

Structural and functional studies on novel oxidoreductases specific to steroids and isoprenoids reduction

Thesis submitted to AcSIR for the award of
the degree of

DOCTOR OF PHILOSOPHY

In

Biological Sciences



By

Anandsukeerthi Sandholu

10BB13A26042

Under the guidance of

Dr. Kiran Kulkarni

CSIR-National Chemical Laboratory

Pune, India

**“We should not give up and we should not
allow the problem to defeat us...”**

.....A. P. J Abdul Kalam

Dedicated to.....

Mom and Dad

*who always stood behind me and
encouraged me to go on every adventure,
especially this one!!*



सीएसआईआर - राष्ट्रीय रासायनिक प्रयोगशाला

(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)

डॉ. होमी भाभा मार्ग, पुणे - 411 008, भारत

CSIR - NATIONAL CHEMICAL LABORATORY

(Council of Scientific & Industrial Research)

Dr. Homi Bhabha Road, Pune - 411 008, India



Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled "**Structural and functional studies on novel oxidoreductases specific to steroids and isoprenoids reduction**" submitted by **Mr. Anandsukeerthi Sandholu** to Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy in Biological Sciences**, embodies original research work under my supervision and guidance. I further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research material obtained from other sources has been duly acknowledged in the thesis. Any text, illustration, table etc., used in the thesis from other sources, have been duly cited and acknowledged.

It is also certified that this work done by the student, under my supervision, is plagiarism free.

S. Anandsukeerthi

(Student)
Anandsukeerthi Sandholu
CSIR-National Chemical Laboratory
Pune, India

Kiran Kulkarni

(Supervisor)
Dr. Kiran Kulkarni
CSIR-National Chemical Laboratory
Pune, India

Communication Channels

NCL Level DID : 2590
NCL Board No. : +91-20-25902000
EPABX : +91-20-25893300
 : +91-20-25893400



FAX

Director's Office : +91-20-25902601
COA's Office : +91-20-25902660
SPO's Office : +91-20-25902664

WEBSITE

www.ncl-india.org

Declaration of Authorship

I, **Anandsukeerthi Sandholu**, hereby declare that the thesis entitled “**Structural and functional studies on novel oxidoreductases specific to steroid and isoprenoids reduction**” submitted for the degree of Doctor of Philosophy in Biological Sciences to the Academy of Scientific & Innovative Research (AcSIR), has been carried out by me at the CSIR-National Chemical Laboratory, Pune under the supervision of Dr. Kiran Kulkarni. Research material obtained from other sources has been duly acknowledged in the thesis. The work is original and has not been submitted in part or full by me for any other degree or diploma to any other Institution or University.

Place: *pune*

Date: *26/06/19*


.....
Anandsukeerthi Sandholu

Acknowledgement

The work presented in this thesis would not have been possible without my close association with many people. I take this opportunity to extend my sincere gratitude and appreciation to all those who made this PhD thesis possible.

First and foremost, I would like to extend my sincere gratitude to my research guide Dr. Kiran Kulkarni. It has been an honor to be his first PhD student. He has taught me and trained me in how good structural biology is done. I appreciate all his contributions of time, ideas and funding to make my PhD experience productive and stimulating. The joy and enthusiasm he has for research was contagious and motivational for me, even during tough times in the PhD pursuit. I am also thankful for the excellent example he has provided as an enthusiastic scientist and teacher.

Besides my advisor, I would like to thank my Doctorial advisory committee: Dr. Mahesh Kulkarni, Dr. Jomon Joseph and Dr. Chetan Gadgil, for their insightful comments and encouragement which incited me to widen my research from various perspectives.

I gratefully acknowledge University grants commission, for providing me financial support and Director, CSIR-NCL to allow me to carry out my research in this esteemed laboratory.

I am grateful to Dr. H.V. Thulasiram from organic chemistry division, CSIR-NCL who provided reagents and GC-MS facility for my research work. I would like to express my sincere thanks to Dr. Durba Sengupta and Madhura Mohole for sharing part of our work on MD simulations. I am grateful to Prof. William L. Duax for his valuable time spent with us in ScanProsite analysis during his visit to NCL

I am extremely grateful to Dr. Ravindra Makde, Dr. Biplab Ghosh and Dr. Ashwani Kumar for their time and help during my experiments at RRCAT Indus II synchrotron, protein crystallography beamline. My sincere gratitude to Dr. Vinothkumar Kutti (NCBS, Bangalore), for his kind help in cryo-EM experiments. I am also thankful to Dr. Saikrishana Kayarat and Dr. Gayathri Pananghat (IISER PUNE) for their help in using mosquito during initial days of my PhD, Dr. Janesh Kumar and Dr. Radha Chauhan (NCCS, PUNE) for allowing me to utilize the MALS facility. I also thankful to Dr. Nishant Kumar Varshney and Dr. Gabriele Giachin for their help during my visit to ESRF synchrotron.

