink) superposes well with the bound *L-malate* (green) from LmFH structure with the C2 hydroxyl of both substrates coming in hydrogen bonding distance from both Fe-S cluster and catalytic acid D135. **C.** C2 hydroxyl group of *meso-tartrate* (blue) points in the opposite direction to the Fe-S cluster compared to the bound *L-malate* (green).

2.4. Summary

L-TTD and class-I FH are highly similar proteins with the both of them sharing 45% similarity. The two enzymes are in fact similar to the extent to which they have been misannotated in NCBI, RCSB and UniProt database. The RCSB structure database in which the PDB structure of MjFH β subunit has been deposited (5dni), shows the molecular component of the protein to be a putative L-tartrate dehydratase subunit beta. While, on the UniProt protein database under the protein putative L(+)-tartrate dehydratase subunit beta, MjFH β subunit structure shows up as an entry. Class-I fumarate hydratase has been shown to catalyze the conversion of D-tartrate to oxaloacetate (Kronen & Berg, 2015; van Vugt-Lussenburg et al., 2009, 2013), thus making the two enzymes, class-I FH and L-TTD, enzymes with specificity for substrates of opposite stereospecificity. The way that nature deals with stereoisomers in most circumstances is to design separate enzymes with different active sites to catalyze reactions on opposite stereoisomers (Lamzin et al., 1995). Dehydrogenases, glycosyl hydrolases and amino oxidases are a few examples where, we have two enzymes with different folds catalyzing reactions on opposite stereoisomers with in fact, some following a similar reaction mechanism (Lamzin et al., 1995). Lipases are the only exceptions in nature where, the same enzyme can process opposite enantiomers although, having a clear preference for one over the other (Cygler et al., 1994). To the best of our knowledge, class-I FH and L-TTD are the first examples of two enzymes having an active site possessing identical residues, yet catalyzing reactions on opposite stereoisomers. To understand what distinguishes the two proteins apart if not the active site, multiple sequence alignment was examined. A highly conserved motif 'KGXGS' in class-I FH was found to be present as an 'AGGGC' motif in L-TTD. This is the first report where a motif that distinguishes class-I FH from its highly identical counterpart L-TTD has been identified. To understand how L-TTD is distributed in comparison to class-I single-subunit and two-subunit FH and identify any evolutionary relationship, a phylogenetic tree was constructed. From the phylogenetic tree, it was observed that all class-I FH had the 'KGXGS' motif and the L-TTD had the 'AGGGC' motif. Sequences annotated as L-TTD or as FH/L-TTD but grouping with characterized FH also had the 'KGXGS' motif confirming the observation that indeed the motif is a distinguishing feature between the two enzymes. Phylogenetic tree clearly reveals that the two-subunit FH and L-TTD share a common ancestral origin suggesting a possible horizontal gene transfer (HGT) event during evolution. Single subunit Bacilli FH could have been an intermediate in the process of HGT from archaeal FH to bacterial

L-TTD as they have the motif ‘KGGGC’ that has residues conserved from both motifs and also share a very high identity with L-TTD in spite of being a class I FH. Overall the phylogenetic tree has confirmed previous observations with higher confidence (Kronen & Berg, 2015) and the identification of an exclusive motif present in class-I FH and L-TTD will now help identify these proteins apart and help solve the issue of misannotation between these two groups. Swapping the ‘KGXGS’ motif from MjFH onto *E. coli* L-TTD in the hope of switching enzyme specificity led to complete absence of expressed protein going to show that the motif may also have other structural roles. Therefore, residues from L-TTD motif and residues flanking the motif were replaced onto MjFH, considering the higher stability of the thermostable archaeal enzyme. Mutating the residues and altering the motif in MjFH did not switch the substrate specificity and only lead to decreased enzyme activity which could be a consequence of the proximity of the residues in the motif to the Fe-S cluster and substrate. Further, the mutant enzymes did not develop an ability to bind L-tartrate, as could be seen from the lack of inhibition of the activity on L-malate. A lack of switch in stereospecificity suggests that the motif may not be the sole specificity conferring element in the two enzymes. Computational approach utilizing decrease in free energy of binding of L-tartrate at the active site of LmFH structure (5l2r) as a parameter to design mutants was carried out using the computational program IPRO. A series of 9 mutant designs were made in MjFH drawing information from the energy guided mutant design obtained from LmFH structure using IPRO. None of the mutants turned out to have activity on L-tartrate. Since some of the residues mutated were substrate coordinating residues shown to be critical for enzyme catalysis (details in chapter 1) on fumarate/ L-malate, they also did not retain activity on L-malate and so its ability to bind L-tartrate could not be tested.

L-TTD has been kinetically characterized for the first time and the catalytic efficiency of the enzyme on L-tartrate was found to be 10 fold higher than that of class-I FH on D-tartrate (van Vugt-Lussenburg et al., 2013) going to show that L-TTD is a kinetically superior enzyme. Substrate screening has revealed the enigmatic nature of specificity in these classes of enzymes. L-TTD and class-I FH were found to have absolutely opposite substrate specificity. L-malate and D-tartrate are substrates for class-I FH while, D-malate and L-tartrate are substrate for L-TTD. Fumarate which is planar and achiral was found to be a substrate for both enzymes albeit with L-TTD exhibiting lower activity. Meso-tartrate acted as a competitive inhibitor, inhibiting both class

of enzymes. Surprising difference between these enzymes was that L-malate inhibited L-TTD but L-tartrate turned out to have no affinity for MjFH even at a concentration of 1 mM. L-tartrate however inhibited *E. coli* class-I FH going to show that there are subtle differences between the single-subunit and two-subunit class-I FH. It is intriguing that single-subunit, two-subunit class-I FH and L-TTD have active sites with identical residues yet behave differently. This hints at the fact that subtle differences in enzyme active site architecture accumulated during evolution must have contributed to enzyme specificity in this class of Fe-S cluster hydrolyases.

Structures of LmFH holoprotein and MjFH apo-protein are the only available class-I FH structures while, L-TTD structure has not been solved to date. In order to understand the difference in the active site between these proteins, modelled L-TTD structures were examined using LmFH holoprotein structure as a reference. L-TTD structure modelled by David Baker and coworkers using evolutionary information was remarkably close to the class-I FH superposing with the LmFH structure with a rmsd of 5.8 Å. Although the two structures were similar with regard to the tertiary fold, the protein was modelled in its apo-protein state. Moreover, the active site residues are located in loops making it hard to consider the modelled structure with confidence. Much information with regard to active site architecture could not be deciphered using the modelled L-TTD structure. Modelled α - and β -subunit structures of L-TTD are also available in the AlphaFold database, and this was used to assemble the functional tetramer. Observing the active site, it was evident that AlphaFold was biased by the deposition of LmFH structure in the database, as most active site residues along with the cysteine residues that ligate the Fe-S cluster superposed exceptionally well with LmFH structure (5l2r). The AlphaFold structures as well did not provide any leads on the differences in active site architecture in these class of enzymes that play a role in enantiomer selectivity.

Classically, enzyme stereospecificity has been studied with regard to attachment points of the substrate to the enzyme (Easson & Stedman, 1933; Ogston, 1948). According to the three point attachment model, for an enantiomer to be bound at the enzyme active site, it has to be held in place by active site residues anchoring the substrate at three separate points called attachment points. It has been proposed that the number of attachment points increases with the increase in chiral centers on the substrate (Abrol & Sundaresan, 2002). For such substrates, some enzymes follow the four-

location model. For many substrates with one chiral group, the three-point attachment model still continues to be well accepted (Abrol & Sundaresan, 2002). According to the three-point attachment model, for a molecule with a single chiral center to be bound at the active site, three attachment points of the enzyme, A, B and C has to be anchored to three points on the substrate, A', B' and C'. One of the points of contact breaks for the opposite enantiomer, thus disallowing proper binding or catalysis to take place. For the opposite enantiomer to bind at the active site of the protein, one of the points of contact has to be located on the opposite side in the protein (**Fig. 2.15**). This is the exact reason why in nature we see enzymes with different folds carrying out catalysis on opposite enantiomers (Lamzin et al., 1995). In case of class-I FH, the substrate carboxylate binding residues, Fe-S cluster, and the catalytic aspartate residue that binds both substrate hydroxyl and carboxyl group are some of the points of attachment. Docking D-tartrate which is a substrate for class-I FH at the active site of LmFH, we see that the C2 hydroxyl group comes in hydrogen bonding distance with the Fe-S cluster and the catalytic aspartate residue (**2.12B**). But docking L-tartrate results in the C2 hydroxyl group pointing away from the Fe-S cluster and catalytic aspartate residue (**Fig. 2.14A**). In case of L-TTD, one of the attachment points, Fe-S cluster or the catalytic aspartate residue could be potentially oriented on the opposite side compared to LmFH to catalyze L-tartrate dehydratase reaction. Flipping the Fe-S cluster to the opposite side of the enzyme however seems unlikely as the location of the cysteine residues are similar to MjFH. However, for activity on L-tartrate it is imperative that there be an alteration, though not major, in the conformation of the active site residues and this awaits structural confirmation.

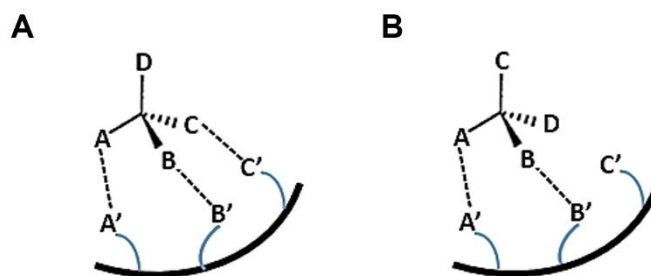


Fig. 2.15. The three-point attachment model A. In the TPA model, three groups, A, B, and C on one enantiomer, simultaneously bind to A', B', and C' on the receptor. (B) Two groups, A and B of the opposite enantiomer, bind to A' and B', but the third group C cannot simultaneously bind to C' (Abrol & Sundaresan, 2002).

2.5. Future directions

Future directions with regard to this part of the study will be to solve the structure of L-TTD to visualize the differences from class-I FH. A snapshot of the active site of the enzyme in complex with substrate/inhibitor will reveal the origin of stereospecificity in these enzymes conclusively. Directed evolution will also help switch substrate specificity in these enzymes and throw novel insights not directly discernible from the structure. Screening for activity/inhibition on/by more dicarboxylic acid substrate analogues will deliver a comprehensive understanding of the enigmatic nature of the active site in these two enzymes.

2.5.1. L-TTD structure solution

Predicted structures of L-TTD by David Baker and AlphaFold have not provided insights into the enigmatic nature of this class of enzymes. In an effort to solve the structure of the enzyme, crystals of L-TTD were obtained in the condition – 1.6 M sodium citrate tri-basic dehydrate, pH 6.5, 25% PEG 3350 (**Fig. 2.16A**). Crystals also turned out to diffract to a resolution of close to 5 Å (**Fig. 2.16B**). Unfortunately, these crystals did not form reproducibly and further standardization in the crystal tray setup along with grid screening has to be carried out to get good quality crystals for diffraction.

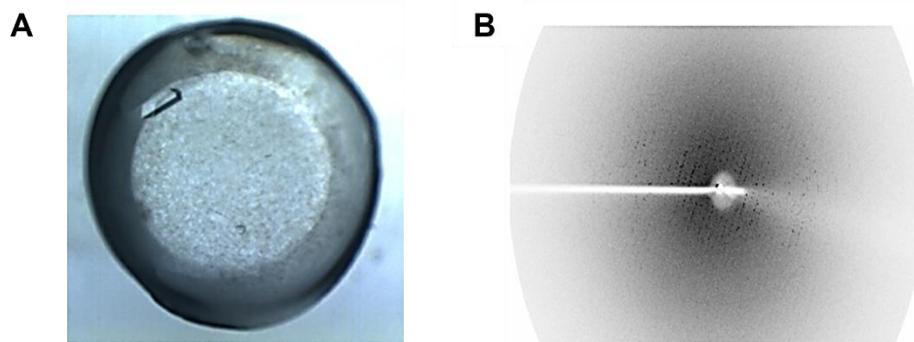


Fig. 2.16. Attempts at L-TTD structure solution. **A.** Crystal of L-TTD obtained from the condition 1.6 M sodium citrate tri-basic dehydrate, pH 6.5, 25% PEG 3350. **B.** Crystals obtained of L-TTD, diffracted only to a best resolution of 5Å with faint diffraction spots disallowing data collection.

2.5.2. Directed evolution to switch substrate specificity

E. coli has three genes that code for FH; *fumA* and *fumB* of class-I type and *fumC* of class-II type. A triple knock out strain of *E. coli* Δ *fumACB* was generated and confirmed in our laboratory earlier (Jayaraman et al., 2018). L-Tartrate dehydratase genes (*orfz1* coding for TTD α and *orfz2* coding

for TTD β) are located in tandem and driven by a common promoter. The two genes were knocked out together in both WT *E. coli* and Δ fumACB strains. Deletion of the genes was validated by PCR using genomic DNA of the mutant strain as a template with appropriate primers (**Fig. 2.17A, B and C**). Genotyping results are provided in **Fig. 2.17D**.

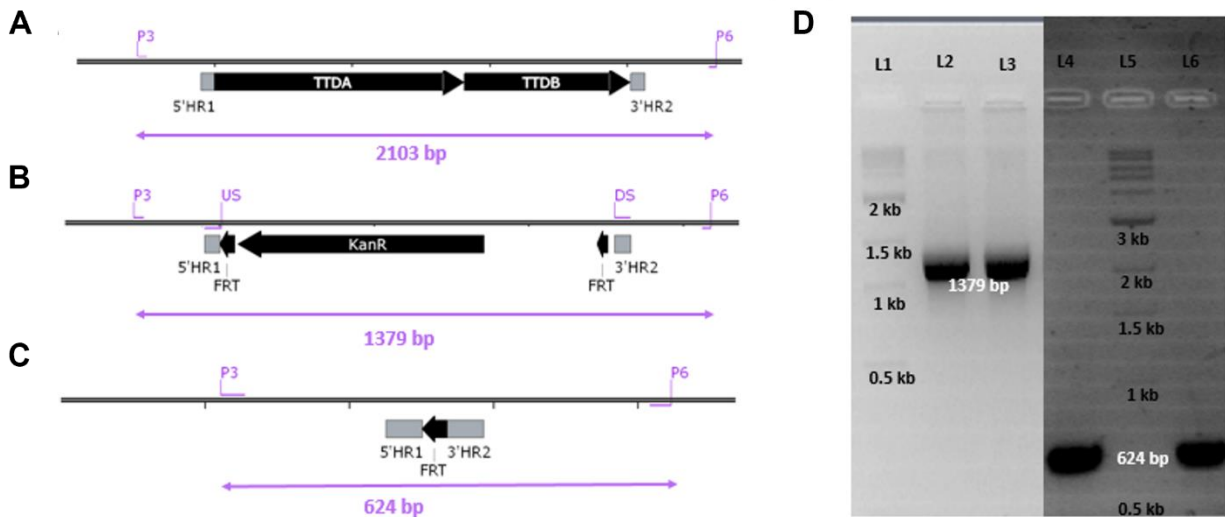


Fig. 2.17. Genotyping of *E. coli* Δ L-TTD and Δ fumACB/ Δ L-TTD strains. **A.** Relative positions of TTD α and TTD β genes in *E. coli* and the location of primers P3 and P6 used for validation of knockout. 5'HR1 and 3'HR2 represent the 30bp homologous regions used for gene replacement with kanamycin resistant marker. **B.** L-TTD $\alpha\beta$ gene loci after homologous recombination and replacement of kanamycin resistance marker flanked by FRT (Flippase recognition target) sites. Primers US and DS were used for the amplification of kanamycin cassette. **C.** TTD $\alpha\beta$ gene loci after FRT recombination and flipping out the kanamycin marker cassette. **D.** PCR amplified products using primers P3 and P6 from genomic DNA of Δ L-TTD and Δ fumACB/ Δ L-TTD strain. L2 and L3, PCR amplicon from P3 and P6 primers in *E. coli* WT strain and Δ fumACB with L-TTD loci replaced with kanamycin resistance marker; L4 and L6, PCR amplicon from P3 and P6 primers in *E. coli* WT strain and Δ fumACB with L-TTD loci knocked out; DNA ladder was loaded in L1 and L5.

Knockout constructs were validated by growing cells on minimal medium with different substrates as sole carbon sources. As expected, the L-TTD KO strain grew on oxaloacetate and Δ fumACB/ Δ L-TTD strain was able to grow on malate and neither strains were able to grow on fumarate (**Fig. 2.18**). To our surprise, the Δ L-TTD strain was also able to grow on L-tartrate going to show that an alternate enzyme capable of using the carbon source is present in *E. coli*. Earlier work carried out in our laboratory has shown that *Plasmodium falciparum* FH was able to complement Δ fumACB growth on L-tartrate (Jayaraman et al., 2018). However, *in vitro* activity testing revealed that the PfFH was inactive on L-tartrate and the cause of the complementation was suggested to be because of the enzyme playing a secondary but critical role required for cell growth. Δ fumACB/ Δ L-TTD strain, which does not grow on L-tartrate medium can be used for

screening error prone PCR driven MjFH or *E. coli FumA* mutants for L-tartrate dehydratase activity.

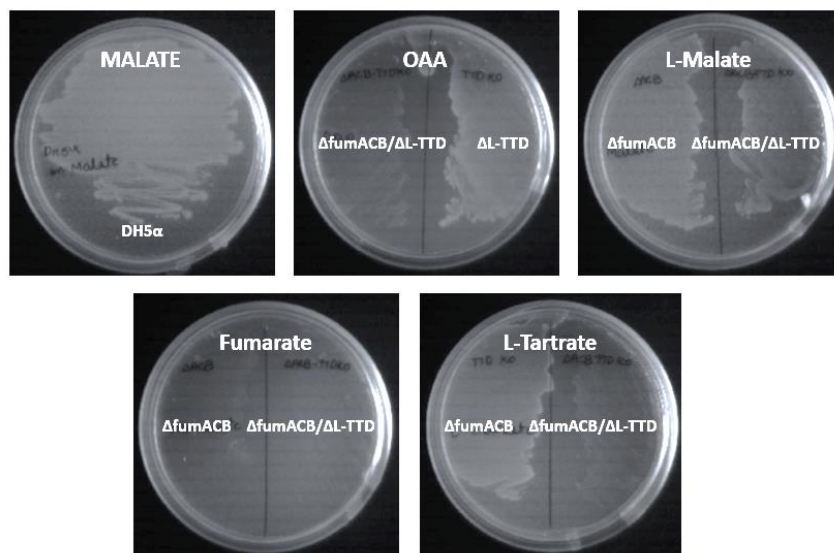


Fig. 2.18. Phenotyping of *E. coli* Δ L-TTD and Δ fumACB/ Δ L-TTD strain. Growth phenotype of Δ L-TTD and Δ fumACB/ Δ L-TTD strain on various substrates.

2.5.3. Wider substrate screen for better understanding of enzyme active site

This is the first report showing three enzyme classes (class-I single-subunit FH, two-subunit FH and L-TTD) having active sites residues that are identical, behaving differently pertaining substrates and inhibitors. More information about how these enzymes are different can be garnered through screening a larger subset of chiral substrates and inhibitors, as has been done in the case of class-II FH (**Table. 2.3**). From the information derived on the inactive substrate analogues and inhibitors, a comprehensive understanding of how similar or different the active site in these enzymes is, can be deciphered.

Table. 2.3. List of substrates, inactive analogues and inhibitors for class-I, class-II FH and L-TTD. Highlighted in blue are the inactive analogues of class-II FH which have not been tested for activity in either class-I FH or L-TTD and red are the inhibitors of class-II FH which have not been tested for inhibition in either class-I FH or L-TTD.

	Class I FH	Class II FH	L-TTD
Substrates	Fumarate L-malate D-tartrate Mesaconate Acetylene dicarboxylate Fluoro-fumarate	Fumarate L-malate D-tartrate Acetylene dicarboxylate Fluoro-fumarate	Fumarate D-malate L-tartrate
Inactive analogues	L-tartrate (two subunit FH) D-malate Maleate Succinate (+) citramalate	Dimethylfumarate Aspartate Crotonate Acetate Acetoacetate Butyrate	D-tartrate
Inhibitors	Thiomalate Meso-tartrate L-tartrate (single subunit FH) Cis-aconitate Trans-aconitate Citraconate Phenyl lactate trans-Glutaconate (-) citramalate Citrate L-Isocitrate Malonate	L-tartrate Meso-tartrate Maleate Mesaconate Succinate Trans-aconitate Citrate L-Isocitrate Malonate L-2-hydroxy-3-sulphopropionate Trans-glutaconate Isopropylmalate Pyruvate Alpha-ketoglutarate Adipate Glycine glutarate	L-malate Meso-tartrate
Reference	<i>Dennis H. Flint (1994), Archives of biochemistry and biophysics, 311, 409-516</i> <i>This study</i>	<i>Enzymes, Dixon and Webb (2014)</i> <i>John W. Teipel, et al., (1968) The journal of biological chemistry, 243, 2684-2694</i>	<i>This study</i>

2.5.4. Interface analysis to understand the tight interaction between subunits of class-I FH and L-TTD

Large majority of the residues at the interface of class-I FH are conserved as has been shown in a previous report (Feliciano et al., 2016). Aligning L-TTD with class-I FH sequences, it was observed that only 10 of the 16 residues at the interface are conserved between class-I FH and L-TTD (**Fig. 2.19**). As majority of the active site residues are located at the interface of the α - and β -subunit (N/C-terminal domains for class-I FH) for the two groups of enzymes, it will be worth examining how differently the two subunits interact to constitute the active site of the enzyme. A Ni-NTA pull down of MjFH α from bacterial cells co-expressing MjFH α and *E. coli fumA* CTD showed that the two proteins interact (**Fig. 2.20**). Since the interface of both class-I FH and L-TTD are well conserved, it will be interesting to see if MjFH α interacts with *E. coli* L-TTD β and the vice versa. If it turns out that there is a cross-interaction between these subunits of the two groups of proteins, activity of the same can be tested on various substrates to obtain insights into substrate specificity in these enzymes.

Methanocaldococcus_janschii_FH	1	..MKTS	SDVVVE	FR	EA	AI	YD	PE	OV	KN	AL	EA	YK	ES	EIS	KNT	LKA	TE	NN	KI	AE	ET	QV	PL	CD	TG	VP	AI																																						
<i>E. coli</i> FumA	43	APEA	LTLL	AR	AF	HD	AS	FL	RP	AH	QQ	VAD	IL	RD	PE	AS	EN	KY	VA	LQ	LR	NS	DI	AA	RG	VLP	TC	CD	TG	TA	AI																																			
Leishmania_major_FH	71	NPEA	LTLL	AR	AF	SD	VH	HF	FR	KD	H	LE	GW	RR	AI	ED	PE	AS	DN	RY	VAT	LL	KN	AC	IA	AG	RV	LP	TC	CD	TG	TA	AI																																	
Shigella_sonnei_FumA	43	APEA	LTLL	AR	AF	HD	AS	FL	RP	AH	QQ	VAD	IL	RD	PE	AS	EN	KY	VA	LQ	LR	NS	DI	AA	RG	VLP	TC	CD	TG	TA	AI																																			
Pseudomonas_putida_FumA	5	KQDD	LI	QV	AD	AL	QFI	SY	HP	V	DF	I	Q	A	M	H	E	A	Y	L	R	E	S	P	A	A	R	D	S	I	A	Q	I	L	I	N	S	R	M	C	A	T	G	H	R	P	I	C	CD	TG	I	V	T													
Mpob	41	DPEA	L	L	A	L	A	N	T	A	M	R	D	V	S	F	M	L	R	P	A	H	L	E	K	V	A	A	I	L	S	D	P	E	S	A	N	D	R	G	V	A	V	A	L	L	R	N	A	E	V	A	A	R	G	V	L	P	I	C	CD	TG	T	A	I	
Burkholderia_xenovorans_FH	18	KQED	LI	Q	S	I	A	D	S	L	Q	I	S	Y	H	P	L	D	Y	I	Q	A	L	G	R	A	Y	E	L	E	Q	S	P	A	A	K	D	A	I	A	Q	I	L	T	N	S	R	M	C	A	E	G	K	R	P	I	C	CD	TG	I	V	T				
Pyrococcus_furiosus_FH	1	..MI	K	V	E	D	I	V	E	A	L	K	L	A	V	N	L	D	V	A	L	A	L	K	N	A	Y	E	R	E	D	V	E	I	A	K	Y	N	L	E	M	L	R	A	V	K	T	S	K	E	K	V	L	P	I	C	CD	TG	T	P	I					
Salmonella_typhimurium_FH	6	NTEI	I	I	P	I	E	K	L	V	K	K	A	C	E	L	D	D	N	L	M	C	S	F	R	K	A	Y	D	K	E	S	K	I	G	K	E	T	I	K	L	I	D	N	G	E	A	K	K	E	Q	L	A	C	H	D	F	G	T	G	I					
Helicobacter_hepaticus_FH	5	RVED	K	S	A	V	A	K	L	I	A	N	A	C	T	I	P	T	K	R	A	S	A	K	S	S	S	A	L	Q	N	L	D	L	L	E	M	G	K	I	A	E	S	N	M	F	L	C	D	D	G	M	T	V												
Archaeoglobus_fulgidus_FH	2	EFD	V	V	N	A	V	D	A	L	K	R	A	H	E	L	P	D	V	L	E	A	K	R	A	Y	G	E	E	T	A	R	K	N	L	E	A	I	E	A	A	K	L	G	V	E	S	C	CD	TG	I	V	T													
Pelotomaculum_thermopropionicum_FH	7	NTS	V	I	T	N	V	A	R	L	C	Q	A	N	L	G	P	D	V	M	E	A	F	K	K	A	Q	E	E	I	S	L	T	G	K	G	L	Q	D	L	I	V	N	A	E	I	A	C	E	E	Q	A	F	M	C	CD	TG	I	V	T						
<i>E. coli</i> TTD	10	AVN	K	L	T	E	I	V	A	N	F	T	A	M	I	S	T	R	M	P	D	V	V	D	K	L	K	L	K	D	A	E	T	S	S	M	G	K	I	I	Y	H	T	M	F	D	N	M	Q	A	I	D	L	N	R	P	A	C	CD	TG	E	I	M			
Shigella_sonnei_TTD	10	AVN	K	L	T	E	I	V	A	N	F	T	A	M	I	S	T	R	M	P	D	V	V	D	K	L	K	L	K	D	A	E	T	S	S	M	G	K	I	I	Y	H	T	M	F	D	N	M	Q	A	I	D	L	N	R	P	A	C	CD	TG	E	I	M			
Pseudomonas_putida_TTD	9	AV	E	S	L	T	D	V	L	A	K	E	T	A	Y	I	G	K	R	L	P	K	D	V	K	E	K	T	A	L	R	A	E	T	N	P	L	A	I	A	V	Y	D	S	M	A	D	N	O	Y	A	D	K	L	N	R	P	A	C	CD	TG	E	I	M		

Methanocaldococcus_janschii_FH	68	V	F	I	K	I	G	K	N	I	S	S	E	I	M	K	I	E	E	T	K	E	G	V	K	K	A	T	E	E	..	V	P	L	R	P	N	V	V	H	P	L	..	T	R	E	N	F	K	T	N	V	G	L	N	S	P	F	T	N	I	E	F	D	E
<i>E. coli</i> FumA	113	I	V	G	K	K	Q	R	V	..	W	T	G	G	..	D	E	A	A	L	A	R	G	V	N	T	Y	I	E	..	D	N	L	R	Y	S	Q	N	A	P	L	D	M	Y	K	E	V	N	T	G	T	N	L	P	A	..	Q	I	D	L	A	V	D		
Leishmania_major_FH	141	V	L	G	K	R	G	E	L	C	..	W	T	G	G	..	D	E	K	Y	L	S	R	G	I	W	N	A	Y	..	H	N	L	R	Y	S	Q	T	A	A	L	D	M	F	K	E	C	N	T	G	D	N	L	P	A	..	Q	I	D	L	L	A	V		
Shigella_sonnei_FumA	113	I	V	G	K	K	Q	R	V	..	W	T	G	G	..	D	E	A	A	L	A	R	G	V	N	T	Y	I	E	..	D	N	L	R	Y	S	Q	N	A	P	L	D	M	Y	K	E	V	N	T	G	T	N	L	P	A	..	Q	I	D	L	A	V			
Pseudomonas_putida_FumA	74	V	F	I	R	V	G	M	D	V	R	W	G	A	T	M	S	V	D	D	M	L	I	N	E	G	V	R	R	A	Y	N	L	P	E	N	V	L	A	S	L	A	D	F	..	A	G	A	R	K	N	R	D	N	P	A	..	V	I	H	S	I	V		
Mpob	117	I	V	G	K	K	Q	R	V	..	W	T	G	G	..	D	E	K	Y	L	S	R	G	V	F	K	T	Y	E	..	E	N	L	S	O	T	L	P	L	T	M	Y	E	V	N	S	T	N	D	N	P	A	..	Q	I	D	L	A	V						
Burkholderia_xenovorans_FH	87	V	F	I	R	V	G	M	D	V	R	W	G	A	T	M	S	V	D	D	M	L	I	N	E	G	V	R	R	A	Y	N	L	P	E	N	V	L	A	S	L	A	D	F	..	A	G	A	R	K	N	R	D	N	P	A	..	V	I	H	S	I			
Pyrococcus_furiosus_FH	68	F	V	E	T	D	G	T	W	N	V	E	I	..	Y	E	T	I	A	E	A	V	K	R	A	T	E	..	I	P	L	R	P	N	A	I	C	L	L	..	S	G	K	V	V	G	..	N	V	P	E	I	H	E	P	..									
Salmonella_typhimurium_FH	75	V	I	M	E	I	G	H	V	C	W	E	G	K	P	..	L	K	D	O	N	O	G	V	R	O	G	Y	E	N	..	G	Y	L	R	K	S	M	V	A	D	P	..	L	E	R	N	T	N	D	N	P	A	..	L	H	T	I	V						
Helicobacter_hepaticus_FH	74	V	F	V	E	I	Q	D	V	H	I	T	G	Y	..	L	E	D	A	I	N	E	G	I	K	E	G	Y	T	..	G	Y	L	R	K	S	V	V	E	E	P	..	L	Y	E	R	K	N	T	N	S	P	A	..	V	I	H	T	R	I					
Archaeoglobus_fulgidus_FH	71	F	I	L	E	V	E	G	R	E	L	C	L	D	..	F	D	L	E	A	I	N	E	G	V	R	K	A	T	E	..	I	P	L	R	P	N	A	V	H	P	I	..	T	R	E	N	S	G	D	N	T	G	L	G	I	P	I	N	E	L	V			
Pelotomaculum_thermopropionicum_FH	76	I	F	L	E	L	Q	D	V	R	I	E	G	D	..	L	Y	E	A	I	N	E	G	V	R	K	Y	T	E	..	G	Y	L	R	K	S	I	V	G	H	P	..	L	E	R	K	N	T	G	D	N	P	A	..	V	I	H	T	I						
<i>E. coli</i> TTD	79	F	F	V	K	V	G	S	R	F	L	L	G	..	E	L	Q	S	I	L	K	Q	A	V	E	A	T	V	..	A	P	L	R	H	N	A	V	E	I	F	..	D	E	V	N	T	G	K	N	T	G	S	G	V	P	W	T	I							
Shigella_sonnei_TTD	79	F	F	V	K	V	G	S	R	F	L	L	G	..	E	L	Q	S	I	L	K	Q	A	V	E	A	T	V	..	A	P	L	R	H	N	A	V	E	I	F	..	D	E	V	N	T	G	K	N	T	G	S	G	V	P	W	T	I							
Pseudomonas_putida_TTD	78	Y	F	I	S	A	G	A	R	F	L	L	G	..	E	M	E	G	L	E	N	A	T	K	E	A	T	I	..	G	P	L	R	H	N	A	V	E	T	F	..	I	E	K	N	T	G	N	T	G	S	R	I	P	W	D	E								

Methanocaldococcus_janschii_FH	132	S	L	D	R	E	I	E	I	A	F	P	K	G	A	G	S	E	N	M	S	A	I	K	M	L	P	S	D	G	I	E	G	I	K	N	F	V	L	E	T	J	A	N	A	G	K	P	C	P	I	V	V	G	I	G	G	T	A
<i>E. coli</i> FumA	175	..	G	D	E	Y	K	F	L	C	I	A	K	G	G	S	A	N	K	T	Y	L	Y	Q	E	T	K	A	L	L	T	P	G																										

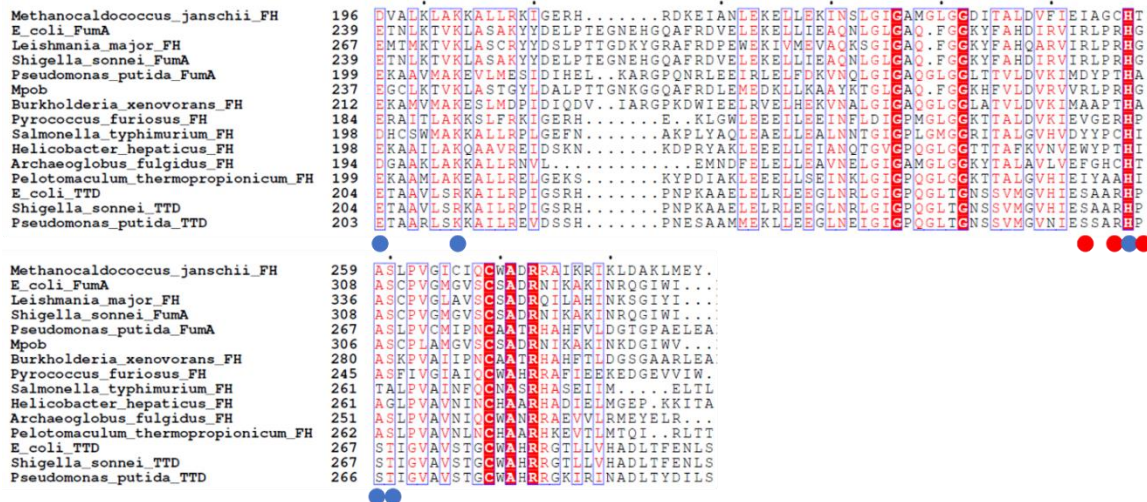


Fig. 2.19. Multiple sequence alignment of biochemically characterized class I FH and L-TTD. Interface residues that are conserved across the two enzymes are highlighted with blue circles while the ones that aren't conserved are highlighted with red circles beneath the respective residue.

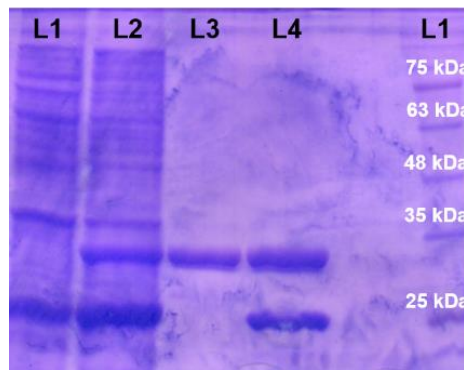


Fig. 2.20. Pull down of *E. coli* fumA CTD with MjFHa subunit. L1, fumA CTD expressing (DE3)-RIL cells lysate; L2, 250 μ M imidazole elute from lysate expressing fumA CTD after incubation with MjFHa subunit and Ni-NTA beads; L3 and L4, MjFHa and MjFHo β purified protein as control; L5, protein molecular weight marker.



(This page is intentionally left blank)



Chapter 3

Understanding structural and functional modulation of IMPDH from the archaeon *Methanocaldococcus jannaschii*

Although the mechanism of catalysis by IMPDH is fairly conserved across prokaryotes and eukaryotes, there are differences with regard to modulation of activity. IMPDH consists of a core catalytic TIM domain and a regulatory CBS domain. Understanding of enzyme modulation by nucleotide modulators is not complete in IMPDH and the same has not been probed into in archaeal counterparts. Previous studies from the laboratory have revealed some interesting modulatory aspects of the archaeal enzyme from *M. jannaschii* where a potassium ion is not required for optimal activity and the enzyme gets modulated by all nucleotide modulators except ATP. In order to understand the modulation better, structure solution of MjIMPDH has been attempted and efforts at obtaining the IMPDH structure is presented in this chapter. Transmission electron microscopy imaging revealed that MjIMPDH forms cytoophidia like filaments, which have been observed in eukaryotes but not prokaryotes going to show that the archaeal enzyme is structurally similar to eukaryotic rather than prokaryotic orthologs.

3.1 Introduction

3.1.1. IMPDH is a key player for nucleotide regulation within cells

Optimal nucleotide levels are maintained within cells across varied metabolic states, mediated by the operation of complex regulatory networks. Nucleotides are ubiquitous and essential biological molecules that play a central role in cellular physiology by participating in processes as an energy source, building blocks, cofactors and effector molecules (Lane & Fan, 2015). To avoid misincorporation of dNTPs into DNA, which could lead to mutations, a properly balanced biosynthesis of the same mediated by a cascade of reactions through enzymes that are precisely regulated take place (Labesse et al., 2013). Living cells appear to take advantage of subcellular localization and more recently discovered, supramolecular organization as a mode of regulation through formation of purinosomes and its dynamic assembly/disassembly for optimal nucleotide biosynthesis (An et al., 2008; Narayanaswamy et al., 2009; Noree et al., 2010). The complex is composed of six enzymes responsible for the biosynthesis of IMP from phosphoribosyl pyrophosphate. Recently, allosteric regulation has been shown to have an important role in this context and has been considered as a unique player in IMP biosynthesis. While the purinosome complex ensures optimal substrate channeling to deliver IMP to the cells, IMPDH is regulated at multiple levels including modulation by allostery. IMPDH controls the de novo biosynthesis of guanosine nucleotides, the inhibition of which causes an imbalance between adenine and guanine nucleotides with large repercussions like cytotoxicity (Hedstrom, 2009a; Shu & Nair, 2008). Both adenine and guanine nucleotides allosterically modulate IMPDH activity by altering either the enzyme activity or oligomeric state, or both. Physiological processes within the cells and their proliferation are tightly linked to this well controlled allosteric regulation of IMPDH by purine nucleotides which in turn maintains the adenine to guanine nucleotide balance within the cells. This makes IMPDH an important target for immunosuppressive, antiviral and cancer chemotherapy (Hedstrom, 2009).

3.1.2. General features of the enzyme IMPDH

IMPDH (EC 1.1.1.205) catalyzes the oxidation of inosine 5'-monophosphate (IMP) to xanthosine 5'-monophosphate (XMP) concomitantly reducing NAD^+ to NADH. IMPDH is ubiquitously distributed across all organisms known to date with the exception of two protozoan parasites, *Giardia lamblia* and *Trichomonas vaginalis* (Carlton et al., 2007; Morrison et al., 2007). IMPDHs

exist as tetramers or octamers with each monomer consisting of 400-500 amino acid residues carrying a core catalytic domain and a regulatory domain (Labesse et al., 2013) (**Fig. 3.1B, C**). The regulatory domain is referred to as the Bateman domain or CBS domain, containing two repeating CBS motifs, which were first discovered in human cystathione beta-synthase (Bateman, 1997) (**Fig. 3.1A**). CBS domains are found in several structurally and functionally unrelated proteins and are involved in maintenance of cellular energy status, metal ion concentration or ionic strength, and regulate enzymatic activity (Baykov et al., 2011; Ereño-Orbea et al., 2013). Removal of CBS domain in IMPDH has been found to have no effect either on activity *in vitro* or perturbation of subunit organization (Nimmegern et al., 1999; Sintchak et al., 1996). The CBS domain is found to be disordered in many of the structures deposited in PDB with the removal of the domain also reported to facilitate crystallization (Hedstrom, 2009). All reported IMPDH to date are found to require a monovalent cation (preferably K^+) for optimal activity with an increase in activity to around 100-folds (Alexandre et al., 2015). The binding site for the monovalent cation comprises the catalytic loop and the C-terminal segment, which are again highly disordered segments in most X-ray structures.