My heartfelt regards to my teachers because of whose teaching has made it possible for me to see this day. Because of their enlightenment I was able to reach a stage where I could write this thesis.

My present and past research group members have always provided a stimulating and fun environment to learn and work. I thank Dr. Srivatsa, Ashwini, Sharmila, Sneha, Zenia, Debopriya, Aishwarya, Gopal, Dr. Pranaya, Raveena, Disha, Simer, Meghana, Shreegauri and Krishna for being a friendly and cheerful group of colleagues. Special thanks to my friends Nalini, Appu, Bindu, Mukesh and Amit for many scientific and non-scientific discussions and for being excellent companion throughout this journey. I would also thank my friends Ananth, Sangeeta, Yashpal, Deepanjan, Vijay, Dr. Ejaj, Shakuntala, Bhagyasri, Sneha, Swamy, Dr. Sagar, Raja, Praveen, Srikanth, Naresh and Dr. Prabhakar for many unforgettable memories and making my stay memorable which I will cherish forever.

Last but not the least, a special thanks to my parents, **John Edward** and **Jyothi**, my brother Elikranth, sister Sameera, brother-in-law John Lanka and my niece Aradhya for their constant support and encouragement throughout these years. Many a times, it was only because of your unconditional belief and confidence that helped me to overcome hard times and accomplish my dreams. Thank you, Nanna and Amma for your motivation that helped and boosted my confidence each times and also for being patient and supportive. Words can't express the gratitude I feel for both of you, I dedicate this thesis to You.

Place: Pune

Anandsukeerthi Sandholu

Date: 27/06/19

Contents

Certificate	i
Declaration	ii
Acknowledgement.....	iii
Contents.....	v
List of figures	ix
List of tables	xii
List of abbreviations	xiii
List of publications	xvi
Thesis abstract	xvii
Thesis overview.....	xviii
CHAPTER 1	
1.Introduction	1
1.1 NADPH-dependent oxidoreductases.....	2
1.2 Structural characteristics of NADPH-dependent oxidoreductases.....	3
1.3 Classes of NADPH-dependent oxidoreductases	4
1.3.1 Alcohol dehydrogenases.....	5
1.4 Applications of NADPH-dependent oxidoreductases.....	10
1.5 Atypical oxidoreductases involved in plant iridoid biosynthesis.....	11
1.6 Atypical oxidoreductase involved in steroid biosynthesis	13
1.6.1 WWOX pathogenesis	15
1.7 Statement of problem	16
1.8 Objectives of the thesis	188
1.9 Bibliography.....	18
CHAPTER 2	
2.Materials & methods	23
2.1 Materials.....	23
2.2 Methods.....	25
2.2.1 Molecular biology methods	25
2.2.2 Analytical methods	29

2.2.3 Biochemical methods	30
2.2.4 Macromolecular X-ray crystallography.....	33
2.2.5 Cryo-electron microscopy	44
2.2.6 Computational methods.....	49
CHAPTER 3	
3.Structural basis of substrate specificity of Iridoid Synthases.....	53
3.1 Background	53
3.2 Methods.....	54
3.2.1 Cloning, protein expression and purification:	54
3.2.2 Site-directed mutagenesis	54
3.2.3 GC-MS assay	55
3.2.4 Spectrophotometric assay:.....	55
3.2.5 MD simulations	55
3.2.6 Homology Modelling	56
3.3 Results:.....	57
3.3.1 Cloning, expression and purification of IS.....	57
3.3.2 Crystallization and structure determination of IS.....	57
3.3.3 Structure of IS.....	59
3.3.4 Identification of sequence-signatures governing substrate specificity.....	60
3.3.5 Site-directed mutagenesis studies to validate “SY”/“EG” as substrate specificity defining residues.....	65
3.3.6 Structural basis of substrate selectivity of SDR family of reductases.....	68
3.3.7 Determinants of IS specificities.....	69
3.3.8 Determinants of P5βR specificities	70
3.3.9 Molecular dynamics simulation of IS reveal altered top-channel dynamics:.....	72
3.4 Discussion	76
3.5 Bibliography.....	79
CHAPTER 4	
4.Structural insights into cascade mechanism of IS and 10-HGO	82
4.1 Background	82
4.2 Methods.....	83
4.2.1 Cloning, protein expression and purification	83

4.2.2 Activity assay	84
4.2.3 Bacterial DHFR PCA survival selection assay.....	84
4.2.4 Western blot.....	85
4.2.5 SEC-MALS	85
4.2.6 Cryo-EM.....	85
4.3 Results and discussion.....	86
4.3.1 10-HGO and IS physical interaction to produce cyclic nepetalactol.....	86
4.3.2 Chimera of 10-HGO and IS as surrogate system	87
4.3.3 Structural basis of 10-HGO-IS interaction	88
4.5 Bibliography.....	90
APPENDIX	
Structural studies of 10-hydroxygeraniol dehydrogenase from <i>C. roseus</i>	92
4.6 Background	92
4.7 Methods.....	92
4.7.1 Cloning, protein expression and purification	92
4.7.2 Substrate docking	93
4.8 Results and discussion.....	93
4.8.1 Cloning, expression and purification of 10-HGO	93
4.8.2 Crystallization and structure determination of 10-HGO	94
4.8.3 Overall structure of 10-HGO	94
4.8.4 Active site and catalytic mechanism	97
4.9 Conclusion.....	99
4.10 Bibliography.....	99
CHAPTER 5	
5. Structural studies of WW domain containing oxidoreductase (WWOX)	101
5.1 Background	101
5.1.1 WW domain structure.....	102
5.2 Methods.....	103
5.2.1 Cloning, protein expression and purification	103
5.2.2 Thermal shift assay.....	104
5.2.3 SEC-MALS	104
5.2.4 Site-directed mutagenesis.....	1055

5.2.4 Activity assay	105
5.2.5 Spectrophotometric assay	105
5.2.6 Homology modelling.....	106
5.3 Results	106
5.3.1 Cloning, expression and purification of WWOX.....	106
5.3.2 Determining WWOX targets	109
5.3.3 Site-directed mutagenesis studies.....	111
5.3.4 Crystallization of WWOX.....	112
5.3.5 Data collection and structure determination.....	113
5.4 Discussion	116
5.5 Bibliography.....	117
CHAPTER 6	
6.Summary and future scope	120
7.Supplementary information.....	123

List of figures

1.1	Reversible redox reaction between NADH/NADPH and NAD ⁺ /NADP ⁺	03
1.2	Schematic and structural representation of Rossmann fold	04
1.3	Active site of SDR family protein: human 17 β -hydroxysteroid dehydrogenase	06
1.4	Active site of MDR family protein: sinapyl alcohol dehydrogenase	07
1.5	Active site of LDR family protein: mannitol dehydrogenase	08
1.6	Active site of AKR family protein: 3 α -hydroxysteroid dehydrogenase	09
1.7	Active site of FeADH family protein: 1,2 Propanediol oxidoreductase	10
1.8	Biosynthesis pathway of various MIAs	12
1.9	Iridoid biosynthesis pathway	13
1.10A	Domain organization and sequence of WW domains of WWOX	16
1.10B	Pathway showing the interactions of WWOX with other signaling proteins	16
2.1	USER cloning method	27
2.2	Workflow of X-ray crystallography	34
2.3	Types of crystallization methods	35
2.4	Argand diagram of isomorphous replacement	40
2.5	Argand diagram of anomalous dispersion	41
2.6	Workflow of single particle reconstruction	45
3.1	Reactions catalyzed by iridoid synthase and progesterone 5 β -reductase	53
3.2	Cloning and purification of IS	57
3.3	Crystals of IS	58
3.4A	Structure of IS and NADP complex.	60
3.4B	Active site of IS	60
3.4C	Interaction of Tyr178 and Lys146 with NADP and 8-oxogeraniol	60
3.4D	Superposition of IS and P5 β R	60
3.5	ScanProsite web server	61
3.6	Sequence alignment of P5 β R hits from ScanProsite	63
3.7	Sequence alignment of IS hits from ScanProsite	64
3.8A	Divergent residues between IS and P5 β R lying within a radius of 5 \AA	65
3.8B	Relative activities of IS and IS ^{7M}	65

3.9A	Superimposition of IS and P5 β R structures showing the difference in conformation of L1	66
3.9B	Relative activity of wild-type IS, P5 β R, IS ^{SY} and P5 β R ^{EG}	66
3.10	Product profiles of GC-MS analysis with 10-oxogeranial and progesterone	67
3.11	Difference in L1 conformation among different crystal structures of IS	68
3.12	Surface representation of IS and P5 β R structures	70
3.12B	Surface representation of IS with closed state conformation	70
3.13	Modeled structure of IS ^{SY} with confirmation of L1 as in P5 β R	71
3.14A	Snapshots of the L1 loop conformational dynamics in the IS and IS ^{SY} simulation	73
3.14B	Difference plot of probability of pair-wise contacts between all protein residues	73
3.14C	The surface representation of wild-type and mutant protein as observed in simulations	73
3.15	Segment of L1 in P5 β R form type-II β turn	74
3.16	Secondary structure content for apo and 10-oxogeranial bound IS and IS ^{SY}	75
3.17	Superimposition of modeled IS homologs	78
3.18	Multiple sequence alignment of IS and P5 β R with other IS	79
4.1	Reactions catalyzed by 10-HGO and IS	82
4.2A	Schematic representation of IS-DHFR3 and 10-HGO-DHFR2 plasmids	87
4.2B	Agar plate showing colonies of IS-DHFR3 and 10-HGO-DHFR2 co-transformation	87
4.3A	SEC profile of 10-HGO and IS <i>in vitro</i> complex	88
4.3B	Agarose gel of chimera PCR and RD	88
4.3C	SDS-PAGE and MALS results of chimera	88
4.3D	GCMS product profile of chimera	88
4.4A	2D class averages of chimera	89
4.4B	2D projection of docked IS-10-HGO structure map	89
4.4C	Putative structure model of chimera	89
4.5A	Coevolution analysis for 10-HGO and IS	90
4.5B	Distance plot of selected sequences from phylogeny	90
4.6A	SDS-PAGE images and gel filtration profile of 10-HGO purification	94
4.6B	Crystals of 10-HGO	94