The core catalytic domain is composed of a TIM barrel in IMPDH which is again sub-divided into five key structural motifs that carry out substrate binding and catalysis (**Fig. 3.1A**). “Catalytic cysteine loop” carries the active site cysteine which is strictly conserved across all IMPDH. Adjacent to the active site loop is the phosphate binding region also called the “phosphate binding loop” that is composed of six residues. A twisted β -sheet that projects out of the carboxy-terminal of TIM barrel is referred to as “finger domain”, which is shown to help transmission of allosteric signals from the Bateman domain to the catalytic domain (Buey, Ledesma-Amaro, Velázquez-Campoy, et al., 2015). Around ten residues at the distal end of the finger domain constitute the “mobile flap”, which is dynamic and moves in and out of the active site during catalysis. The open conformation allows the dehydrogenase reaction to occur followed by hydrolysis in closed conformation. Finally, the “C-terminal segment” couples the active site through a monovalent cation. The catalytic cysteine loop, mobile flap, finger loop and about 20 residues at the C-terminal end of IMPDH are found to be highly flexible and thus missing from the X-ray crystal structures of various IMPDHs (Hedstrom, 2009; Morrow et al., 2012; Prosis et al., 2002; Rao et al., 2013). IMP binding site is constituted of residues from the catalytic cysteine loop, phosphate binding

motif, and mobile flap. Residues from the catalytic loop and mobile flap make up NAD^+ binding site. While IMP binding residues are identified to be invariant, large divergence has been observed with regard to NAD^+ binding site (Hedstrom, 2009).

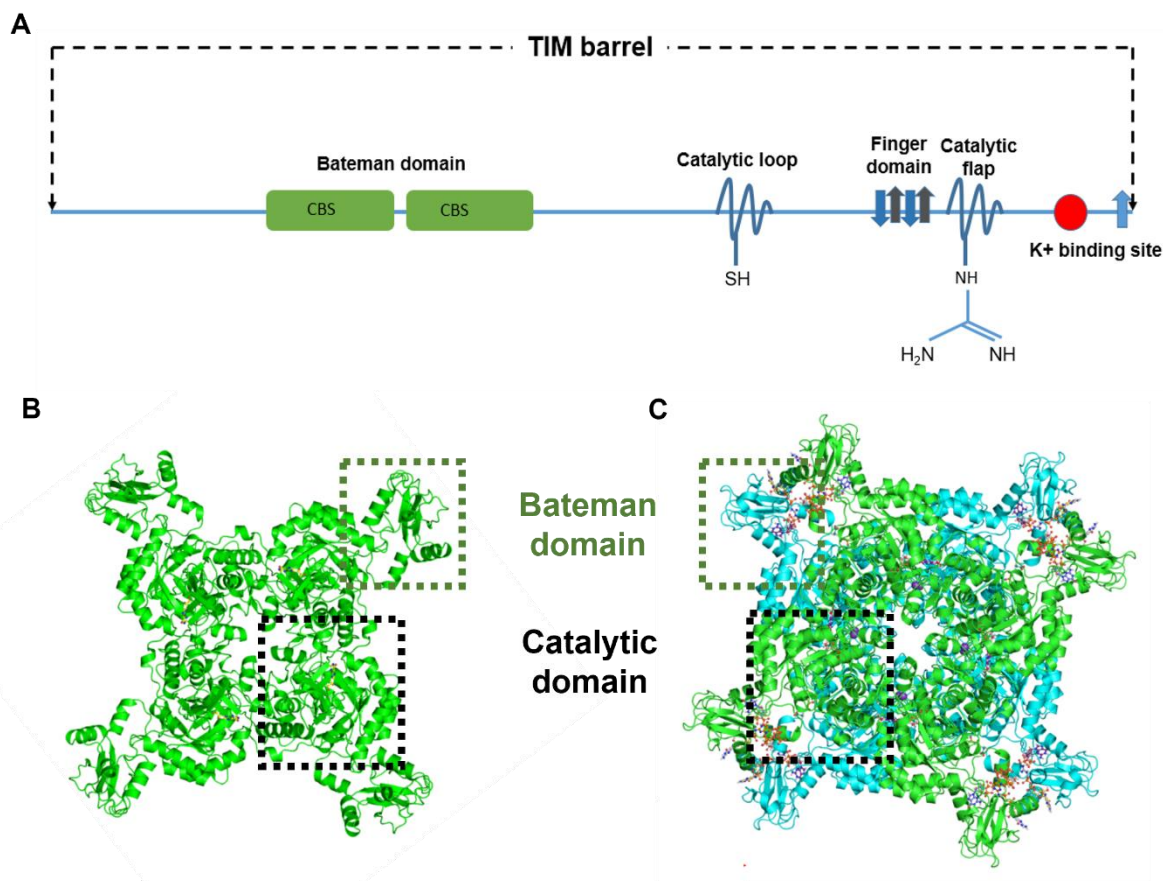


Fig. 3.1. Domain architecture and spatial arrangement of IMPDH. *A.* Schematic bar representation of IMPDH key structural elements and CBS subdomain. *B.* Tetrameric structure of *S. pyogenes* IMPDH (1zfi) shown in ribbon diagram. CBS subdomain and catalytic domain are highlighted within dashed boxes. *C.* Octameric structure of *A. gossypium* IMPDH (4z87) shown in ribbon diagram. CBS subdomain and catalytic domain are highlighted within dashed boxes.

The chemical transformations catalyzed by IMPDH include a dehydrogenase reaction in which the catalytic cysteine attacks the C2 of IMP to form a covalent thiomidate intermediate, E-XMP* with simultaneous hydride transfer onto NAD^+ to form NADH. This reaction is followed by a subsequent hydrolyase reaction on E-XMP* yielding XMP and free enzyme (**Fig. 3.2**). IMPDH assumes an open conformation allowing NAD^+ to bind during dehydrogenase reaction while the mobile flap moves into the cofactor site during the hydrolyase reaction, bringing the conserved Arg-Tyr dyad into the active site to allow the arginine to act as the catalytic base (Hedstrom, 2009).

Presence of K^+ is found to increase the rate of association of NAD^+ with E-IMP complex and accelerate the flap closure (Riera et al., 2011).

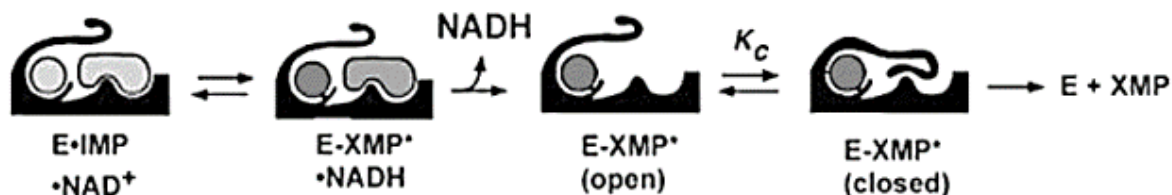


Fig. 3.2. Mechanism of IMPDH reaction. Adapted from Hedstrom 2009.

3.1.3. Understanding IMPDH enzyme modulation in prokaryotes and eukaryotes

Although the reaction rate catalyzed by IMPDH varies slightly between the prokaryotic and eukaryotic IMPDH, both the enzymes have similar structure and mechanism of catalysis. However, differences with regard to allosteric regulation has been observed within and between prokaryotic and eukaryotic IMPDH (Alexandre et al., 2015). IMPDH have been grouped into two classes, class-I and class-II based on the allosteric modulation taking place in these enzymes. Class-I IMPDHs show cooperativity for IMP binding, which is activated upon MgATP binding and the enzyme is octameric in all conditions. Class-II IMPDH on the other hand behave as Michaelis-Menten enzymes and are tetramers in the apo-state and shift to octamers upon MgATP binding (Alexandre et al., 2015). While binding of MgATP to class-I IMPDH directly influences activity, MgATP binding in class-II IMPDH only result in structural changes without any direct influence on activity. Prokaryotic IMPDH is shown to have both class-I and class-II IMPDH, whereas all eukaryotic IMPDHs are shown to be only class-II IMPDH (Buey et al., 2017).

Prokaryotic and Eukaryotic IMPDH also display different modes of inhibition, with prokaryotic IMPDH being inhibited only by GMP in a competitive fashion, while eukaryotic IMPDH are inhibited by GDP and GTP in a mixed fashion by the binding of the same to a non-canonical nucleotide binding site in the CBS domain that is absent in prokaryotic IMPDH. GTP and GDP bind to the non-canonical binding site in eukaryotic IMPDH and induce a tail to tail dimerization of the tetramers leading to a more compacted octameric conformation with compromised activity (Buey et al., 2017). The compact IMPDH structure becomes relaxed on MgATP binding, which displaces the bound GDP/GTP from the non-canonical site on CBS. It is believed that class-I

IMPDHs are kept auto-inhibited *in vitro* in the absence of nucleotides and achieve full activity upon MgATP binding. In contrast, eukaryotic class-II IMPDH are kept inactivated by GDP/GTP binding which is relieved upon ATP binding. With the available information, it is speculated that a conformational switch controlled by ATP was already present in the common ancestor for prokaryotic and eukaryotic IMPDH, which likely got modified with the incorporation of non-canonical nucleotide binding site in eukaryotic IMPDH (Buey et al., 2017).

3.1.4. Cytoophidia and its role in physiology

Adding more complexity to the mechanism of IMPDH regulation, which is already known to associate with purinosomes and also display allostery, the enzyme has also been reported to form micro-metric long structures referred to as cytoophidia (Carcamo et al., 2011; Chang et al., 2015; Ji et al., 2006). IMPDH cytoophidia filaments however are an exclusive feature of the eukaryotes and has not been observed in prokaryotic IMPDH (Burrell et al., 2022). A number of non-cytoskeletal proteins form filaments which are dynamic assemblies polymerizing upon nutrient stress (Alberti et al., 2009; Aughey & Liu, 2016; Liu, 2016; Narayanaswamy et al., 2009; Noree et al., 2010; O'Connell et al., 2012, 2014; Q. J. Shen et al., 2016). This phenotype has been however, disproportionately associated more with metabolic enzymes (O'Connell et al., 2012). From the few characterized metabolic filaments, polymerization is known to regulate enzyme activity and maintain homeostasis with its role not determined in most cases (Barry et al., 2014; Beaty & Lane, 1985; Lynch et al., 2017; Petrovska et al., 2014). IMPDH reversibly assembles into helical polymers in an ATP dependent fashion forming stacked octamers *in vitro* (Anthony et al., 2017; Burrell et al., 2022; Labesse et al., 2013). IMPDH filaments colocalize with the pyrimidine biosynthetic enzyme cytidine triphosphate synthase (CTPS) assemblies in some circumstances suggesting a potential coordination between purine and pyrimidine biosynthesis (Keppeke et al., 2015). The physiological function of IMPDH cytoophidia remarkably remains unclear although its polymerization has been correlated with rapid cell proliferation (Chang et al., 2015; Duong-Ly et al., 2018; Keppeke et al., 2018). IMPDH generally does not polymerize in cell lines grown in rich media which however gets triggered by depleting culture media of essential purine precursors or adding inhibitors such as mycophenolic acid or ribavirin (Calise et al., 2014; Carcamo et al., 2014). Responsiveness of IMPDH to reduced flux through guanine nucleotide biosynthetic

pathway has suggested that the role of the assembly could be to maintain homeostasis by restoring guanine nucleotide pools (Anthony et al., 2017).

3.1.5. Unique regulatory features of an archaeal MjIMPDH

Previous studies carried out at our laboratory have revealed unique regulatory features of an archaeal IMPDH from *Methanocaldococcus jannaschii*. The archeal IMPDH has all adenine and guanine nucleotides inhibiting the enzyme except for ATP, which has no effect on enzyme activity similar to the eukaryotic IMPDHs. MjIMPDH does not require a monovalent cation for optimal activity which is not the case with any other IMPDH reported to date. The fact that MjIMPDH gets inhibited by GDP and GTP with no increase in activity in the presence of ATP particularly suggests that the archeal enzyme has eukaryotic IMPDH like features. Deletion of the CBS domain relived inhibition by purine nucleotides significantly suggesting that a non-canonical nucleotide binding site also exists in MjIMPDH.

3.1.6. Objectives

Considering the unique regulatory features that the archaeal MjIMPDH displays, the objective of the study was to solve the structure of the enzyme to understand its modulatory aspects better and conduct transmission electron microscopy imaging to observe if the protein forms a cytoophidia like filament similar to that observed in eukaryotic IMPDH.

3.2 Experimental procedures

3.2.1. Chemicals, strains, and molecular biology reagents

Media components for growing *E. coli* cultures were from Himedia, Mumbai, India. Akta HPLC and Q-Sepharose and Superdex resins were from GE Healthcare Life Sciences, USA. Crystallization cocktails were from Hampton Research, Aliso Viejo, CA.

3.2.2. Sequence alignment and phylogenetic tree construction

Sequence of IMPDH from various organisms was obtained from NCBI BLAST-P tool using *Methanocaldococcus jannaschii* IMPDH (MjIMPDH) sequence as query. Clustal Omega (Sievers et al., 2011) was used for multiple sequence alignment and output was processed using ESPRIPT (Robert & Gouet, 2014) for visualization. Sequence identity and similarity server (SIAS) was used for obtaining the % similarity and identity between sequences.

3.2.3. Protein expression and purification

MjIMPDH expressing clone in pET21b+ with a C-terminal (His)₆-tag fusion (pET21bIMPDH) and core catalytic domain expressing clone in pET21b+ with a C-terminal (His)₆-tag fusion (pET21bΔCBS) were already generated in the laboratory. BL21(DE3) strain of *E. coli* lacking the endogenous IMPDH enzyme, Δ*guaB*^K previously generated at the lab was co-transformed with pLySs (expressing ROS tRNA) and pET21bIMPDH or pET21bΔCBS. The culture was grown overnight at 37 °C and 1% of inoculum was added to 800 ml of Terrific broth (TB). The cells were grown at 37 °C to an absorbance of 0.6, induced with 0.05 mM IPTG and grown further for 12 hours at 16 °C. The cells were harvested by centrifugation at 6000g for 10 minutes at 4 °C and resuspended in lysis buffer containing 50 mM Tris HCl, pH 7.0, 10% glycerol, 2 mM DTT and 0.1 mM PMSF. Cell lysis was achieved by five cycles of French press (Thermo IEC Inc., USA) at 1000 psi, and the lysate was clarified by centrifugation at 30,000 x g for 30 minutes. Supernatant was heated at 70 °C for 30 minutes to precipitate the bacterial proteins and again clarified by centrifugation at 30,000 x g for 30 minutes. Supernatant was then treated with 0.01% PEI to precipitate nucleic acids and clarified by centrifugation at 30,000 x g for 30 minutes. The clarified lysate was filtered through a 0.44 micron filter and loaded onto Q-sepharose anion-exchange column. The protein was eluted using a linear gradient of KCl in buffer containing 50 mM Tris HCl, pH 7.0, 10 % glycerol and 2 mM DTT. Fractions collected at concentrations of KCl above 300 mM were examined by SDS-PAGE, pooled according to purity, and precipitated using

ammonium sulfate at 90% saturation. Precipitate obtained was resuspended in lysis buffer and subjected to size-exclusion chromatography on Superdex 200 column of dimension 1.6 cm x 60 cm (GE healthcare, USA). The purified protein was analyzed by SDS-PAGE (Laemmli, 1970) and protein concentration was determined by the Bradford method (Bradford, 1976) with bovine serum albumin (BSA) as standard. Expression of MjIMPDPH^{ΔCBS} was carried out in a similar fashion to that of MjIMPDPH using 50mM Tris HCl, pH 8.4 for anion exchange and 50mM Tris HCl, pH 8.0 for size-exclusion chromatography and final storage.

3.2.4 Crystallization, data collection and structure solution

Purified MjIMPDPH and MjIMPDPH^{ΔCBS}, concentrated to ~5 mg mL⁻¹ were set up for crystallization using the Microbatch method (Chayen et al., 1992). A 72 multi-well plate from Grenier-Bio was used. Conditions for crystallization were obtained by using all the conditions in the commercially available crystallization kit from Hampton (Hampton Research, USA). The crystallization droplet contained 3 μl of protein and 3 μl of buffer from different conditions placed under a 50% mix of paraffin and silicone oil at room temperature. Numerous crystals were obtained as observed under the light microscope post 1 week of placing for crystal trays. Crystals of MjIMPDPH^{ΔCBS} obtained in the condition containing 0.05 M magnesium chloride hexahydrate, 0.1 M HEPES, pH 7.5, 30% v/v polyethylene glycol monomethyl ether 550 (**Fig. 3.3**) were diffracted using a Rigaku RU200 X-ray diffractometer with a monochromatic light source of wavelength 1.54179 Å. For detection, an image plate of type MAR scanner 345 mm was used.

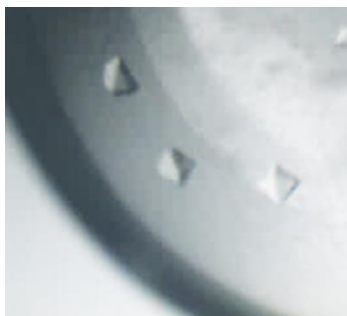


Fig. 3.3. *MjIMPDPH^{ΔCBS} protein crystal morphology.* Crystals were obtained under the condition containing 0.05 M Magnesium chloride hexahydrate, 0.1 M HEPES, pH 7.5, 30% v/v polyethylene glycol monomethyl ether 550

For structure solution Various packages available in CCP4 suite (Winn et al., 2011) and Phenix modules (Adams et al., 2010) were used. Briefly, iMOSFLM (Battye et al., 2011) was used for

data processing, PHASER MR (McCoy, 2006) for phasing and REFMAC 5.0 (Murshudov et al., 2011) and Phenix.refine (Afonine et al., 2012) were used for refinement. Manual refinements were carried out using COOT (Emsley et al., 2010). Refinement statistics are summarized in (Table 3.1). All the structure-related figures were created using PyMol software (Schrödinger, 2015).

Table 3.1. Data collection and refinement statistics of MjIMPDH^{ΔCBS} protein structure

PDB ID	7XLO
Wavelength	1.54
Resolution range	53.37 - 2.6 (2.74 - 2.6)
Space group	I 4
Unit cell	97.65 97.65 168.22 90 90 90
Total reflections	106781 (15333)
Unique reflections	24178 (3511)
Multiplicity	4.4 (4.4)
Completeness (%)	100.0 (100.0)
Mean I/sigma(I)	4.0 (2.3)
Wilson B-factor	14.89
R-merge	0.357 (0.632)
R-meas	0.407 (0.722)
R-pim	0.195 (0.346)
CC1/2	0.845 (0.155)
Reflections used for R-free	1233 (146)
R-work	0.262
R-free	0.278
RMS(bonds)	0.013
RMS(angles)	1.72
Ramachandran favored (%)	91.30
Ramachandran allowed (%)	8.05
Ramachandran outliers (%)	0.64
Rotamer outliers (%)	6.86
Clashscore	8.23
Average B-factor	16.98

3.2.5. Transmission electron microscopy imaging

Aliquots of purified MjIMPDH protein were diluted to 2.5 μ M in buffer containing 50mM Tris HCl, pH 7.0, 2 mM DTT and incubated with or without ATP and GTP at 1mM concentration for 15 minutes at room temperature. Following the incubation, the aliquots were applied to carbon film EM grids and negatively stained using 1% uranyl acetate. Grids were imaged using a FEI Tecnai G2 transmission electron microscope operating at 200 kV and a Gatan ultrascan CCD. Image processing was carried out using the software Gatan digital micrograph.

3.3. Results and discussion

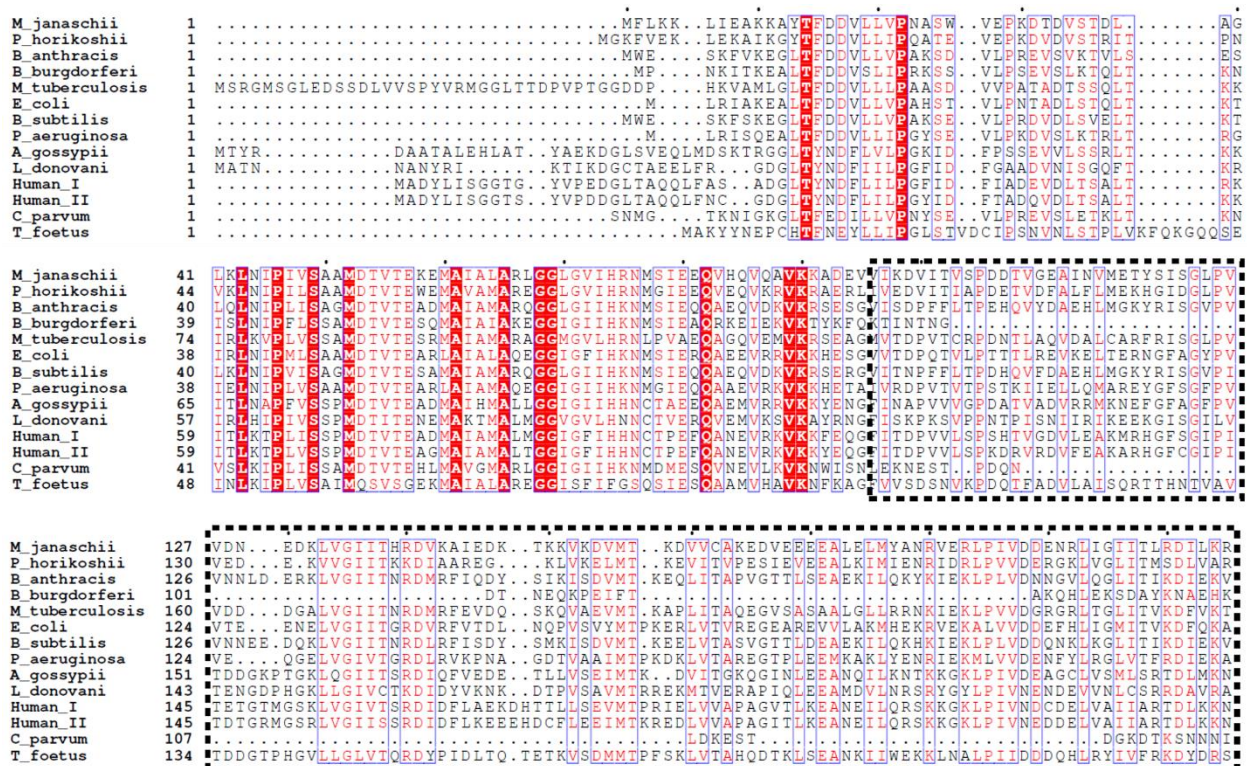
3.3.1. Sequence analysis

MjIMPDH shares greater biochemical similarities with eukaryotic IMPDH rather than with prokaryotic IMPDH, particularly those pertaining to modulation of enzyme activity. To understand the extent of similarity MjIMPDH sequence shares with prokaryotic and eukaryotic IMPDH, sequences from representative organisms for which the structure has been solved and characterized were taken and aligned with MjIMPDH. Alignment was used to obtain the sequence similarity and identity scores (**Table. 3.2**). Among the sequences considered, on an average, MjIMPDH shares around 65% similarity with prokaryotic IMPDH and 50% similarity with eukaryotic IMPDH in contrast to biochemical features of the enzyme.

Table. 3.2. Comparison of sequence similarity and identity of MjIMPDH with IMPDH from organisms for which structures are available

	Organism	Identity %	Similarity %
Eukaryotes	<i>A. gossypium</i>	38	49
	<i>L. donovani</i>	33	45
	<i>Human I</i>	36	48
	<i>Human II</i>	36	48
	<i>C. parvum</i>	50	61
	<i>T. foetus</i>	32	44
Prokaryotes	<i>B. anthracis</i>	53	65
	<i>B. burgdorferi</i>	47	59
	<i>M. tuberculosis</i>	51	63
	<i>E. coli</i>	51	61
	<i>B. substilis</i>	53	67
	<i>P. aeruginosa</i>	51	63
Archaea	<i>P. horikoshii</i>	63	75

Five structural features involved in determining ligand binding and catalysis which include the catalytic cysteine loop, phosphate binding loop, finger loop, C-terminal loop and mobile flap loop were identified in MjIMPDH and found to be similar to IMPDH from other organisms. The key catalytic residues, Cys306, Thr308, Arg405, Tyr406 and other residues that interact with IMP were found to be conserved in MjIMPDH. Conservation of these residues and motifs in MjIMPDH are highlighted in the multiple sequence alignment (Fig. 3.4). Residues interacting with the adenine ring of NAD⁺ however, vary to a large extent and are difficult to identify in sequence alignments (Hedstrom, 2009). IMPDH from all organisms characterized to date are found to require a monovalent cation for catalysis. Potassium ion binding sites within the catalytic cysteine loop and C-terminal segment of the protein were found to be well conserved in MjIMPDH. GDP and GTP inhibit the eukaryotic but not the prokaryotic IMPDH in a mixed fashion by binding to a non-canonical nucleotide binding site. These residues were found to be highly conserved in eukaryotic IMPDH and not prokaryotic IMPDH as has been reported earlier (Buey, Ledesma-Amaro, Velázquez-Campoy, et al., 2015). These residues are poorly conserved in archaeal IMPDH as well, including MjIMPDH (Fig. 3.5).



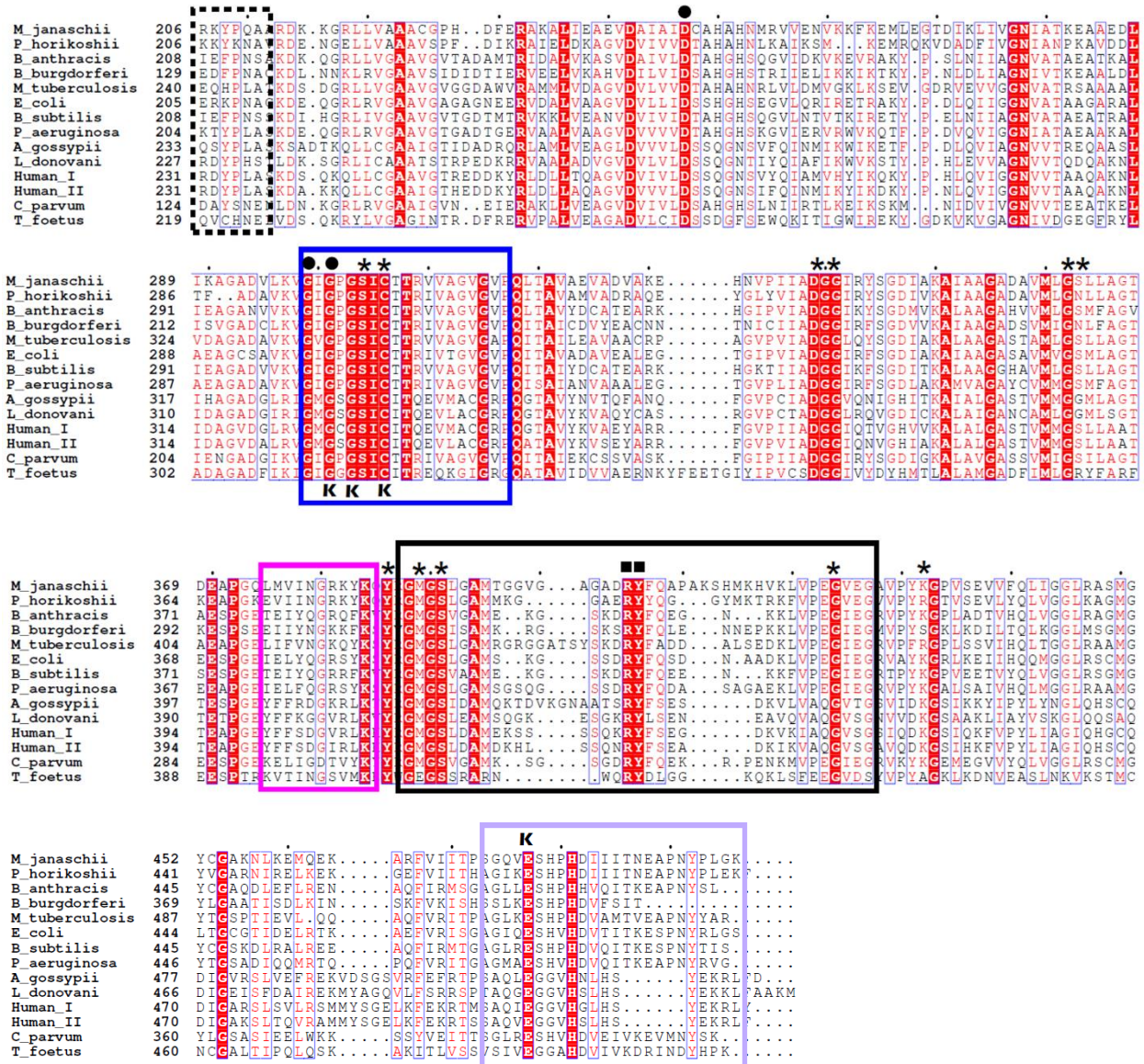


Fig. 3.4. Multiple sequence alignment of IMPDH from various organisms show that MjIMPDH has all the catalytic residues conserved. Sequences of IMPDHs from prokaryotic, eukaryotic and archaeal sources for which crystal structures are available were aligned using Clustal Omega (Sievers et al., 2011) and rendered using ESPRIPT (Robert & Gouet, 2014). Consensus >70 % was found across all sequences analyzed. Residues from the catalytic cysteine loop are enclosed in a blue box, phosphate binding loop with a black dashed box, finger loop with a pink box, mobile flap loop with a black box and c-terminal loop with a purple box. Residues that interact with IMP are highlighted with a black asterisk. Conserved NAD interacting residues are highlighted with a black dot. Catalytic residues in the mobile flap are highlighted with a black square. The residue that interacts with monovalent cations is marked with the letter 'K'.

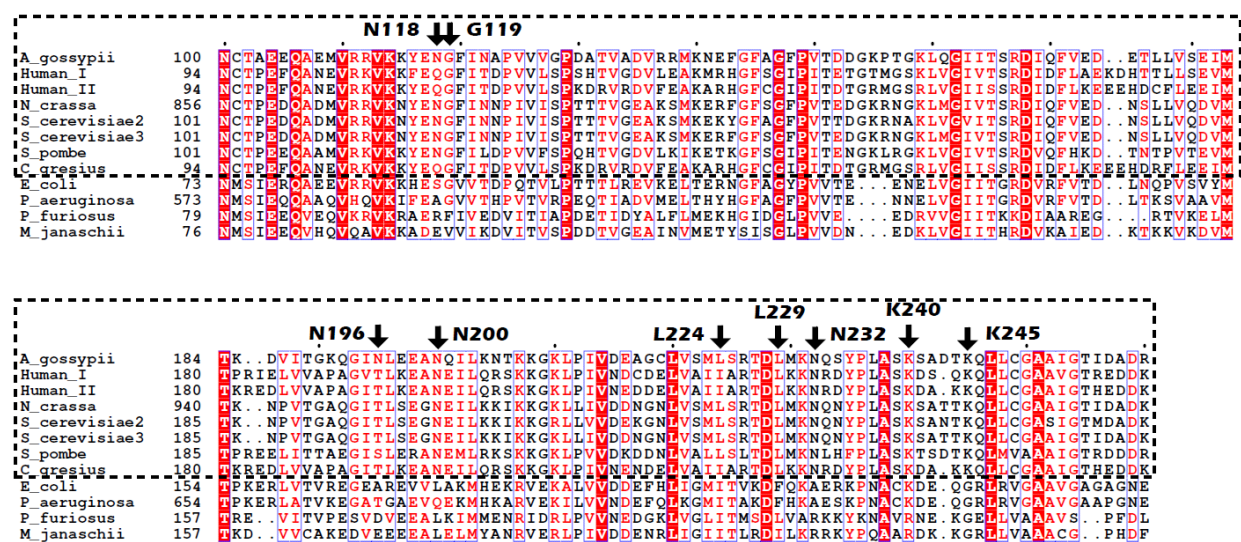


Fig. 3.5. Multiple sequence alignment of Bateman domain of IMPDH from various organisms highlighting the non-canonical nucleotide binding site observed in eukaryotic IMPDH. Sequences were aligned using Clustal Omega (Sievers et al., 2011) and rendered using ESPRIPT (Robert & Gouet, 2014). Key residues that interact with GDP in the non-canonical nucleotide binding site of eukaryotic IMPDH (enclosed within a dotted black box) are indicated.

3.3.2. Structure of IMPDH TIM domain

To elucidate the nucleotide binding site in the Bateman domain and the cation binding site of MjIMPDH, full-length protein was purified, and attempts were made at obtaining crystals of the same. Most of the population of full-length protein migrates at a high molecular mass as observed on preparative size-exclusion chromatography, suggesting that the protein is prone to the formation of higher order oligomers in solution (**Fig. 3.6A**). To maintain homogeneity, the higher order oligomers and the fractions corresponding to octamers were used for setting up crystal trays separately. While almost all conditions in which the higher order oligomers placed for crystallization precipitated, octamers formed crystals under numerous conditions (**Appendix. Table D.1**). All the crystals formed were small and lacked sufficient water content (**Appendix. Fig. D.1**). None of the crystals were of diffraction quality and crystals diffracting to a high resolution could not be obtained to collect data for structure solution. With the intent to understand the potassium ion binding site of the protein better, MjIMPDH core catalytic domain (MjIMPDH^{ACBS}) was expressed (**Fig. 3.6A**) and numerous crystals of the same were obtained as well (**Appendix. Table D.2**). Structure of the core catalytic domain was solved using *Pyrococcus horikoshii* IMPDH structure (2cuo) as a MR model template and refined to a resolution of 3.1 Å (**Table 3.1**). Unit cell of the solved structure consisted of two protomers of catalytic subdomains,

although the functional form of the enzyme is a tetramer as evident from previous biochemical characterization of the protein at the laboratory. Solved structure of the catalytic domain superposed well with the MR model used, *Pyrococcus horikoshii* full length IMPDH structure with an RMSD of 0.609 Å (**Fig. 3.6B**). The catalytic cysteine loop, phosphate binding loop and the finger domain have been mapped in both structures (**Fig. 3.6C**), but the mobile flap region and C-terminal segment which were partially mapped in *P. horikoshii* IMPDH structure were missing in the MjIMPDH structure. Superposing MjIMPDH structure on *Streptococcus pyogenes* IMPDH (1zfv) having bound IMP, showed that the IMP binding site is structurally well conserved as it is across all classes of IMPDH. Majority of the residues in the pocket that accommodate IMP are present in loops. Loops from both MjIMPDH and SpIMPDH superpose very well despite MjIMPDH structure being solved in the apo-state, suggesting that these loops are not dynamic, but rather pre-organized to bind to the ligand (**Fig. 3.4D**). C-terminal segment of IMPDHs is coupled to the active site through a monovalent cation. Superposition of MjIMPDH structure on *Ashbya gossypium* IMPDH (4z87) in which the bound potassium ion has been mapped, revealed that the conserved residues in the catalytic cysteine loop that bind a potassium ion (G329, G331 and C334 from AgIMPDH) are structurally similar in both proteins (**Fig. 3.4E**). These residues were observed to be well poised to accommodate the potassium ion in MjIMPDH. Remainder of the residues that interact with potassium ion in AgIMPDH, which are conserved at the C-terminal segment of the protein were not mapped in the MjIMPDH catalytic TIM domain structure and hence couldn't be analyzed.

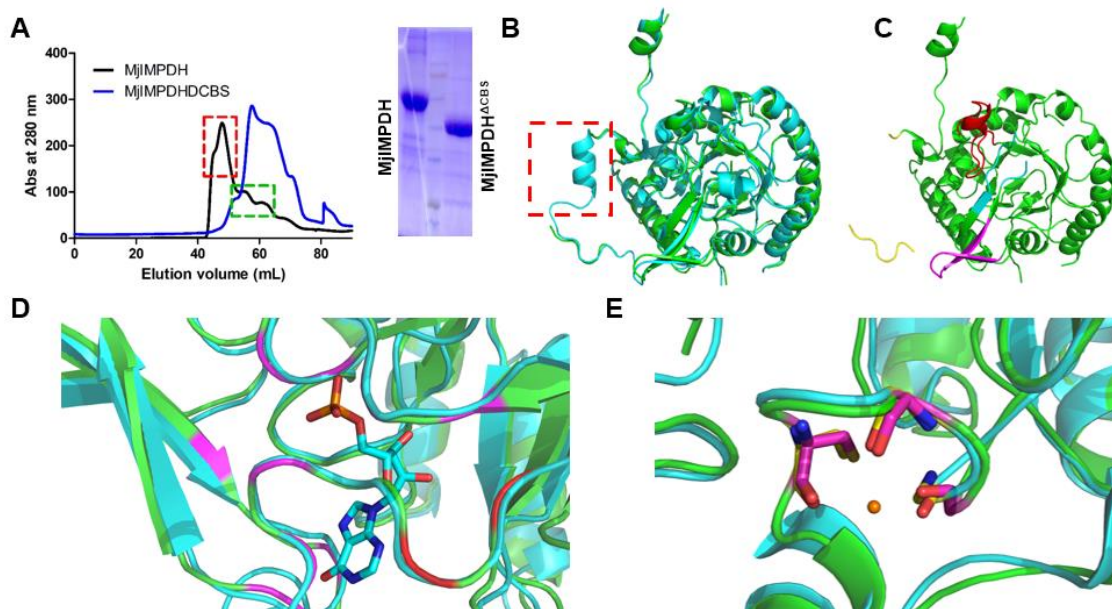


Fig. 3.6. MjIMPDPH protein purification and MjIMPDPH^{ACBS} structure analysis. **A.** Elution profile of MjIMPDPH and MjIMPDPH^{ACBS} on preparative grade size-exclusion chromatography (left panel). Enclosed in red dotted box is the fraction corresponding to higher order oligomers of MjIMPDPH migrating at a higher molecular mass. Enclosed in green dotted box is the fraction corresponding to octamers of IMPDH which was collected separately for crystallization and TEM imaging. Purified fractions of MjIMPDPH (55 kDa) and MjIMPDPH^{ACBS} (43 kDa) were examined on SDS-PAGE (right panel). **B.** Superposition of MjIMPDPH^{ACBS} (green) on *Pyrococcus horikoshii* IMPDH (cyan) (2cuo) that was used as the MR model, revealed a good alignment. A small region around the C-terminal segment of the protein which couldn't be mapped in MjIMPDPH but was mapped in PhIMPDPH is indicated within a red box. **C.** Mapped catalytic cysteine loop of the MjIMPDPH structure is highlighted in red; finger domain in magenta; partially mapped mobile flap and C-terminal segment in cyan and yellow, respectively. **D.** Superposition of MjIMPDPH (green) over *Streptococcus pyogenes* IMPDH (cyan) (1zfi) shows that the IMP binding residues (highlighted in pink) located in loops are pre-organized in MjIMPDPH to accommodate the substrate. **E.** Superposition of MjIMPDPH (green) over *Ashbya gossypium* IMPDH (cyan) (4z87) shows that one cysteine and two glycine residues (highlighted in magenta) in MjIMPDPH are well poised to hold the potassium ion (orange sphere) in place.

3.3.3. Transmission electron microscopy imaging of MjIMPDPH

MjIMPDPH forms higher order oligomers which is apparent from its migration at higher molecular mass during the size-exclusion chromatography step of protein purification (**Fig. 3.6A**). A large proportion of the protein in size-exclusion chromatography also goes into the void volume. To understand if there is a structural rearrangement of MjIMPDPH mediated by nucleotide modulators as has been reported for eukaryotic IMPDPHs, TEM imaging was carried out. Images were acquired for the apo-protein and protein incubated with ATP and GTP. Cytophidia filaments were

observed in case of the protein incubated with ATP (**Fig. 3.7A**), while the same was not observed with protein without modulator or with addition of GTP (**Fig. 3.7B, C**).