4.7A	Overall structure of 10-HGO	97
4.7B	Active site of 10-HGO contains one catalytic Zn coordination	97
4.7C	Structural Zn coordinated by C103, C106, C109, C117	97
4.8A	Docked structure of 10-HGO with 10-hydroxygeraniol	99
4.8B	Hydrophobic cavity of 10-HGO	99
4.8C	Proton shuttling mechanism during the reduction process in the active site of 10-HGO	99
5.1A	Consensus sequence of WW domains	103
5.1B	Structure of WW domain of YAP65 with three anti-parallel β -sheets	103
5.1C	Complex structure of WW domain (cyan) and PPxY motif	103
5.1D	Superimposition of atypical WWOX WW2 domain with Pin1WW1 domain	103
5.2A	Cloning of WWOX	107
5.2B	WWOX purification	107
5.2C	RONN disorder prediction of WWOX	107
5.3	Thermal shift assay melt curve plot	108
5.4A	WWOX purification	109
5.4B	MALS results fitting profile	109
5.5A	Modelled structure of WWOX SDR domain	111
5.5B	Product profiles of GC-MS analysis with substrates	111
5.5C	Relative activities of Δ 51-WWOX and Δ 110-WWOX towards Estradiol	111
5.6A	WWOX mutants purification	112
5.6B	Product profiles of GC-MS analysis with SDR and mutants	112
5.7	Crystals of WWOX	113
5.8	Self-rotation function of WWOX data sets	114
5.9A	SDS PAGE of selenomethionine containing WWOX	114
5.9B	Selenomethionine containing WWOX crystals	114

List of tables

1.1	Oxidoreductases subclasses and their function	02
2.1	Source of chemicals	23
3.1	Diffraction data collection and refinement statistics of IS	58
3.2	Example of hits from ScanProsite with search string containing SY	62
3.3	Example of hits from ScanProsite with search string containing EG	64
3.4	Kinetic parameters of NADPH consumption with 10-oxogeranial	66
3.5	Kinetic parameters of NADPH consumption with progesterone	66
3.6	Kinetic parameters of IS and P5βR with other plant ISs	77
4.1	Diffraction data collection and refinement statistics of 10-HGO	96
5.1	Summarized list of WWOX clones and their fate after purification	109
5.2	Kinetic parameters of NADP⁺ consumption with estradiol	111
5.3	Data collection statistics of WWOX diffraction data	115
5.4	Summarized molecular replacement runs with WWOX X-ray diffraction data	115

List of abbreviations

10HGO	10-hydroxygeraniol dehydrogenase
XOG	10-Oxogeraniol
Å	Angstrom
ADP	Atomic displacement factor
AKR	Aldo-keto reductase
ASU	Asymmetric unit
CCD	Charge coupled devices
Cryo-EM	Cryo-Electron microscopy
C-terminus	Carboxyl terminus
CTF	Contrast transfer function
DDD	Direct detection device
DMAPP	Dimethylallyl diphosphate
DOEP	Discrete optimized protein energy
DTT	Dithiothritol
dU	Deoxy uridine nucleotide
EDTA	Ethylene diamine tetra acetic acid disodium salt
FAD	Flavin adenine dinucleotide
FeADH	Iron dependent alcohol dehydrogenase
FEG	Field emission gun
FMN	Flavin mononucleotide
G10H	Geraniol 10-hydroxylase
GC	Gas chromatography
GDS	Geranyl diphosphate synthase
GPP	Geranyl diphosphate
GS	Geraniol synthase
GST	Glutathione S-transferase
HPC	Hybrid Photon Counting
IPP	Isopentenyl diphosphate