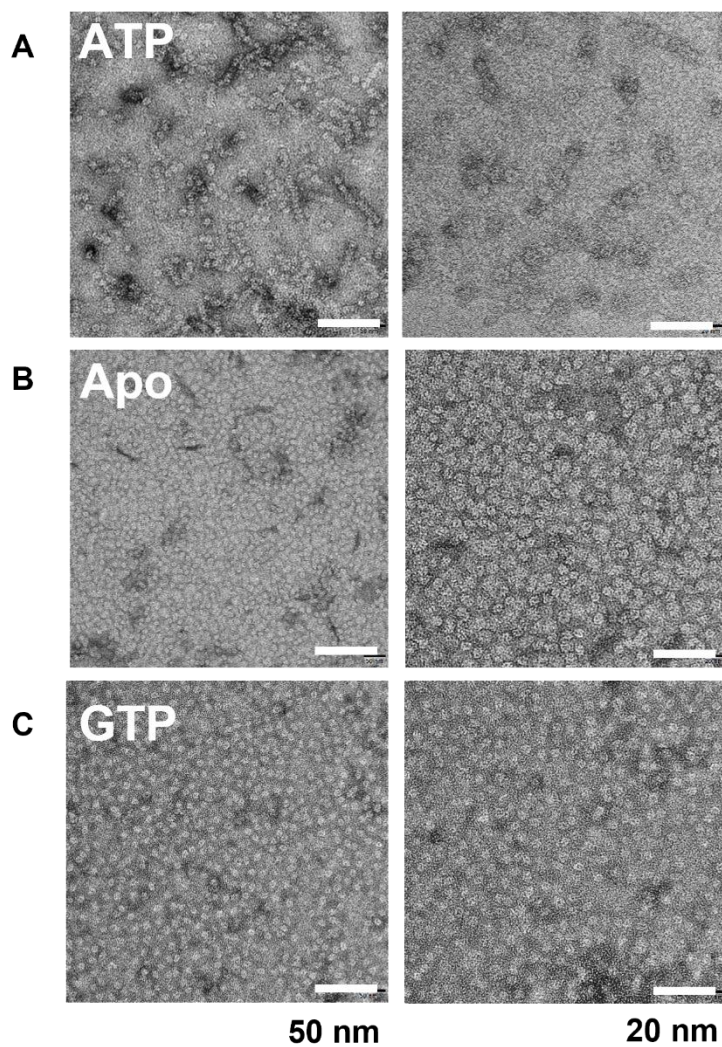


Fig. 3.7. Negative-stain electron microscopy imaging of purified MjIMPDH. A. MjIMPDH incubated with 1mM ATP. B. MjIMPDH apo-protein. C. MjIMPDH incubated with 1mM GTP. Images of scale bar 50 nm and 20 nm are presented.

3.4. Summary

Although biochemical, kinetic, and structural characterization has been carried out in IMPDH from prokaryotes and eukaryotes to great detail, archaeal IMPDH remains to be characterized to date. Previous work from our laboratory had revealed interesting biochemical aspects in IMPDH from the archaea, *M. jannaschii* that has not been observed in this enzyme from other organisms. Although MjIMPDH shares a much higher identity with prokaryotic IMPDHs than eukaryotic IMPDHs with reports of phylogenetic tree showing archaeal IMPDH grouping with prokaryotic proteins (Tamura et al., 2011), MjIMPDH has revealed inhibition by guanine nucleotides and a lack of altered catalytic activity on ATP binding, which are exclusive features of the eukaryotic type IMPDH. A non-canonical nucleotide binding site in the CBS domain of the eukaryotic IMPDH to which guanine nucleotides bind and inhibit the enzyme is absent in prokaryotic IMPDH. A sequence alignment of IMPDH from various organisms revealed that the non-canonical nucleotide binding site seen in eukaryotic IMPDH is not conserved in MjIMPDH (**Fig. 3.4**), which does not explain how the enzyme is inhibited by guanine nucleotides. CBS domain of IMPDH displays an exceedingly high degree of disorder, making it difficult to map the domain in X-ray structures. Only a small proportion of the X-ray structures of IMPDH solved have the CBS domain mapped, disallowing easy understanding of enzyme modulation. The only archaeal IMPDH structure solved from *Pyrococcus horikoshii* also has the CBS domain missing. With this view, structure solution of full-length MjIMPDH was attempted with the hope of mapping the CBS domain. Numerous crystals formed under various conditions go to show that the protein is prone to crystallization. Unfortunately, however, none of the crystals were of diffraction quality and multiple efforts at obtaining large crystals failed. Disorder of the CBS domain could have been a possible reason why well-formed crystals weren't obtained. All IMPDHs reported are found to require K^+ ions for optimal activity with its binding increasing the activity of the enzyme by ~100 folds (Hedstrom, 2009). MjIMPDH was found to have a unique feature of not requiring K^+ or any other monovalent cation for full activity which has not been reported in any other IMPDHs studied to date. Removal of the CBS sub-domain has been reported earlier to facilitate IMPDH crystallization (Hedstrom, 2009), hence, structure of the MjIMPDH core catalytic domain (MjIMPDH^{ΔCBS}) was solved to understand differences in the monovalent cation binding site of the protein. Potassium ion binding site within the protein comprises of six main-chain carbonyl interactions, including three residues from the catalytic cysteine loop and three residues from the

C-terminal segment of the protein (Riera et al., 2011). Four of the total six residues involved in K^+ binding are well conserved in MjIMPDH (**Fig. 3.4**). The C-terminal segment of the protein was not mapped in the MjIMPDH^{ΔCBS} structure, but carbonyl groups from the three residues of the catalytic cysteine loop (G301, G303 and C306) were well poised to bind the K^+ ion as was observed from superposition of the structure over *Ashbya gossypium* IMPDH structure (4z87) having the bound K^+ ion mapped (**Fig. 3.6E**). Further cues on the lack of requirement of a K^+ ion for activity could not be deciphered from the MjIMPDH^{ΔCBS} structure. Comparison of the IMP binding site of MjIMPDH^{ΔCBS} with that of *Streptococcus pyogenes* IMPDH having a bound IMP mapped, also revealed that the loops of the IMP binding site in apo-protein of MjIMPDH are preorganized to bind IMP.

IMPDH have been shown to form filamentous structures in vertebrate cells (Calise et al., 2014; Carcamo et al., 2014; Ji et al., 2006) and in human IMPDH1 (Fernández-Justel et al., 2019) and IMPDH2 (Johnson & Kollman, 2020) isoforms *in vitro*. These filaments called cytoophidia which form upon ATP binding is observed only in eukaryotes and has been a subject of study in the recent past (Burrell et al., 2022). To understand if such structural modulation also takes place in the archaeal IMPDH, negative-stain electron microscope images of MjIMPDH protein was acquired in the presence and absence of modulators ATP and GTP. Cytoophidia filaments formed in MjIMPDH in the presence of ATP supports the kinetic data which as well suggests that archaeal IMPDH have features similar to eukaryotic rather than prokaryotic homologs (**Fig. 3.7A**). Recently it was shown that CTPS, one of the other proteins which is known to form filaments in both prokaryotes and eukaryotes, also formed cytoophidia filaments in archaea (Zhou et al., 2020). This is the first study to reveal that IMPDH from archaea form cytoophidia filaments *in vitro*.

3.5. Future directions

To answer the question as to why MjIMPDH has biochemical features closer to the eukaryotic counterparts compared to prokaryotes, although sharing a higher sequence identity with prokaryotes; nucleotide mediated modulation, which is the distinguishing feature between the prokaryotic and eukaryotic IMPDH has to be studied in much greater detail. The number of nucleotide binding sites in the CBS domain and the presence or absence of a non-canonical nucleotide binding site has to be probed. Perhaps one of the bottlenecks in understanding IMPDH protein in detail is the difficulty in mapping all its domains in the X-ray crystal structure due to the large degree of disorder it displays. Resorting to cryo-electron microscopy might help answer some of the questions posed by TEM imaging and kinetic studies on MjIMPDH. Filaments formed on ATP binding in human IMPDH can take up either a relaxed active conformation or a compressed inactive conformation based on the binding of GTP to the non-canonical nucleotide binding site (Burrell et al., 2022). In the absence of complete conservation of residues in the non-canonical nucleotide binding site, elucidation of the mode of binding of nucleotides that lead to filament formation in MjIMPDH requires availability of the structure of the full-length enzyme. Previous studies on human IMPDH have shown that certain mutations can lead to disruption of filament formation entirely (non-assembly mutants) or an overly tight filament formation (continuous assembly mutants) (Anthony et al., 2017; Burrell et al., 2022). Residue Y12 present in the N-terminus loop extension, at the human IMPDH tetramer interface interacts with R356 from the adjacent protomer. Mutations Y12A and R356A have been shown to disrupt filament formation in human IMPDH. Mutation of R224P in the CBS domain, a retinal disease-causing mutation, also disrupted filament formation. A continuous filament assembly has been observed in the mutant S275L, a residue located at the active site and tetramerization interface of IMPDH. N-terminus loop is not mapped in most prokaryotic and archaeal IMPDH. Also, none of the above residues are conserved. How these residues have an impact on filament assembly and disassembly is not understood. Mutational analysis of the interface residues in MjIMPDH can be carried out taking leads from sequence alignment and structure to understand residues critical for polymerization similar to what has been done in human IMPDH. Obtaining more non-assembly mutants and continuous assembly mutants will throw further light on the peculiar feature of filament formation in IMPDHs in general.

Appendix A

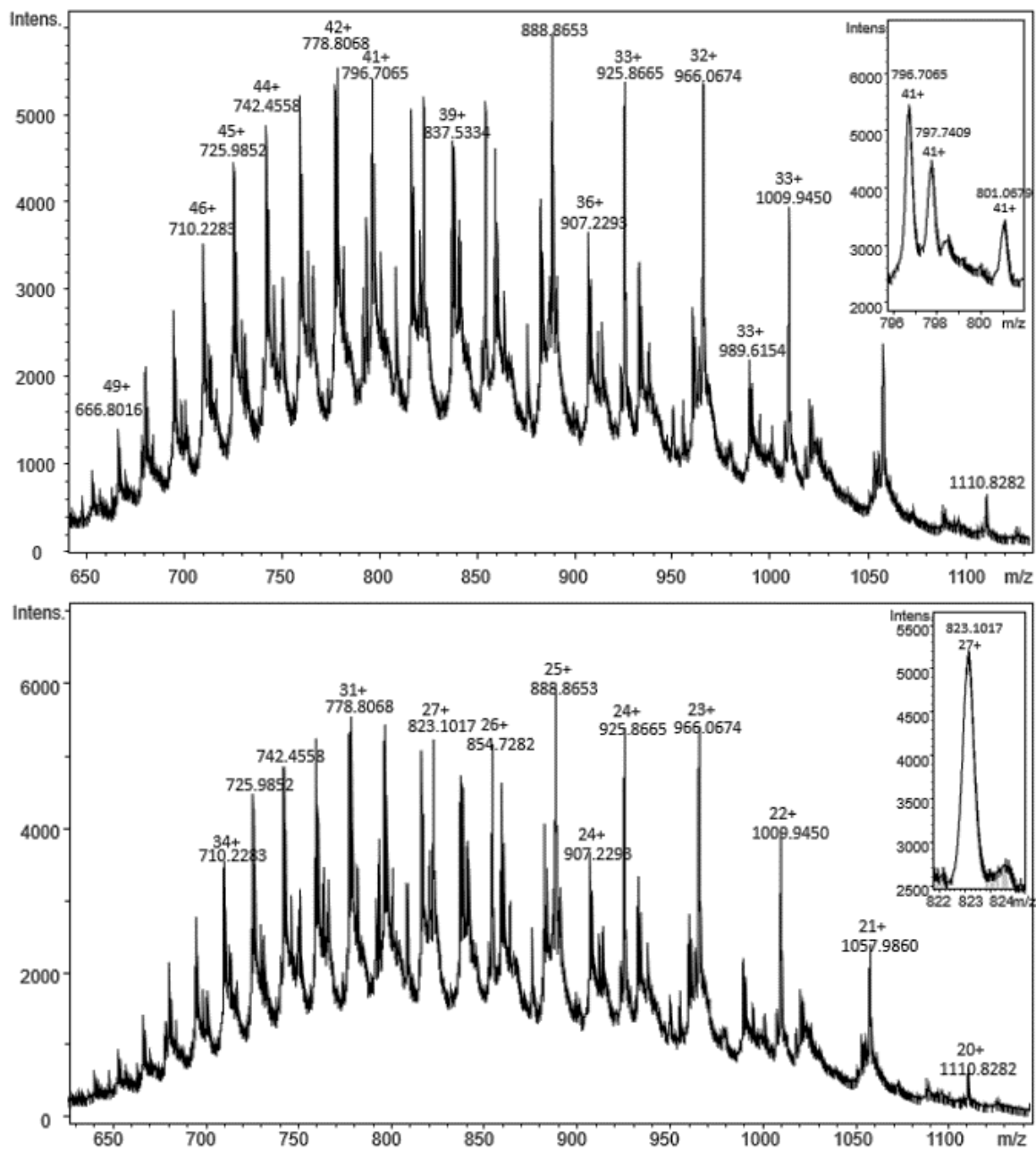


Fig. A.1. LC-ESI-MS of selenomethionine incorporated MjFH subunits. Top panel, mass spectrum of α -subunit. Inset, +41 charge state corresponding to a mass of 32624.99, which is 187 Da more than that of the native protein indicating replacement of all four methionines with selenomethionines. Bottom panel, mass spectrum of β -subunit. Inset, +27 charge state corresponding to a mass of 22197, which is 234 Da more than native protein mass indicating replacement of all seven methionines with selenomethionines.

Appendix B

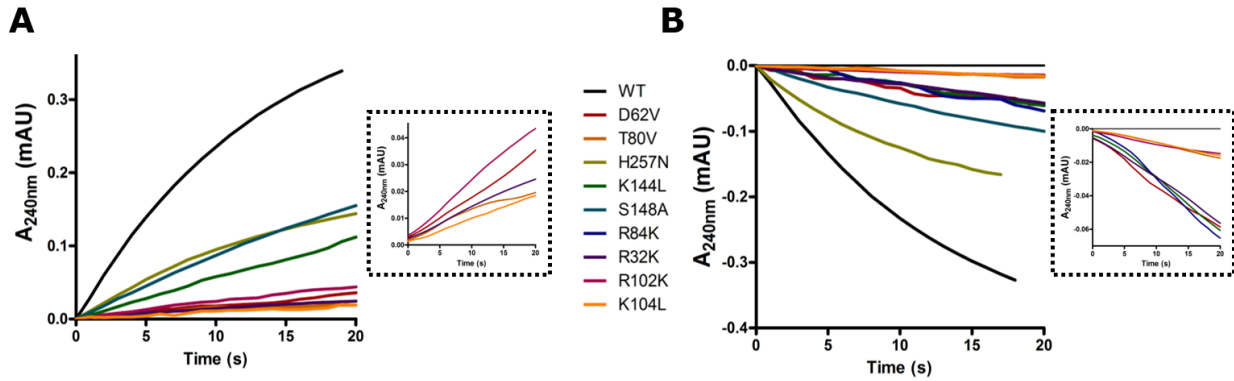


Fig. B.1. Representative time course curves for MjFH WT and active mutants on A. 1 mM malate and B. 500 μ M fumarate. Inset: expansion of Y-axis to highlight the presence of activity even in the weakly active mutants.

Appendix C

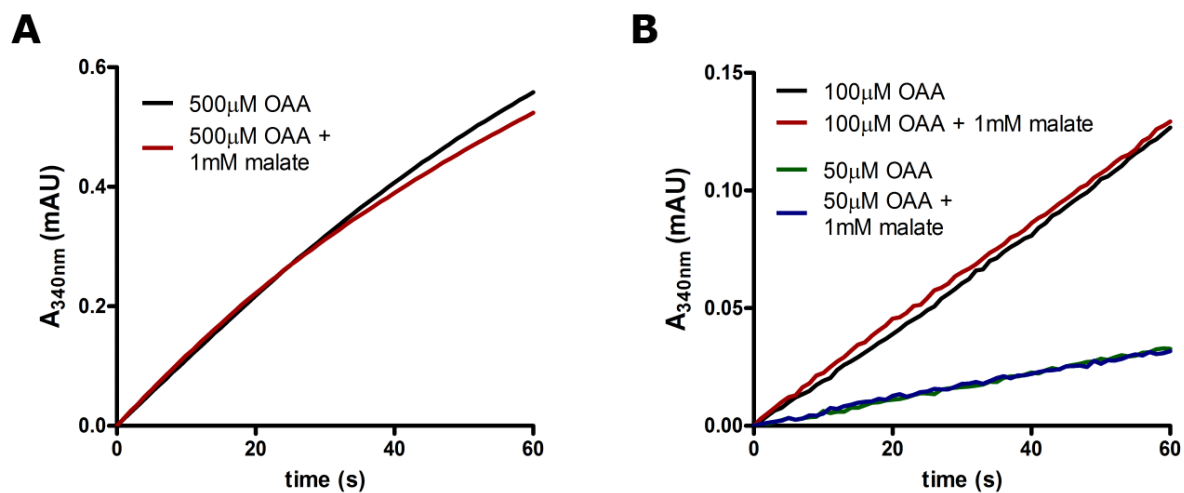


Fig. C.1. Time course curves for conversion of OAA to L-malate by MDH in the presence of L-malate in the assay mixture (pH 7.4). The progress curves show that the enzyme is not inhibited by 1 mM L-malate with 500 μM (A), 100 μM and 50 μM (B) concentrations of OAA as substrate.

Appendix D

Table D.1. Conditions under which full-length MjIMPDH crystals were obtained

Sl. No.	Condition Components		
	Salt (Concentration in M)	Buffer pH (Concentration in M)	Precipitant (Concentration in %)
1	DL-Malic acid	-	20 % w/v PEG 3350
2	0.1 M Potassium thiocyanate	-	30 % w/v PEG 2000 MME
3	0.15 M Potassium bromide	-	30 % w/v PEG 2000 MME
4	0.2 M Sodium acetate trihydrate	-	20 % w/v PEG 3350
5	0.2 M Sodium citrate tribasic dihydrate	-	20 % w/v PEG 3350
6	0.2 M Sodium formate	-	20 % w/v PEG 3350
7	0.2 M Potassium sodium tartrate tetrahydrate	-	20 % w/v PEG 3350
8	0.2 M Potassium thiocyanate	-	20 % w/v PEG 3350
9	-	0.1 M PCTP pH 5.0	25 % w/v PEG 1500
10	0.2 M Lithium chloride	0.1 M HEPES pH 7.0	20 % w/v PEG 6000
11	-	0.1 M MMT pH 6.0	25 % w/v PEG 1500
12	0.2 M Sodium bromide	-	20 % w/v PEG 3350
13	0.2 M Sodium iodide	-	20 % w/v PEG 3350
14	0.2 M Potassium thiocyanate	-	20 % w/v PEG 3350
15	0.2 M Sodium nitrate	-	20 % w/v PEG 3350
16	0.2 M Sodium formate	-	20 % w/v PEG 3350
17	0.2 M Sodium acetate trihydrate	-	20 % w/v PEG 3350
18	0.2 M Potassium sodium tartrate tetrahydrate	-	20 % w/v PEG 3350
19	0.2 M Potassium iodide	pH 7	20% w/v Polyethylene glycol 3,350
20	0.2 M Potassium thiocyanate	pH 7	20% w/v Polyethylene glycol 3,350
21	0.2 M Sodium acetate trihydrate	pH 8	20% w/v Polyethylene glycol 3,350
22	0.15 M Cesium chloride	-	15% w/v Polyethylene glycol 3,350
23	1% w/v Tryptone, 0.001 M Sodium azide	-	12% w/v Polyethylene glycol 3,350
24	-	0.1 M MES monohydrate pH 6.0	22% v/v Polyethylene glycol 400
25	-	0.1 M MES monohydrate pH 6.0	20% v/v Jeffamine® M-600® pH 7.0
26	-	0.1 M Sodium acetate trihydrate pH 4.0	10% w/v Polyethylene glycol monomethyl ether 2,000
27	-	0.1 M Citric acid pH 3.5	25% w/v Polyethylene glycol 3,350
28	-	0.1 M HEPES pH 7.5	12% w/v Polyethylene glycol 3,350
29	-	0.1 M MES monohydrate pH 6.0	14% w/v Polyethylene glycol 4,000
30	-	0.1 M Sodium citrate tribasic dihydrate pH 5.0	18% w/v Polyethylene glycol 20,000