IPTG	Isopropyl 1-thio-D-galactopyranoside
IS	Iridoid synthase
kDa	Kilo Dalton
LB	Luria Bertani broth
LDR	Long chain dehydrogenase/reductase
MD	Molecular dynamics
MDR	Medium chain dehydrogenase/reductase
mg	Milligram
MIA	Monoterpene indole alkaloids
min	Minute(s)
MR	Molecular Replacement
MRA	Multi-reference alignment
MS	Mass spectrophotometer
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide (phosphate)
NADPH	Nicotinamide adenine dinucleotide (phosphate) reduced
nm	Nano meter
N-terminus	Amino terminus
OD	Optical density
P5βR	Progesterone 5 beta reductase
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Poly ethylene glycol
PH	Pleckstrin homology
PME	Particle-Mesh Ewald
PTB	Phosphorylated tyrosine binding domains
PTS	Protein thermal shift
RE	Restriction enzyme
RMSD	Root mean square deviation

SDR	Short chain dehydrogenase/reductase
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Sec	Seconds
SEC-MALS	Size exclusion chromatography-Multi angle light scattering
TAE	Tris-acetate EDTA
TBME	Tert-Butyl Methyl Ether
TCA	Tri carboxylic acid
TEM	Transmission electron microscopy
TEV	Tobacco etch virus
UDG	Uracil DNA glycosylase
USER	Uracil-Specific Excision Reagent
WWOX	WW domain containing oxidoreductase
μg	Microgram
μl	Microliter
μM	Micro molar

List of publications

- **Dynamics of loops at the substrate entry channel determine the specificity of iridoid synthases**

Anand S. Sandholu, Madhura Mohole, William L. Duax, Hirekodathakallu V. Thulasiram, Durba Sengupta and Kiran Kulkarni* *FEBS Letters* 592 (2018) 2624–2635.

- **Structural Insights into Hand-shake Mechanism of 10-HGO and IS**

Anand S. Sandholu, S. Mujawar, Hirekodathakallu V. Thulasiram, Kiran Kulkarni* (Manuscript under preparation)

- **Structural basis of substrate specificity of 10-hydroxygeraniol dehydrogenase**

Anand S. Sandholu, S. Mujawar, R. Krithika, Hirekodathakallu V. Thulasiram, Kiran Kulkarni* (Manuscript under preparation)

- **Structural basis of specificity of WW domain containing oxidoreductase (WWOX) and its role as tumor suppressor**

Anand S. Sandholu, Kiran Kulkarni* (Manuscript under preparation)

Thesis abstract

Oxidoreductases constitute a large class of enzymes, present in all three domains of life. They use a variety of electron acceptors and electron-donating substrates, yielding many products of pharmacological and biotechnological interest. Aberrations in their function are implicated in various diseases and posing as drug targets. We employed X-ray crystallography and Cryo-electron microscopy techniques for structural and functional investigation of three atypical NADPH-dependent oxidoreductase systems which possess non-canonical oxidoreductase mechanism. Two of them govern the iridoid biosynthesis, a precursor for vincristine and vinblastine (iridoid synthase and 10-hydroxygeraniol dehydrogenase) and the third one, WW domain containing oxidoreductases, is a newly identified drug target for breast cancer.

Iridoid synthases (IS), a short-chain dehydrogenase/reductase, involved in the biosynthesis of iridoids, exhibit stringent substrate specificity. Unlike the canonical terpene synthases, IS catalyzes the NADPH-dependent reduction of 10-oxogeraniol resulting in the formation of an *enol* or *enolate* intermediate, which further cyclizes *via* Michael addition to yield napetalactol. To understand the structural basis of the catalytic mechanism and substrate specificity of IS we determined the crystal structure of IS. From structural, biochemical and molecular dynamics studies we show that a complex interplay of interactions between the loops at the substrate entry confers the substrate specificity to the enzyme. Furthermore, to understand the stereo selectivity production of napetalactol in a concerted chemical reaction of IS and its precursors enzyme in the pathway, 10-HGO, we probed the structural basis of their interaction using single particle cryo-EM and other biophysical studies. Here we show that 10-HGO and IS indeed interact physically to form a channel between the molecules to facilitate passage of former's (10-HGO) product as a substrate to the later (IS) in the “handshake” manner. Additionally, the crystal structure of 10-HGO provided insight into its catalytic mechanism of the enzyme. Continuing in the same direction, we addressed the structural basis of catalytic mechanism of a recently identified SDR, WW domain-containing oxidoreductase (WWOX), shown to be implicated in several human cancers. Our studies unravel putative substrates of WWOX, its specificity. We have further successfully crystallized this protein to elucidate the mechanism of the protein. The outcome of these studies has significant implications in biotechnology of alkaloid synthesis and development of novel drugs targeting WWOX.

Thesis overview

Chapter1: Introduction

This chapter provides general introduction and literature review of NADPH-dependent oxidoreductases and their classification. It also describes the structural features and catalytic aspects of various types of alcohol dehydrogenases and their current applications. Concluding part of this chapter comprises of the biological questions being addressed and objectives of this thesis.