Table D.2. Conditions in which MjIMPDH^{ΔCBS} crystals were obtained

Sl. No.	Condition Components		
	Salt (Concentration in M)	Buffer pH (Concentration in M)	Precipitant (Concentration in %)
1	0.2 M Magnesium chloride hexahydrate	0.1M Tris-7.0	10% w/v PEG 8000
2	0.2 M Magnesium chloride hexahydrate	0.1M Tris-7.0	10% w/v PEG 8000
3	3.2 M Ammonium sulphate	0.1 M Citrate- 5.0	-
4	0.2 M Ammonium sulphate	0.1 M BIS-Tris pH 5.5	25% w/v PEG 3350
5	1.0 M Sodium citrate tribasic dihydrate	0.1 M Sodium cacodylate pH 6.5	-
6	0.2 M Sodium chloride	0.1 M HEPES pH 7.5	10 % v/v 2-Propanol
7	2.1 M DL-Malic acid	pH 7.0	-
8	0.2M Lithium sulphate	0.1 M Sodium HEPES pH 7.5	25% w/v PEG 3350
9	0.2 M Ammonium sulfate	0.1 M Sodium HEPES pH 7.5	25 % w/v PEG 3350
10	0.2 M Ammonium tartrate dibasic	-	20 % w/v PEG 3350
11	0.2 M Lithium chloride	0.1 M Tris pH 8.0	20% w/v PEG 6000
12	0.2 M Sodium fluoride	0.1 M BIS-Tris propane pH 6.5	20% w/v PEG 3350
13	0.2 M Sodium iodide	0.1 M BIS-Tris propane pH 6.5	20% w/v PEG 3350
14	0.2 M Sodium nitrate	0.1 M BIS-Tris propane pH 6.5	20% w/v PEG 3350
15	0.2 M Sodium formate	0.1 M Bis-Tris propane pH 8.5	20 % w/v PEG 3350
16	0.2 M Sodium acetate trihydrate	0.1 M Bis-Tris propane pH 8.5	20 % w/v PEG 3350
17	0.02 M Sodium/potassium phosphate	0.1 M Bis-Tris propane pH 8.5	20 % w/v PEG 3350
18	0.05 M Magnesium chloride hexahydrate	0.1 M HEPES pH 7.5	30% v/v Polyethylene glycol monomethyl ether 550
19	0.2 M Ammonium sulfate	0.1 M BIS-TRIS pH 5.5	25% w/v Polyethylene glycol 3,350
20	0.2 M Sodium chloride	0.1 M HEPES pH 7.5	25% w/v Polyethylene glycol 3,350
21	0.2 M Lithium sulfate monohydrate	0.1 M HEPES pH 7.5	25% w/v Polyethylene glycol 3,350
22	0.2 M Magnesium nitrate hexahydrate	pH 5.9	20% w/v Polyethylene glycol 3,350

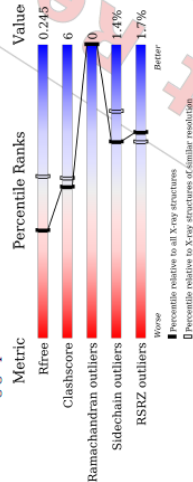
**Fig. D.1. The crystals of MjIMPDH obtained from multiple conditions were small and not of diffraction quality.**

1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.46 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Rfree	130704	1544 (2.48-2.44)
Clashscore	141614	1613 (2.48-2.44)
Ramachandran outliers	138981	1598 (2.48-2.44)
Sidechain outliers	138945	1598 (2.48-2.44)
RSRZ outliers	127900	1523 (2.48-2.44)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	298	<div style="display: flex; align-items: center;"> <div style="width: 4%; height: 10px; background-color: red; margin-right: 2px;"></div> <div style="width: 81%; height: 10px; background-color: orange; margin-right: 2px;"></div> <div style="width: 15%; height: 10px; background-color: yellow; margin-right: 2px;"></div> <div style="width: 0%; height: 10px; background-color: green; margin-right: 2px;"></div> <div style="width: 0%; height: 10px; background-color: grey; margin-right: 2px;"></div> <div style="width: 0%; height: 10px; background-color: white; margin-right: 2px;"></div> <div style="margin-left: 2px;">4%</div></div>

Full wwPDB X-ray Structure Validation Report i

Apr 22, 2022 - 03:16 PM JST

PDB ID : 7XKY

Title : Crystal structure of Methanocaldococcus jamareschii two-subunit Fumarate hydratase apo-protein complex
 Deposited on : 2022-04-20
 Resolution : 2.46 Å (reported)

This wwPDB validation report is for manuscript review

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB bioautomation pipeline after annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the i symbol.

The following versions of software and data (see references i) were used in the production of this report:

MolProbity : 4.02b-467
 Mogul : 1.8.5 (37.4361), CSD as541be (2020)
 Xtrige (Phenix) : 1.13
 EDS : 2.27
 Poromonte statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
 Refmac : 5.8.0158
 CCP4 : 7.0.044 (Garagrow)
 Ideal geometry (proteins) : Engh & Huber (2001)
 Ideal geometry (DNA, RNA) : Parkin et al. (1996)
 Validation Pipeline (wwPDB-VI) : 2.27

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3567 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy; the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Fumarate hydratase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	282	2068	1328	351	380	5	4	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-12	MSE	-	expression tag	UNP A0A832S1UM4
A	-11	GLY	-	expression tag	UNP A0A832S1UM4
A	-10	SER	-	expression tag	UNP A0A832S1UM4
A	-9	SER	-	expression tag	UNP A0A832S1UM4
A	-8	HIS	-	expression tag	UNP A0A832S1UM4
A	-7	HIS	-	expression tag	UNP A0A832S1UM4
A	-6	HIS	-	expression tag	UNP A0A832S1UM4
A	-5	HIS	-	expression tag	UNP A0A832S1UM4
A	-4	HIS	-	expression tag	UNP A0A832S1UM4
A	-3	HIS	-	expression tag	UNP A0A832S1UM4
A	-2	SER	-	expression tag	UNP A0A832S1UM4
A	-1	GLN	-	expression tag	UNP A0A832S1UM4
A	0	ASP	-	expression tag	UNP A0A832S1UM4
A	1	PRO	-	expression tag	UNP A0A832S1UM4

- Molecule 2 is a protein called Fumarate hydratase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
2	B	133	1434	926	233	266	2	7	0	0	0


- Molecule 3 is water.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	N	O	S	Se		
3	A	43	43	43	43	0	0	0	0	0

Continued on next page...

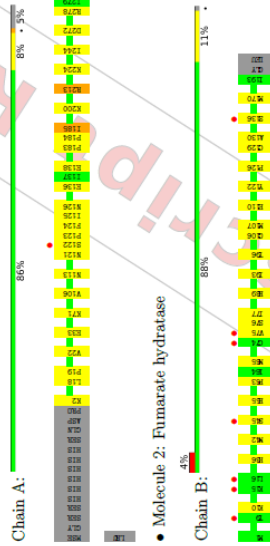
Continued from previous page...

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	N	O	S	Se		
3	B	22	22	22	22	0	0	0	0	


3 Residue-property plots 

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Fumarate hydratase



- Molecule 2: Fumarate hydratase

4 Data and refinement statistics 

Property	Value	Source
Space group	C1 2 1	Depositor ¹
Cell constants	116.35Å 61.70Å 83.88Å	Depositor
a, b, c, α, β, γ	90.00° 123.90° 90.00°	Depositor
Resolution (Å)	52.13 – 2.46	EDS
% Data completeness (in resolution range)	52.07 – 2.46	Depositor
	98.1 (52.13-2.46)	EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$	3.33 (at 2.45Å)	Xtriage
Refinement program	REFMAC 5.8.0238	Depositor
R, R_{free}	0.184 0.224	Depositor
R_{free} test set	0.205 0.245	DCC
Wilson B-factor (Å ²)	936 reflections (5.21%)	wwPDBB-VP
Anisotropy	30.3	Xtriage
Bulk solvent $k_{vol}(e/\text{Å}^3)$, $B_{sol}(Å^2)$	0.483	Xtriage
	0.37, 53.7	EDS
L-test for twinning ²	$\langle I \rangle = 0.47$, $\langle I^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o/F_c correlation	0.93	EDS
Total number of atoms	3567	wwPDBB-VP
Average B, all atoms (Å ²)	34.0	wwPDBB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.47% of the height of the origin peak. No significant pseudotranslation is detected.

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |I| \rangle >$, $\langle I^2 \rangle >$ for accurate reflections are 0.5, 0.333 respectively for unmerged datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.79	0/2091	0.90	0/2825
2	B	0.68	0/1455	0.90	0/1969
All	All	0.74	0/3546	0.90	0/4794

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2068	0	2107	22	0
2	B	1434	0	1381	17	0
3	A	43	0	0	1	0
3	B	22	0	0	0	0
All	All	3567	0	3488	39	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 6.

All (39) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:122:TYR:CG	2:B:170:MSE:HE3	2.16	0.81

Continued on next page...



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:33:GLU:HB3	1:A:244:ILE:HD13	1.64	0.79
1:A:33:GLU:HB3	1:A:244:ILE:CD1	2.12	0.79
2:B:63:PRO:HB3	2:B:77:ILE:HD12	1.67	0.77
2:B:76:SER:HB2	2:B:130:ALA:HB1	1.78	0.65
2:B:106:GLY:HA2	2:B:170:MSE:HE2	1.83	0.59
2:B:122:TYR:CD1	2:B:170:MSE:HE3	2.36	0.59
2:B:63:PRO:HB2	2:B:65:MSE:HE2	1.86	0.57
1:A:126:ASN:OD1	1:A:200:LYS:HE2	2.05	0.56
1:A:183:PRO:HA	1:A:185:ILE:HG23	1.87	0.56
1:A:123:PRO:O	1:A:125:ILE:HG13	2.09	0.53
1:A:183:PRO:HA	1:A:184:PRO:C	2.29	0.53
2:B:93:ILE:O	2:B:96:THR:O	2.27	0.53
1:A:2:LYS:HG2	1:A:2:LYS:O	2.10	0.52
2:B:106:GLY:HA2	2:B:170:MSE:CE	2.41	0.50
2:B:126:PRO:HB2	2:B:129:CYS:SG	2.52	0.49
1:A:122:SER:HB3	1:A:123:PRO:HD3	1.94	0.49
2:B:107:MSE:HG3	2:B:170:MSE:HE1	1.95	0.48
1:A:244:ILE:HD12	1:A:244:ILE:N	2.28	0.48
1:A:213:ARG:HE	1:A:224:LYS:HG3	1.79	0.47
1:A:19:PRO:HD3	3:A:318:HO:O	2.15	0.47
2:B:110:GLU:OE1	2:B:110:GLU:N	2.47	0.47
1:A:18:LEU:CD1	1:A:272:ASP:O	2.63	0.47
1:A:213:ARG:HH11	1:A:213:ARG:HA	1.80	0.46
2:B:10:LYS:HE2	2:B:55:GLU:HG2	1.98	0.45
1:A:7:LYS:HB2	1:A:138:GLU:HB3	2.00	0.44
1:A:22:VAL:HG21	1:A:272:ASP:HB3	2.00	0.44
1:A:18:LEU:HD12	1:A:272:ASP:O	2.18	0.43
1:A:121:ASN:O	1:A:124:PHE:CE1	2.71	0.43
2:B:36:HIS:NE2	2:B:89:GLU:OE1	2.51	0.43
1:A:136:GLU:HG2	1:A:280:LYS:HA	2.00	0.43
2:B:107:MSE:N	2:B:170:MSE:HE1	2.34	0.43
2:B:77:ILE:HG23	2:B:77:ILE:O	2.19	0.42
1:A:122:SER:O	1:A:124:PHE:N	2.52	0.42
1:A:106:VAL:HG12	1:A:113:ASN:HA	2.01	0.42
1:A:136:GLU:CG	1:A:280:LYS:HA	2.50	0.41
2:B:42:MSE:O	2:B:45:SER:O	2.38	0.41
2:B:10:LYS:HG2	2:B:55:GLU:OE2	2.19	0.41
1:A:126:ASN:OD1	1:A:200:LYS:CE	2.70	0.40

There are no symmetry-related clashes.



5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1 A	280/298 (94%)	270 (96%)	10 (4%)	0	100 100
2 B	191/195 (98%)	183 (96%)	8 (4%)	0	100 100
All	471/493 (96%)	453 (96%)	18 (4%)	0	100 100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol Chain	Analysed	Rotameric	Outliers	Percentiles
1 A	207/246 (84%)	204 (99%)	3 (1%)	67 77
2 B	143/162 (88%)	141 (99%)	2 (1%)	67 77
All	350/408 (86%)	345 (99%)	5 (1%)	67 77

All (5) residues with a non-rotameric sidechain are listed below:

Mol Chain	Res	Type
1 A	185	ILE
1 A	213	ARG
1 A	278	ARG
2 B	75	VAL
2 B	136	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol Chain	Res	Type
2 B	173	HIS

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(A ²)	Q<0.9
1	A	278/298 (93%)	-0.24	1 (0%)	92 93	17, 30, 45, 61
2	B	186/195 (95%)	0.05	7 (3%)	40 37	20, 37, 53, 83
All	All	464/493 (94%)	-0.13	8 (1%)	70 67	17, 32, 50, 83

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	122	SER	4.1
2	B	15	LYS	3.3
2	B	16	LEU	3.3
2	B	74	CYS	2.8
2	B	75	VAL	2.3
2	B	136	SER	2.2
2	B	45	SER	2.1
2	B	9	THR	2.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

(This page is intentionally left blank)

Rights and permissions



Home



Help ▾



Live Chat



Sign in



Create Account

The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters



Author: Arren Bar-Even, Elad Noor, Yonatan Savir, et al

Publication: Biochemistry

Publisher: American Chemical Society

Date: May 1, 2011

Copyright © 2011, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)

[CLOSE WINDOW](#)

Iron–Sulfur Proteins with Nonredox Functions



Author: Dennis H. Flint, Ronda M. Allen

Publication: Chemical Reviews

Publisher: American Chemical Society

Date: Jan 1, 1996

Copyright © 1996, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)

[CLOSE WINDOW](#)

IMP Dehydrogenase: Structure, Mechanism, and Inhibition

Author: Lizbeth Hedstrom

Publication: Chemical Reviews

Publisher: American Chemical Society

Date: Jul 1, 2009

Copyright © 2009, American Chemical Society



PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)

[CLOSE WINDOW](#)

(This page is intentionally left blank)

References

- Abrol, R., & Sundaresan, V. (2002). Towards a general model for protein-substrate stereoselectivity. *Protein Science*, *11*(6), 1330–1339.
- Adam, J., Yang, M., Bauerschmidt, C., Kitagawa, M., O’Flaherty, L., Maheswaran, P., Özkan, G., Sahgal, N., Baban, D., Kato, K., Saito, K., Iino, K., Igarashi, K., Stratford, M., Pugh, C., Tennant, D. A., Ludwig, C., Davies, B., Ratcliffe, P. J., Pollard, P. J. (2013). A Role for Cytosolic Fumarate Hydratase in Urea Cycle Metabolism and Renal Neoplasia. *Cell Reports*, *3*(5), 1440–1448.
- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, *66*(2), 213–221.
- Adamu, A., Aliyu, F., Huyop, F., & Wahab, R. A. (2018). Molecular Basis and Engineering of Enzymes Stereospecificity. *Journal of Molecular Biology and Methods*, *1:1*.
- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography*, *68*, 352–367.
- Agarwal, P. K., Billeter, S. R., & Hammes-Schiffer, S. (2002). Nuclear Quantum Effects and Enzyme Dynamics in Dihydrofolate Reductase Catalysis. *The journal of Physical Chemistry*, *106*(12), 3283–3293.
- Ajalla Aleixo, M. A., Rangel, V. L., Rustiguel, J. K., de Pádua, R. A. P., & Nonato, M. C. (2019). Structural, biochemical and biophysical characterization of recombinant human fumarate hydratase. *FEBS Journal*, *286*(10), 1925–1940.
- Alberti, S., Halfmann, R., King, O., Kapila, A., & Lindquist, S. (2009). A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell*, *137*(1), 146–158.
- Alexandre, T., Rayna, B., & Munier-Lehmann, H. (2015). Two classes of bacterial IMPDHs according to their quaternary structures and catalytic properties. *PLoS ONE*, *10*(2), 1–17.
- Amyes, T. L., Wood, B. M., Chan, K., Gerlt, J. A., & Richard, J. P. (2008). Formation and Stability of a Vinyl Carbanion at the Active Site of Orotidine 5'-Monophosphate Decarboxylase: pK_a of the C-6 Proton of Enzyme-Bound UMP. *Journal of the American Chemical Society*, *130*, 1574–1575.

-
- An, S., Kumar, R., Sheets, E. D., & Benkovic, S. J. (2008). Reversible compartmentalization of de novo purine biosynthetic complexes in living cells. *Science*, *320*(5872), 103–106.
- Anthony, S. A., Burrell, A. L., Johnson, M. C., Duong-Ly, K. C., Kuo, Y.-M., Simonet, J. C., Michener, P., Andrews, A., Kollman, J. M., & Peterson, J. R. (2017). Reconstituted IMPDH polymers accommodate both catalytically active and inactive conformations. *Molecular Biology of the Cell*, *28*(20), 2600–2608.
- Aughey, G. N., & Liu, J. L. (2016). Metabolic regulation via enzyme filamentation. *Critical Reviews in Biochemistry and Molecular Biology*, *51*(4), 282.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*, 2006-2008.
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, *98*(18), 10037-10041.
- Bar-even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Taw, D. S., & Milo, R. (2011). The Moderately Efficient Enzyme: Evolutionary and Physicochemical. *Biochemistry*, *50*(21), 4402–4410.
- Barry, R. M., Bitbol, A. F., Lorestani, A., Charles, E. J., Habrian, C. H., Hansen, J. M., Li, H. J., Baldwin, E. P., Wingreen, N. S., Kollman, J. M., & Gitai, Z. (2014). Large-scale filament formation inhibits the activity of CTP synthetase. *ELife*, *3*, 1–19.
- Bartsch, S., Kourist, R., & Bornscheuer, U. T. (2008). Complete inversion of enantioselectivity towards acetylated tertiary alcohols by a double mutant of a Bacillus subtilis esterase. *Angewandte Chemie - International Edition*, *47*(8), 1508 - 1511.
- Bateman, A. (1997). The structure of a domain common to archaeobacteria and the homocystinuria disease protein. *Trends in Biochemical Sciences*, *22*(1), 12–13.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., & Leslie, A. G. W. (2011). iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography*, *67*, 271-281.
- Baykov, A. A., Tuominen, H. K., & Lahti, R. (2011). The CBS domain: A protein module with an emerging prominent role in regulation. In *ACS Chemical Biology*, *6*(11), 1156-1163.
- Bearne, S. L., & Wolfenden, R. (1995). Enzymatic Hydration of an Olefin: The Burden Borne by Fumarase. *Journal of the American Chemical Society*, *117*(37), 9588–9589.
- Bearne, S. L., & Wolfenden, R. (1997). Mandelate racemase in pieces: Effective concentrations of enzyme functional groups in the transition state. *Biochemistry*, *36*(7), 1646–1656.
-

-
- Beatty, N. B., & Lane, M. D. (1985). Kinetics of Citrate-induced Activation and Polymerization of Chick Liver Acetyl-CoA Carboxylase. *Annals of the New York Academy of Sciences*, 447(1), 23–37.
- Beinert, H., Kennedy, M. C., & Stout, C. D. (1996). Aconitase as Iron–Sulfur Protein, Enzyme, and Iron-Regulatory Protein. *Chemical Reviews*, 96(7), 2335–2374.
- Blanchard, J. S., & Cleland, W. W. (1980). Use of Isotope Effects To Deduce the Chemical Mechanism of Fumarase. *Biochemistry*, 19(19), 4506–4513.
- Borgis, D., & Hynes, J. T. (1993). Dynamical theory of proton tunneling transfer rates in solution: general formulation. *Chemical Physics*, 170(3), 315–346.
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(2), 248–254.
- Britton, H. G. (1973). Methods of determining rate constants in single-substrate–single-product enzyme reactions. Use of induced transport: limitations of product inhibition. *Biochemical Journal*, 133(2), 255–261.
- Buey, R. M., Fernández-Justel, D., Marcos-Alcalde, Í., Winter, G., Gómez-Puertas, P., De Pereda, J. M., & Luis Revuelta, J. (2017). A nucleotide-controlled conformational switch modulates the activity of eukaryotic IMP dehydrogenases. *Scientific Reports*, 7(1), 1–12.
- Buey, R. M., Ledesma-Amaro, R., Velázquez-Campoy, A., Balsera, M., Chagoyen, M., De Pereda, J. M., & Revuelta, J. L. (2015). Guanine nucleotide binding to the Bateman domain mediates the allosteric inhibition of eukaryotic IMP dehydrogenases. *Nature Communications*, 6, 8923.
- Bulusu, V., Srinivasan, B., Bopanna, M. P., & Balaram, H. (2009). Elucidation of the substrate specificity, kinetic and catalytic mechanism of adenylosuccinate lyase from *Plasmodium falciparum*. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1794(4), 642–654.
- Burrell, A. L., Nie, C., Said, M., Simonet, J. C., Fernández-Justel, D., Johnson, M. C., Quispe, J., Buey, R. M., Peterson, J. R., & Kollman, J. M. (2022). IMPDH1 retinal variants control filament architecture to tune allosteric regulation. *Nature Structural & Molecular Biology*, 29(1), 47–58.
- Calise, S. J., Carcamo, W. C., Krueger, C., Yin, J. D., Purich, D. L., & Chan, E. K. L. (2014). Glutamine deprivation initiates reversible assembly of mammalian rods and rings. *Cellular and Molecular Life Sciences : CMLS*, 71(15), 2963–2973.
- Carcamo, W. C., Calise, S. J., von Mühlen, C. A., Satoh, M., & Chan, E. K. L. (2014). Molecular cell biology and immunobiology of mammalian rod/ring structures. *International Review of Cell and Molecular Biology*, 308, 35–74.
-

-
- Carcamo, W. C., Satoh, M., Kasahara, H., Terada, N., Hamazaki, T., Chan, J. Y. F., Yao, B., Tamayo, S., Covini, G., von Mühlen, C. A., & Chan, E. K. L. (2011). Induction of Cytoplasmic Rods and Rings Structures by Inhibition of the CTP and GTP Synthetic Pathway in Mammalian Cells. *PLOS ONE*, 6(12), 29690.
- Carlton, J. M., Hirt, R. P., Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., Bidwell, S. L., Alsmark, U. C. M., Besteiro, S., Sicheritz-Ponten, T., Noel, C. J., Dacks, J. B., Foster, P. G., Simillion, C., Van De Peer, Y., Miranda-Saavedra, D., Barton, G. J., Westrop, G. D., Johnson, P. J. (2007). Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science*, 315(5809), 207–212.
- Cer, R. Z., Mudunuri, U., Stephens, R., & Lebeda, F. J. (2009). IC50-to-Ki: A web-based tool for converting IC50 to Ki values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Research*, 37, 441–445.
- Chakravorty, D. K., & Hammes-Schiffer, S. (2010). Impact of Mutation on Proton Transfer Reactions in Ketosteroid Isomerase: Insights from Molecular Dynamics Simulations. *Journal of the American Chemical Society*, 132(21), 7549–7555.
- Chakravorty, D. K., Soudackov, A. V., & Hammes-Schiffer, S. (2009). Hybrid quantum/classical molecular dynamics simulations of the proton transfer reactions catalyzed by ketosteroid isomerase: Analysis of hydrogen bonding, conformational motions, and electrostatics. *Biochemistry*, 48(44), 10608–10619.
- Chan, H. C. S., Pan, L., Li, Y., & Yuan, S. (2019). Rationalization of stereoselectivity in enzyme reactions. *Computational Molecular Science*, 9(4), 1403.
- Chan, H. C. S., Wang, J., Palczewski, K., Filipek, S., Vogel, H., Liu, Z. J., & Yuan, S. (2018). Exploring a new ligand binding site of G protein-coupled receptors. *Chemical Science*, 9(31), 6480–6489.
- Chang, C.-C., Lin, W.-C., Pai, L.-M., Lee, H.-S., Wu, S.-C., Ding, S.-T., Liu, J.-L., & Sung, L.-Y. (2015). Cytoophidium assembly reflects upregulation of IMPDH activity. *Journal of Cell Science*, 128(19), 3550–3555.
- Chang, C. H., & Kim, K. (2009). Density functional theory calculation of bonding and charge parameters for molecular dynamics studies on [FeFe] hydrogenases. *Journal of Chemical Theory and Computation*, 5(4), 1137–1145.
- Chayen, N. E., Shaw Stewart, P. D., & Blow, D. M. (1992). Microbatch crystallization under oil - a new technique allowing many small-volume crystallization trials. *Journal of Crystal Growth*, 122(1–4), 176–180.
- Cherrier, M. V., Chan, A., Darnault, C., Reichmann, D., Amara, P., Ollagnier De Choudens, S., & Fontecilla-Camps, J. C. (2014). The crystal structure of Fe4S4 quinolinate synthase unravels
-

-
- an enzymatic dehydration mechanism that uses tyrosine and a hydrolase-type triad. *Journal of the American Chemical Society*, 136(14), 5253-5256.
- Cygler, M., Schrag, J. D., Bouthillier, F., Grochulski, P., Kazlauskas, R. J., Serreq, A. N., Gupta, A. K., & Rubin, B. (1994). A Structural Basis for the Chiral Preferences of Lipases. *Journal of the American Chemical Society*, 116(8), 3180–3186.
- Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), 6640-6645.
- De Pádua, R. A. P., Kia, A. M., Costa-Filho, A. J., Wilkinson, S. R., & Nonato, M. C. (2017). Characterisation of the fumarate hydratase repertoire in *Trypanosoma cruzi*. *International Journal of Biological Macromolecules*, 102, 42–51.
- Dik, E., Naamati, A., Asraf, H., Lehming, N., & Pines, O. (2016). Human Fumarate Hydratase Is Dual Localized by an Alternative Transcription Initiation Mechanism. *Traffic*, 17(7), 720–732.
- Doublié, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods in Enzymology*. 276, 523-530.
- Duong-Ly, K. C., Kuo, Y. M., Johnson, M. C., Cote, J. M., Kollman, J. M., Soboloff, J., Rall, G. F., Andrews, A. J., & Peterson, J. R. (2018). T cell activation triggers reversible inosine-5'-monophosphate dehydrogenase assembly. *Journal of Cell Science*, 131(17).
- Easson, L. H., & Stedman, E. (1933). Studies on the relationship between chemical constitution and physiological action. *Biochemical Journal*, 27(4), 1257–1266.
- Ema, T., Fujii, T., Ozaki, M., Korenaga, T., & Sakai, T. (2005). Rational control of enantioselectivity of lipase by site-directed mutagenesis based on the mechanism. *Chemical Communications*. 37, 4650-4651.
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(D), 486-501.
- Eprintsev, A. T., Fedorin, D. N., Sazonova, O. V., & Igamberdiev, A. U. (2018). Expression and properties of the mitochondrial and cytosolic forms of fumarase in sunflower cotyledons. *Plant Physiology and Biochemistry*, 129, 305–309.
- Ereño-Orbea, J., Oyenarte, I., & Martínez-Cruz, L. A. (2013). CBS domains: Ligand binding sites and conformational variability. *Archives of Biochemistry and Biophysics*, 540(2), 70-81.
- Erickson, B. L. E & Alberty, R. A. (1959). Kinetics and mechanism of the base-catalyzed hydration of fumarate to malate. *Journal of Physical Chemistry*, 63, 705-709.
- Feliciano, P. R., & Drennan, C. L. (2019). Structural and Biochemical Investigations of the [4Fe-
-

-
- 4S] Cluster-Containing Fumarate Hydratase from *Leishmania major*. *Biochemistry*, 58(49), 5011–5021.
- Feliciano, P. R., Drennan, C. L., & Nonato, M. C. (2016). Crystal structure of an Fe-S cluster-containing fumarate hydratase enzyme from *Leishmania major* reveals a unique protein fold. *Proceedings of the National Academy of Sciences*, 113(35), 9804–9809.
- Feliciano, P. R., Drennan, C. L., & Nonato, M. C. (2019). Crystal structures of fumarate hydratases from *Leishmania major* in a complex with inhibitor 2-thiomalate Crystal structures of fumarate hydratases from *Leishmania major* in a complex with inhibitor 2-thiomalate. *ACS Chemical Biology*. 14(2), 266-275.
- Feliciano, P. R., Gupta, S., Dyszy, F., Dias-Baruffi, M., Costa-Filho, A. J., Michels, P. A. M., & Nonato, M. C. (2012). Fumarate hydratase isoforms of *Leishmania major*: Subcellular localization, structural and kinetic properties. *International Journal of Biological Macromolecules*, 51(1–2), 25–31.
- Felsenstein, J. (1985). Confidence limits on phylogenies- an approach using the bootstrap. *Evolution*, 5(2), 113-129.
- Fernández-Justel, D., Núñez, R., Martín-Benito, J., Jimeno, D., González-López, A., Soriano, E. M., Revuelta, J. L., & Buey, R. M. (2019). A Nucleotide-Dependent Conformational Switch Controls the Polymerization of Human IMP Dehydrogenases to Modulate their Catalytic Activity. *Journal of Molecular Biology*, 431(5), 956–969.
- Flint, D. H. (1993). *Escherichia coli* Fumarase A Catalyzes the Isomerization of Enol and Keto Oxalacetic Acid. *Biochemistry*, 32(3), 799–805.
- Flint, D. H. (1994). Initial kinetic and mechanistic characterization of *escherichia coli* fumarase A. *Archives of Biochemistry and Biophysics*, 311(2), 509-516.
- Flint, D. H., & Allen, R. M. (1996). Iron-sulfur proteins with nonredox functions. In *Chemical Reviews*, 96(7), 2315–2334.
- Flint, D. H., Emptage, M. H., & Guest, J. R. (1992). Fumarase A from *Escherichia coli*: Purification and Characterization as an Iron-Sulfur Cluster Containing Enzyme. *Biochemistry*, 31(42), 10331–10337.
- Fox, N. K., Brenner, S. E., & Chandonia, J. M. (2014). SCOPe: Structural Classification of Proteins - Extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Research*, 42(D1), 304-309.
- Fried, S. D., & Bagchi, S. (2014). Extreme electric fields power catalysis in the active site of ketosteroid isomerase. *Science*, 346(6216), 1510-1514.
- Gajewski, E., Goldberg, R.N & Steckler, D.K. (1985). Thermodynamics of the conversion of
-

-
- fumarate to L-(-)-malate. *Biophysical Chemistry*, 22, 187-195.
- Genda, T., Watabe, S., & Ozaki, H. (2006). Purification and characterization of fumarase from *Corynebacterium glutamicum*. *Bioscience, Biotechnology and Biochemistry*, 70(5), 1102–1109.
- Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, 10(5), 449.
- Goldberg, R. N., & Tewari, Y. B. (1995). Thermodynamics of Enzyme-Catalyzed Reactions: Part 4. Lyases. *Journal of Physical and Chemical Reference Data*, 24(5), 1669–1698.
- Grisewood, M. J., Hernández-Lozada, N. J., Thoden, J. B., Gifford, N. P., Mendez-Perez, D., Schoenberger, H. A., Allan, M. F., Floy, M. E., Lai, R. Y., Holden, H. M., Pflieger, B. F., & Maranas, C. D. (2017). Computational Redesign of Acyl-ACP Thioesterase with Improved Selectivity toward Medium-Chain-Length Fatty Acids. *ACS Catalysis*, 7(6), 3837-3849.
- Haldane, J. (2003). *Enzymes* (p. 235). MIT Press.
- Hammer, S. C., Knight, A. M., & Arnold, F. H. (2017). Design and evolution of enzymes for non-natural chemistry. *Current Opinion in Green and Sustainable Chemistry*, 7, 23–30.
- Hammes-Schiffer, S. (2012). Catalytic Efficiency of Enzymes: A Theoretical Analysis. *Biochemistry*, 52(12), 2012–2020.
- Hammes-Schiffer, S., & Benkovic, S. J. (2006). Relating Protein Motion to Catalysis. *Annual Review of Biochemistry*, 75, 519-541.
- Hanoian, P., Sigala, P. A., Herschlag, D., & Hammes-Schiffer, S. (2010). Hydrogen Bonding in the Active Site of Ketosteroid Isomerase: Electronic Inductive Effects and Hydrogen Bond Coupling. *Biochemistry*, 49(48), 10339-10348.
- Hansen, J. N., Dinovo, E. C., & Boyer, P. D. (1969). Initial and equilibrium ¹⁸O, ¹⁴C, ³H, and ²H exchange rates as probes of the fumarase reaction mechanism. *Journal of Biological Chemistry*, 244(22), 6270–6279.
- Harris, T. K., & Turner, G. J. (2002). Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites. *IUBMB Life*, 53(2), 85–98.
- Hedstrom, L. (2009). IMP Dehydrogenase: Structure, Mechanism, and Inhibition. *Chemical Reviews*. 109(7), 2903-2928.
- Hedstrom, L. (2010). *Enzyme Specificity and Selectivity*. John Wiley & Sons Ltd.
- Ignoul, S., & Eggermont, J. (2005). CBS domains: structure, function, and pathology in human proteins. *American Journal of Physiology-Cell Physiology*, 189(6), 1369-1378.
- Ivancic, M., Valinger, G., Gruber, K., & Schwab, H. (2007). Inverting enantioselectivity of
-

-
- Burkholderia gladioli esterase EstB by directed and designed evolution. *Journal of Biotechnology*, 129(1), 109-122.
- Jayaraman, V., Suryavanshi, A., Kalale, P., Kunala, J., & Balaram, H. (2018). Biochemical characterization and essentiality of Plasmodium fumarate hydratase. In *Journal of Biological Chemistry*, 293(16), 5878-5894.
- Jencks, W. (1987). *Catalysis in chemistry and enzymology*. 2nd edition. Dover Publications Inc.
- Ji, Y., Gu, J., Makhov, A. M., Griffith, J. D., & Mitchell, B. S. (2006). Regulation of the interaction of inosine monophosphate dehydrogenase with mycophenolic acid by GTP. *Journal of Biological Chemistry*, 281(1), 206–212.
- Jiang, Y., Qian, X., Shen, J., Wang, Y., Li, X., Liu, R., Xia, Y., Chen, Q., Peng, G., Lin, S. Y., & Lu, Z. (2018). Author Correction: Local generation of fumarate promotes DNA repair through inhibition of histone H3 demethylation. *Nature Cell Biology*, 17(9), 1.
- Jin, J., & Hanefeld, U. (2011). The selective addition of water to C≡C bonds; enzymes are the best chemists. *Chemical Communications*, 47(9), 2502–2510.
- Johnson, M. C., & Kollman, J. M. (2020). Cryo-EM structures demonstrate human IMPDH2 filament assembly tunes allosteric regulation. *ELife*, 9, 53243.
- Jolla, L. (1996). Expression and characterization of the flavoprotein subcomplex composed of 50-kDa (NQO1) and 25-kDa (NQO2) subunits of the proton-translocating NADH-quinone oxidoreductase of *Paracoccus denitrificans*. *Journal of Biological Chemistry*, 271(10), 5907–5913.
- Jorgensen, W. L., & Thomas, L. L. (2008). Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *Journal of Chemical Theory and Computation*, 4(6), 869–876.
- Kelly, J. M., & Scopes, R. K. (1986). L-(+)-Tartrate dehydratase from *Pseudomonas putida* is an iron-sulphur enzyme. *FEBS Letters*, 202(2), 274-276.
- Keppeke, G. D., Calise, S. J., Chan, E. K. L., & Andrade, L. E. C. (2015). Assembly of IMPDH2-based, CTPS-based, and mixed rod/ring structures is dependent on cell type and conditions of induction. *Journal of Genetics and Genomics*, 42(6), 287–299.
- Keppeke, G. D., Chang, C. C., Peng, M., Chen, L. Y., Lin, W. C., Pai, L. M., Andrade, L. E. C., Sung, L. Y., & Liu, J. L. (2018). IMP/GTP balance modulates cytoophidium assembly and IMPDH activity. *Cell Division*, 13(1), 1–18.
- Khersonsky, O., & Tawfik, D. S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual Review of Biochemistry*, 79, 471–505.
- Kim, O. Bin, Lux, S., & Uden, G. (2007). Anaerobic growth of *Escherichia coli* on D-tartrate depends on the fumarate carrier DcuB and fumarase, rather than the L-tartrate carrier TtdT
-

-
- and L-tartrate dehydratase. *Archives of Microbiology*, 188(6), 583–589.
- Klesmith, J. R., Bacik, J. P., Michalczyk, R., & Whitehead, T. A. (2015). Comprehensive Sequence-Flux Mapping of a Levoglucosan Utilization Pathway in *E. coli*. *ACS Synthetic Biology*, 4(11), 1235–1243.
- Klinman, J. P., & Kohen, A. (2014). Evolutionary aspects of enzyme dynamics. *Journal of Biological Chemistry*, 289(44), 30205–30212.
- Knox, C., Sass, E., Neupert, W., & Pines, O. (1998). Import into mitochondria, folding and retrograde movement of fumarase in yeast. *Journal of Biological Chemistry*, 273(40), 25587–25593.
- Koga, Y., Kato, K., Nakano, H., & Yamane, T. (2003). Inverting enantioselectivity of Burkholderia cepacia KWI-56 lipase by combinatorial mutation and high-throughput screening using single-molecule PCR and in vitro expression. *Journal of Molecular Biology*, 331(3), 585–592.
- Krebs, H. A. (1953). The equilibrium constants of the fumarase and aconitase systems. *The Biochemical Journal*, 54(1), 78–82.
- Kronen, M., & Berg, I. A. (2015). Meseconase/fumarase FumD in *Escherichia coli* O157:H7 and promiscuity of *Escherichia coli* class I fumarases FumA and FumB. *PLoS ONE*, 10(12), 1–18.
- Kronen, M., Sasikaran, J., & Berg, I. A. (2015). Meseconase activity of class I fumarase contributes to meseconate utilization by *Burkholderia xenovorans*. *Applied and Environmental Microbiology*, 81(16), 5632–5638.
- Kumar, S., Stecher, G., Tamura, K., & Dudley, J. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets Downloaded from. *Molecular Biology and Evolution*, 33(7), 1870–1874.
- Labesse, G., Alexandre, T., Vaupré, L., Salard-Arnaud, I., Him, J. L. K., Raynal, B., Bron, P., & Munier-Lehmann, H. (2013). MgATP regulates allostery and fiber formation in IMPDHs. *Structure*, 21(6), 975–985.
- Labourel, F., & Rajon, E. (2021). Resource Uptake and the Evolution of Moderately Efficient Enzymes. *Molecular Biology and Evolution*, 38(9), 3938–3952.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227, 680–685.
- Lamzin, V. S., Dauter, Z., & Wilson, K. S. (1994). Dehydrogenation through the looking-glass. *Nature Structural Biology*, 1(5), 281–282.
- Lamzin, V. S., Dauter, Z., & Wilson, K. S. (1995). How nature deals with stereoisomers. *Current*
-