Chapter 2: Materials and methods

This chapter consists of materials and methodologies that are used in this study. Sections mainly include methods applied for structure determination and biochemical characterization of proteins. Principle and adaptation of two popular methods for the structure determination, X-ray crystallography and Cryo-Electron microscopy, are discussed. Additionally, various other methods including molecular, analytical, biochemical and computational methods have also been discussed. The complete list of materials used in this study are listed.

Chapter 3: Structural basis of substrate specificity of iridoid synthases

This chapter comprehends the structural and biochemical characterization of iridoid synthase enzyme. Additionally, this chapter reveals the structural aspects of substrate specificity determinants of IS enzyme through various biophysical, biochemical and computational methods. This chapter provides details on the substrate specificity of other SDR superfamily members.

Chapter 4: Structural insights into the “hand-shake” mechanism of 10-HGO and IS

This chapter dwells on the physical interaction of 10-HGO and IS through biochemical characterization of the surrogate system, 10-HGO and IS (Chimera). It also describes the structural aspects of chimera and also highlights the mode of 10-HGO-IS interaction. Additionally, this chapter outlines the structural and biochemical characterization of 10-HGO.

Chapter 5: Structural studies of WW domain containing oxidoreductase (WWOX)

This chapter deals with the heterologous expression and purification of soluble WWOX. Furthermore, it unravels the putative substrates of WWOX. The chapter concludes with discussion on the efforts made on the crystallization of WWOX and their optimization. It concludes by highlighting the challenges involved in the structure determination of WWOX.

Chapter 6: Summary and future scope

This chapter summarizes the thesis by expounding on the important findings and future scope of this research work presented here.

Chapter1

"Be an enzyme -a catalyst for change"

-Pierre Omidyar

Introduction

Oxidoreductases control myriad of metabolic pathways and vital cellular functions such as digestion, cellular respiration and energy storage *etc*¹. This class of enzymes primarily involved in the maintenance of cell homeostasis and cellular integrity through various oxidation-reduction reactions². The target and action of these enzymes are regulated by numerous external stimuli. Oxidoreductases are the second largest class of enzymes, emerged during the early biogenesis and are found in all three domains of life that include archaea, prokaryotes and eukaryotes³. Oxidoreductases drive several metabolic pathways such as glycolysis, citric acid cycle and oxidative phosphorylation⁴. Furthermore, these enzymes are involved in DNA synthesis and repair, signal transduction pathways and ROS detoxification⁴. Additionally, some secondary metabolic pathways in plants and microbes including terpenoids, alkaloids, phenylpropanoids and polyketides biosynthesis pathways are also catalyzed by various oxidoreductases⁵. All these diverse niches and tasks of oxidoreductases show that they are essential for life.

Oxidoreductases specifically catalyze the transfer of electrons from one molecule (the oxidant) to another (the reductant), schematically, $A^- + B \rightarrow A + B^-$, where A is the oxidant and B⁻ is the reductant⁶. Oxidoreductases act on a numerous organic substrates including amines alcohols, ketones, inorganic molecules like sulfite and also metals such as iron and mercury *etc*^{7,8}. Based on the substrates, electron donor and acceptor, oxidoreductases are further classified as oxidases, dehydrogenases, peroxidases, hydroxylases, oxygenases and reductases *etc* (Table1.1) (<http://www.brendaenzymes.org>). Oxidoreductases always depend on cofactors as an electron carrier for their catalytic activity. Some of the cofactors include NAD, NADP, FAD, dihydrobiopterin, molybdopterin, lipoate, ubiquinone, heme-Fe, glutathione *etc*⁹. Among all oxidoreductases, 50% of them depend upon NADPH for their activity⁶.

Table 1.1: Oxidoreductases subclasses and their function

Subclass	Enzyme function
Oxidases	Catalyzes an oxidation-reduction reaction, where molecular oxygen acts as an acceptor
Dehydrogenases	Substrate oxidation carried out by reducing an electron acceptor like NAD ⁺ /NADP ⁺ or FAD/FMN
Peroxidases	Catalyzes oxidation of a particular substrate by hydrogen peroxide.
Hydroxylases	Adds hydroxyl group to its substrate
Oxygenases	Incorporates oxygen from molecular oxygen into organic substrates
Reductases	Catalyzes reduction reaction

1.1 NADPH-dependent oxidoreductases:

These enzymes carry out the oxidation of a substrate by donating a hydride (H⁻) group to a nicotinamide adenine dinucleotide (phosphate) cofactor (NAD⁺ or NADP⁺), reduced to NADH or NADPH. On the other hand, some enzymes are able to reduce substrate by transferring the hydride (H⁻) group of NADPH to substrate resulting in oxidized NAD⁺ or NADP⁺ (Fig 1.1)¹⁰. So far, there are over 150,000 sequences reported and annotated as NADPH-dependent oxidoreductases¹¹. NADPH has a ribosylnicotinamide 5'-phosphate coupled with a pyrophosphate linkage to 5'-phosphate adenosine 2',5'-bisphosphate. The nicotinamide ring of NADH/NAD⁺ or NADPH/NADP⁺ is directly involved in transferring electrons during the catalysis, while the C4 carbon of nicotinamide ring serves as a proton acceptor/ donor¹⁰. Electron transport capability of NADPH/NADP⁺ is not affected by the addition of phosphate to 2'-OH group of adenine ribose ring, as it is present far from the region of electron transfer. However, the phosphate group changes the structure of cofactor, that allows enzymes to have specificities towards NADH/NAD⁺ or NADPH/NADP⁺. The pool of NADH/NAD⁺ and NADPH/NADP⁺ acts as redox systems in maintaining the cellular homeostasis¹².

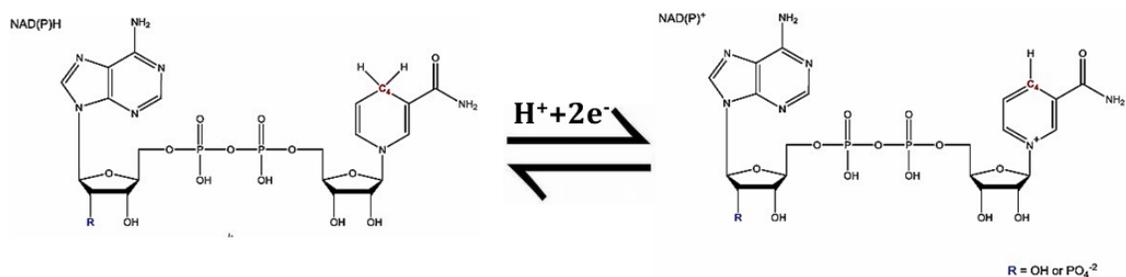


Fig 1.1: Reversible redox reaction between NADH/NADPH and NAD⁺/NADP⁺

1.2 Structural characteristics of NADPH-dependent oxidoreductases:

Over the past decades, a number of high-resolution crystal structures of NADPH-dependent oxidoreductases have been deposited. The structural data enables us to understand the catalytic, physical and chemical properties of these enzymes. Moreover, the structural information provides insights into their substrate specificity and promiscuity. This information has turned out to be vital in the engineering of these enzymes and has opened up new avenues in synthetic biology. X-ray crystallography has revealed two domains in NADPH-dependent oxidoreductases; a nucleotide binding domain and a catalytic domain. The nucleotide binding domain helps in the recruitment of NADPH/NADP⁺ cofactor through a conserved Rossmann fold. This structural motif is also found in several nucleotide-binding proteins, first discovered by Michael Rossmann in lactate dehydrogenase¹³. It contains one α -helix sandwiched by two parallel β -strands ($\beta\alpha\beta$ motif) (Fig 1.2). A tight compact loop that connects the first β -strand and α -helix contain consensus sequence Gly-x-x-x-Gly-x-Gly/Ala (wherein X represents any amino acid), which stabilizes the cofactor binding through direct contacts with pyrophosphate of dinucleotides¹⁴. Additionally, a conserved water molecule is present between the pyrophosphate and glycine-rich loop to bridge their interactions¹⁵. Several NADPH-dependent oxidoreductases are usually composed of two Rossmann folds, one stabilizes the nicotinamide ring and the other stabilizes the adenine moiety. The core $\beta\alpha\beta$ can be extended sometimes to several β -strands in order to form a larger β -sheet. Generally, all β -strands are parallel, but in some proteins, the strands are anti-parallel (For example, methylenetetrahydromethanopterin dehydrogenase and homoserine dehydrogenase)¹⁶. Secondary structures between the additional β -strands can be varied by random coils, α -helices or combinations of coiled segments and small helices.