-
- Opinion in Structural Biology*, 5(6), 830–836.
- Lane, A. N., & Fan, T. W. M. (2015). Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Research*, 43(4), 2466–2485.
- Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. *Journal of Chemical Information and Modeling*, 51(10), 2778–2786.
- Lauble, H., Kerinedy, M. C., Beinert, H., & Stout, C. D. (1992). Crystal structures of aconitase with isocitrate and nitroisocitrate bound. *Biochemistry*, 31(10), 2735–2748.
- Lee, Y. K. (2021). Density Functional Theory (DFT) Calculations and Catalysis. *Catalysts*, 11(4), 454.
- Li, G., & Reetz, M. T. (2016). Learning lessons from directed evolution of stereoselective enzymes. *Organic Chemistry Frontiers*, 3(10), 1350–1358.
- Liu, J. L. (2016). The Cytoophidium and Its Kind: Filamentation and Compartmentation of Metabolic Enzymes. *Annurev-Cellbio*, 32, 349–372.
- Lloyd, S. J., Lauble, H., Prasad, G. S., & Stout, C. D. (2008). The mechanism of aconitase: 1.8 Å resolution crystal structure of the S642A:citrate complex. *Protein Science*, 8(12), 2655–2662.
- López-Blanco, J. R., Aliaga, J. I., Quintana-Ortí, E. S., & Chacón, P. (2014). IMODS: Internal coordinates normal mode analysis server. *Nucleic Acids Research*, 42(W1), 271–276.
- Lynch, E. M., Hicks, D. R., Shepherd, M., Endrizzi, J. A., Maker, A., Hansen, J. M., Barry, R. M., Gitai, Z., Baldwin, E. P., & Kollman, J. M. (2017). Human CTP synthase filament structure reveals the active enzyme conformation. *Nature Structural and Molecular Biology*, 24(6), 507–514.
- Martin, W. F. (2011). Early evolution without a tree of life. *Biology Direct*, 6(1), 1–25.
- McCarter, J. D., & Stephen Withers, G. (1994). Mechanisms of enzymatic glycoside hydrolysis. *Current Opinion in Structural Biology*, 4(6), 885–892.
- McCoy, A. J. (2006). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallographica Section D: Biological Crystallography*, 63(1), 32–41.
- Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B., Adam, R. D., Olsen, G. J., Best, A. A., Cande, W. Z., Chen, F., Cipriano, M. J., Davids, B. J., Dawson, S. C., Elmendorf, H. G., Hehl, A. B., Holder, M. E., Huse, S. M., Kim, U. U., Lasek-Nesselquist, E., Manning, G., Sogin, M. L. (2007). Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science*, 317(5846), 1921–1926.
- Morrow, C. A., Valkov, E., Stamp, A., Chow, E. W. L., Lee, I. R., Wronski, A., Williams, S. J.,
-

-
- Hill, J. M., Djordjevic, J. T., Kappler, U., Kobe, B., & Fraser, J. A. (2012). De novo GTP Biosynthesis Is Critical for Virulence of the Fungal Pathogen *Cryptococcus neoformans*. *PLOS Pathogens*, *8*(10), 1002957.
- Mu, R., Wang, Z., Wamsley, M. C., Duke, C. N., Lii, P. H., Epley, S. E., Todd, L. C., & Roberts, P. J. (2020). Application of Enzymes in Regioselective and Stereoselective Organic Reactions. *Catalysts*, *10*(8), 832.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, *67*(D), 355-367.
- Narayanaswamy, R., Levy, M., Tsechansky, M., Stovall, G. M., O'Connell, J. D., Mirrielees, J., Ellington, A. D., & Marcotte, E. M. (2009). Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(25), 10147–10152.
- Newton, M. S., Arcus, V. L., Gerth, M. L., & Patrick, W. M. (2018). Enzyme evolution: innovation is easy, optimization is complicated. *Current Opinion in Structural Biology*, *48*, 110–116.
- Nimmegern, E., Black, J., Futer, O., Fulghum, J. R., Chambers, S. P., Brummel, C. L., Raybuck, S. A., & Sintchak, M. D. (1999). Biochemical analysis of the modular enzyme inosine 5'-monophosphate dehydrogenase. *Protein Expression and Purification*, *17*(2), 282–289.
- Noree, C., Sato, B. K., Broyer, R. M., & Wilhelm, J. E. (2010). Identification of novel filament-forming proteins in *Saccharomyces cerevisiae* and *Drosophila melanogaster*. *The Journal of Cell Biology*, *190*(4), 541–551.
- Nunes-Nesi, A., Fernie, A. R., & Stitt, M. (2010). Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Molecular Plant*, *3*(6), 973–996.
- O'Connell, J. D., Tsechansky, M., Royall, A., Boutz, D. R., Ellington, A. D., & Marcotte, E. M. (2014). A proteomic survey of widespread protein aggregation in yeast. *Molecular BioSystems*, *10*(4), 851.
- O'Connell, J. D., Zhao, A., Ellington, A. D., & Marcotte, E. M. (2012). Dynamic Reorganization of Metabolic Enzymes into Intracellular Bodies. *Annual Review of Cell and Developmental Biology*, *28*, 89-111.
- Ogston, A. G. (1948). Interpretation of experiments on metabolic processes, using isotopic tracer elements. *Nature*, *162*(4129), 963.
- Ovchinnikov, S., Kinch, L., Park, H., Liao, Y., Pei, J., Kim, D. E., Kamisetty, H., Grishin, N. V., & Baker, D. (2015). Large-scale determination of previously unsolved protein structures using evolutionary information. *ELife*, *4*, 1–25.
-

- Pasteur, Louis (1864). *Mémoire sur la fermentation acétique. Annales scientifiques de l'É.N.S. Ire série, tome 1*, 113-158
- Patel, R., & Patel, R. (2010). Stereoselective Biocatalysis for Synthesis of Some Chiral Pharmaceutical Intermediates. *Stereoselective Biocatalysis*, 1st edition, CRC Press.
- Pauling, L. (1946). Molecular architecture and biological reactions. *Chemical and Engineering News*, 24(10), 1375–1377.
- Pauling, L. (1948). Nature of forces between large molecules of biological interest. *Nature*, 161(4097), 707–709.
- Petrovska, I., Nüske, E., Munder, M. C., Kulasegaran, G., Malinowska, L., Kroschwald, S., Richter, D., Fahmy, K., Gibson, K., Verbavatz, J. M., & Alberti, S. (2014). Filament formation by metabolic enzymes is a specific adaptation to an advanced state of cellular starvation. *ELife*, 3(3).
- Pettersson, G. (1989). Effect of evolution on the kinetic properties of enzymes. *European Journal of Biochemistry*, 184(3), 561–566.
- Porter, D. J., & Bright, H. J. (1980). 3-Carbanionic substrate analogues bind very tightly to fumarase and aspartase. *Journal of Biological Chemistry*, 255(10), 4772–4780.
- Pracharoenwattana, I., Zhou, W., Keech, O., Francisco, P. B., Udomchalothorn, T., Tschoep, H., Stitt, M., Gibon, Y., & Smith, S. M. (2010). Arabidopsis has a cytosolic fumarase required for the massive allocation of photosynthate into fumaric acid and for rapid plant growth on high nitrogen. *Plant Journal*, 62(5), 785–795.
- Prosise, G. L., Zhen Wu, J., & Luecke, H. (2002). Crystal structure of Tritrichomonas foetus inosine monophosphate dehydrogenase in complex with the inhibitor ribavirin monophosphate reveals a catalysis-dependent ion-binding site. *The Journal of Biological Chemistry*, 277(52), 50654–50659.
- Puthan Veetil, V., Fibriansah, G., Raj, H., Thunnissen, A. M. W. H., & Poelarends, G. J. (2012). Aspartase/fumarase superfamily: A common catalytic strategy involving general base-catalyzed formation of a highly stabilized aci - Carboxylate intermediate. *Biochemistry*, 51(21), 4237–4243.
- Rao, V. A., Shepherd, S. M., Owen, R., & Hunter, W. N. (2013). Structure of Pseudomonas aeruginosa inosine 5'-monophosphate dehydrogenase. *Acta Crystallographica Section F: Structural biology*, 69(3), 243–247.
- Reaney, S. K., Begg, C., Bungard, S. J., & Guest, J. R. (2009). Identification of the L-tartrate dehydratase genes (ttdA and ttdB) of Escherichia coli and evolutionary relationship with the Class I fumarase genes. *Journal of General Microbiology*, 139(7), 1523-1530.

-
- Reaney, S. K., Bungard, S. J., & Guest, J. R. (2009). Molecular and Enzymological Evidence for Two Classes of Fumarase in *Bacillus Stearotherophilus* (var. Non-Diastaticus). *Journal of General Microbiology*, *139*(3), 403-416.
- Reed, A. E., Weinstock, R. B., & Weinhold, F. (1998). Natural population analysis. *The Journal of Chemical Physics*, *83*(2), 735.
- Reetz, M. T. (2011). Laboratory evolution of stereoselective enzymes: A prolific source of catalysts for asymmetric reactions. *Angewandte Chemie*, *50*(1), 138–174.
- Reetz, M. T., Wilensek, S., Zha, D., & Jaeger, K. E. (2001). Directed evolution of an enantioselective enzyme through combinatorial multiple-cassette mutagenesis. *Angewandte Chemie*, *40*(19), 3589-3591.
- Riera, T. V., Zheng, L., Josephine, H. R., Min, D., Yang, W., & Hedstrom, L. (2011). Allosteric activation via kinetic control: Potassium accelerates a conformational change in IMP dehydrogenase. *Biochemistry*, *50*(39), 8508–8518.
- Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, *42*(1), 320-324.
- Robinson, P.J, Holbrook, K. A. (1972). On adsorption catalysis. *Z Elektrochem*, *27*, 142–150.
- Rose, I. A. (1998). How fumarase recycles after the malate - Fumarate reaction. Insights into the reaction mechanism. *Biochemistry*, *37*(51), 17651–17658.
- Rose, I. A., Warms, J. V. B., & Yuan, R. G. (1993). Role of Conformational Change in the Fumarase Reaction Cycle. *Biochemistry*, *32*(33), 8504–8511.
- Rose, I. A., & Weaver, T. M. (2004). The role of the allosteric B site in the fumarase reaction. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(10), 3393–3397.
- Rotticci, D., Rotticci-Mulder, J. C., Denman, S., Norin, T., & Hult, K. (2001). Improved enantioselectivity of a lipase by rational protein engineering. *ChemBioChem*, *2*(10), 766-770.
- Rui, L., Cao, L., Chen, W., Reardon, K. F., & Wood, T. K. (2005). Protein engineering of epoxide hydrolase from *Agrobacterium radiobacter* AD1 for enhanced activity and enantioselective production of (R)-1-phenylethane-1, 2-diol. *Applied and Environmental Microbiology*, *71*, 7.
- Ruiter, A. De, & Oostenbrink, C. (2016). Extended Thermodynamic Integration: Efficient Prediction of Lambda Derivatives at Nonsimulated Points. *Journal of Chemical Theory and Computation*, *12*(9), 4476–4486.
- Sambrook, J., & Russell, D. W. (2001). *Molecular Cloning - Sambrook & Russel - Vol. 1, 2, 3. CSH Press.*
-

-
- Saunders, A. H., Griffiths, A. E., Lee, K. H., Cicchillo, R. M., Tu, L., Stromberg, J. A., Krebs, C., & Booker, S. J. (2008). Characterization of quinolinate synthases from *Escherichia coli*, *Mycobacterium tuberculosis*, and *Pyrococcus horikoshii* indicates that [4Fe-4S] clusters are common cofactors throughout this class of enzymes. *Biochemistry*, *47*(41), 10999-11012.
- Schink, B. (1984). Fermentation of tartrate enantiomers by anaerobic bacteria, and description of two new species of strict anaerobes, *Ruminococcus pasteurii* and *Ilyobacter tartaricus*. *Archives of Microbiology*, *139*(4), 409–414.
- Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., Scheer, M., & Schomburg, D. (2013). BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Research*, *41*(D), 764-772.
- Schramm, V. L. (1998). Enzymatic transition states and transition state analog design. *Annual Review of Biochemistry*, 693–720.
- Schrödinger. (2015). The PyMOL Molecular Graphics System. *Schrödinger LLC*, *26*(5), 708-715.
- Seiffert, G. B., Ullmann, G. M., Messerschmidt, A., Schink, B., Kroneck, P. M. H., & Einsle, O. (2007). Structure of the non-redox-active tungsten/[4Fe:4S] enzyme acetylene hydratase. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(9), 3073–3077.
- Shen, Q. J., Kassim, H., Huang, Y., Li, H., Zhang, J., Li, G., Wang, P. Y., Yan, J., Ye, F., & Liu, J. L. (2016). Filamentation of Metabolic Enzymes in *Saccharomyces cerevisiae*. *Journal of Genetics and Genomics*, *43*(6), 393.
- Shen, Z., Lv, C., & Zeng, S. (2016). Significance and challenges of stereoselectivity assessing methods in drug metabolism. *Journal of Pharmaceutical Analysis*, *6*(1), 1.
- Shibata, H., Gardiner, W. E., & Schwartzbach, S. D. (1985). Purification, characterization, and immunological properties of fumarase from *Euglena gracilis* var. *bacillaris*. *Journal of Bacteriology*, *164*(2), 762–768.
- Shimoyama, T., Rajashekhara, E., Ohmori, D., Kosaka, T., & Watanabe, K. (2007). MmcBC in *Pelotomaculum thermopropionicum* represents a novel group of prokaryotic fumarases. *FEMS Microbiology Letters*, *270*(2), 207–213.
- Shu, Q., & Nair, V. (2008). Inosine monophosphate dehydrogenase (IMPDH) as a target in drug discovery. *Medicinal Research Reviews*, *28*(2), 219–232.
- Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., St.Clair, J. L., Gallaher, J. L., Hilvert, D., Gelb, M. H., Stoddard, B. L., Houk, K. N., Michael, F. E., & Baker, D. (2010). Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science*, *329*(5989), 309–313.
-

-
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 539.
- Silas, Y., Singer, E., Das, K., Lehming, N., & Pines, O. (2021). A combination of Class-I fumarases and metabolites (α -ketoglutarate and fumarate) signal the DNA damage response in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 118(23), 2026595118.
- Singh, R., Kumar, M., Mittal, A., & Mehta, P. K. (2016). Microbial enzymes: industrial progress in 21st century. *3 Biotech*, 6, 174.
- Sintchak, M. D., Fleming, M. A., Futer, O., Raybuck, S. A., Chambers, S. P., Caron, P. R., Murcko, M. A., & Wilson, K. P. (1996). Structure and mechanism of inosine monophosphate dehydrogenase in complex with the immunosuppressant mycophenolic acid. *Cell*, 85(6), 921–930.
- Srinivasan, B. (2021a). A guide to the Michaelis–Menten equation: steady state and beyond. *FEBS Journal*. <https://doi.org/10.1111/febs.16124>.
- Srinivasan, B. (2021b). Words of advice: teaching enzyme kinetics. *FEBS Journal*, 288(7), 2068–2083.
- Stryer L, Berg JM & Tymoczko JL. (2002). *Biochemistry* (5th Edition). Published by New York : *W.H. Freeman*.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28(10), 2731-2739.
- Tang, L., Van Merode, A. E. J., Spelberg, J. H. L., Fraaije, M. W., & Janssen, D. B. (2003). Steady-State Kinetics and Tryptophan Fluorescence Properties of Halohydrin Dehalogenase from *Agrobacterium radiobacter*. Roles of W139 and W249 in the Active Site and Halide-Induced Conformational Change. *Biochemistry*, 42(47), 14057-14065.
- Teipel, J. W., Hass, G. M., & Hill, R. L. (1968). The substrate specificity of fumarase. *Journal of Biological Chemistry*, 243(21), 5684-5694.
- Terao, Y., Ijima, Y., Miyamoto, K., & Ohta, H. (2007). Inversion of enantioselectivity of arylmalonate decarboxylase via site-directed mutation based on the proposed reaction mechanism. *Journal of Molecular Catalysis B: Enzymatic*, 45(2), 15-20.
- Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., & Adams, P. D. (2007). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section*
-

-
- D: Biological Crystallography*, 64, 61-69.
- Thoden, J. B., Holden, H. M., & Grant, G. A. (2014). Structure of L-serine dehydratase from *Legionella pneumophila*: Novel use of the C-terminal cysteine as an intrinsic competitive inhibitor. *Biochemistry*, 53(48), 7615-7624.
- Tomlinson, I. P. M., Alam, N. A., Rowan, A. J., Barclay, E., Jaeger, E. E. M., Kelsell, D., Leigh, I., Gorman, P., Lamlum, H., Rahman, S., Roylance, R. R., Olpin, S., Bevan, S., Barker, K., Hearle, N., Houlston, R. S., Kiuru, M., Lehtonen, R., Karhu, A., ... Aaltonen, L. A. (2002). Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer the multiple leiomyoma consortium. *Nature Genetics*, 30(4), 406-410.
- Tsai, M., Koo, J., Yip, P., Colman, R. F., Segall, M. L., & Howell, P. L. (2007). Substrate and Product Complexes of *Escherichia coli* Adenylosuccinate Lyase Provide New Insights into the Enzymatic Mechanism. *Journal of Molecular Biology*, 370(3), 541-554.
- Tuboi, S., Sato, M., Ono, H., Kobayashi, K., & Hiraga, K. (1986). Mechanism of synthesis and localization of mitochondrial and cytosolic fumarases in rat liver. *Advances in Enzyme Regulation*, 25(C), 461-478.
- Ulusu, N. N. (2015). Curious Cases of the Enzymes. *Journal of Medical Biochemistry*, 34(3), 271.
- Van Kuijk, B. L. M., & Stams, A. J. M. (1996). Purification and characterization of malate dehydrogenase from the syntrophic propionate-oxidizing bacterium strain MPOB. *FEMS Microbiology Letters*, 144(3), 141-144.
- Van Loo, B., Spelberg, J. H. L., Kingma, J., Sonke, T., Wubbolts, M. G., & Janssen, D. B. (2004). Directed evolution of epoxide hydrolase from *A. radiobacter* toward higher enantioselectivity by error-prone PCR and DNA shuffling. *Chemistry and Biology*, 11(7), 981-990.
- van Vugt-Lussenburg, B. M. A., van der Weel, L., Hagen, W. R., & Hagedoorn, P. L. (2009). Identification of two [4Fe-4S]-cluster-containing hydro-lyases from *Pyrococcus furiosus*. *Microbiology*, 155(9), 3015-3020.
- van Vugt-Lussenburg, B. M. A., van der Weel, L., Hagen, W. R., & Hagedoorn, P. L. (2013). Biochemical Similarities and Differences between the Catalytic [4Fe-4S] Cluster Containing Fumarases FumA and FumB from *Escherichia coli*. *PLoS ONE*, 8(2), 1-7.
- Vargesson, N. (2015). Thalidomide-induced teratogenesis: History and mechanisms. *Birth Defects Research*, 105(2), 140.
- Warshel, A., Sharma, P. K., Kato, M., Xiang, Y., Liu, H., & Olsson, M. H. M. (2006). Electrostatic basis for enzyme catalysis. *Chemical Reviews*, 106(8), 3210-3235.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R.
-

-
- M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, *67*, 235-242.
- Wolfenden, R. (2002). Analog approaches to the structure of the transition state in enzyme reactions. *Accounts of Chemical Research*, *5*(1), 10–18.
- Wolfenden, R. (2006). Degrees of difficulty of water-consuming reactions in the absence of enzymes. *Chemical Reviews*, *106*(8), 3379–3396.
- Wombacher, R., Keiper, S., Suhm, S., Serganov, A., Patel, D. J., & Jäschke, A. (2006). Control of Stereoselectivity in an Enzymatic Reaction by Backdoor Access. *Angewandte Chemie*, *45*(15), 2469.
- Woods, S. A., Schwartzbach, S. D., & Guest, J. R. (1988). Two biochemically distinct classes of fumarase in *Escherichia coli*. *Biochimica et Biophysica Acta*, *954*(C), 14–26.
- Xu, J., Cen, Y., Singh, W., Fan, J., Wu, L., Lin, X., Zhou, J., Huang, M., Reetz, M. T., & Wu, Q. (2019). Stereodivergent Protein Engineering of a Lipase To Access All Possible Stereoisomers of Chiral Esters with Two Stereocenters. *Journal of American Chemical Society*, *141*(19), 7934-7945.
- Zell, M. B., Fahnenstich, H., Maier, A., Saigo, M., Voznesenskaya, E. V., Edwards, G. E., Andreo, C., Schleifenbaum, F., Zell, C., Drincovich, M. F., & Maurino, V. G. (2010). Analysis of Arabidopsis with Highly Reduced Levels of Malate and Fumarate Sheds Light on the Role of These Organic Acids as Storage Carbon Molecules. *Plant Physiology*, *152*(3), 1251–1262.
- Zheng, L., Kennedy, M. C., Beinert, H., & Zalkin, H. (1992). Mutational analysis of active site residues in pig heart aconitase. *Journal of Biological Chemistry*, *267*(11), 7895–7903.
- Zhou, S., Xiang, H., & Liu, J. L. (2020). CTP synthase forms cytophidia in archaea. *Journal of Genetics and Genomics*, *47*(4), 213–223.