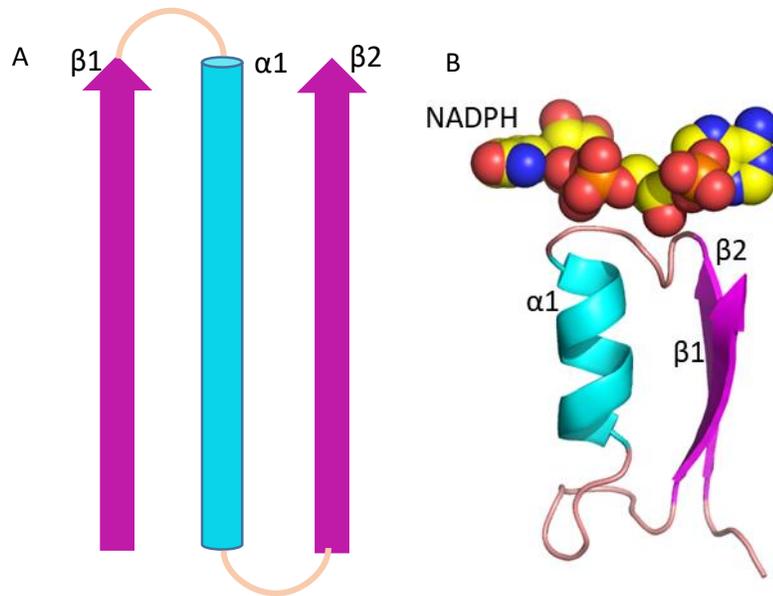


Fig 1.2: A) Representation of Rossmann fold motif B) Structure of Rossmann motif, NADPH showed in yellow spheres (PDB:5EMH)

The second domain of the enzyme is a catalytic domain, composed of residues essential for the reduction reaction and it also facilitates the substrate entry or exit. Unlike, the nucleotide binding domain, the oxidoreductases catalytic domain are variable due to diversity in their catalytic reactions. Enzymes like short-chain alcohol dehydrogenases/reductases lack distinct catalytic domain rather they possess a single domain, which contains a scaffold of Rossmann-fold and a highly variable C-terminal substrate binding site. The catalytic domains of these enzymes are responsible for the redox mechanism and substrate specificity. These oxidoreductases mainly vary with respect to their cofactor, metal ion binding and their C-terminus chain length¹⁷.

1.3 Classes of NADPH-dependent oxidoreductases:

NADPH-dependent oxidoreductases involved in broad range of redox reactions where they differ in the entities participating in electron flow. Although electron transfers occur directly between nicotinamide cofactor and the substrate, some class of enzymes such as alcohol dehydrogenase additionally depend on intermediary redox cofactors like metal ions for catalysis¹⁸. Alcohol dehydrogenases are one of the largest groups of NADPH-dependent oxidoreductases.

1.3.1 Alcohol dehydrogenases:

Alcohol dehydrogenases are ubiquitous and present in all forms of life. They catalyze reversible oxidation of alcohols to aldehydes or ketones in the presence of NADP⁺/NADPH¹⁸. These enzymes are majorly involved in anabolic and catabolic pathways. In some instances, they interact with downstream enzymes to perform thermodynamically favorable cascades¹⁹. Alcohol dehydrogenases majorly encompass five superfamilies, which include:

- Short-chain dehydrogenase/reductase (SDR)
- Medium-chain dehydrogenase/reductase (MDR)
- Long-chain dehydrogenase/reductase (LDR)
- Aldo-keto reductase (AKR)
- Iron-activated alcohol dehydrogenase (FeADH)

1.3.1.1 Short-chain dehydrogenase/reductase (SDR) superfamily:

The SDR superfamily is one of the largest protein families that comprises of more than 20000 sequences known from databanks till date. Genome investigations showed that about 1/4 of all dehydrogenases are found in SDRs²⁰. This superfamily is present in all three domains of life. These enzymes have defined physiological roles in amino acid, fatty acid, sugar and steroid metabolism. Additionally, in plants, these enzymes play a key role in secondary metabolite pathways such as terpenoids and alkaloids biosynthesis²¹. These enzymes are further divided into two types, denoted as classical and extended. The classical type has a chain length of about 250 amino acids (17 β -hydroxysteroid dehydrogenase and WW domain containing oxidoreductase), while the C-terminal of extended type has an additional 100 residue domain (GDP-fucose synthetase)^{20,22}.

1.3.1.1.1 Structural features of SDR superfamily proteins:

Structural information on SDRs has increased tremendously over the last few years, with over 200 structures deposited in the PDB (protein data bank). Although they are characterized by low sequence conservation, they share several common structural motifs:

- A conserved $\beta\alpha\beta$ folding pattern characterized by a 'Rossmann-fold'
- N-terminal cofactor binding motif-GxxxGxG

- A conserved catalytic triad-Sx₁₂YxxxK
- Variable substrate binding region at the C-terminal region

During the catalysis, a key catalytic Tyr residue acts as a proton donor or acceptor from the substrate²³. This is facilitated by an adjacent Lys residue and the nicotinamide ring of NADP⁺, which lowers pKa of the hydroxyl group of Tyr residue. During the catalysis Tyr withdraw a proton from substrate where a hydride is transferred to the oxidized cofactor (Fig1.3). However, in few SDRs, the Tyr residue present in active site do not act as a proton/donor-acceptor, or is absent in some enzymes such as human peroxisomal enoyl-CoA reductase. Instead, protons transfer is facilitated by the solvent. The role of Ser in the catalytic triad is less apparent but, it may be involved in stabilizing the orientation of substrate¹⁷.

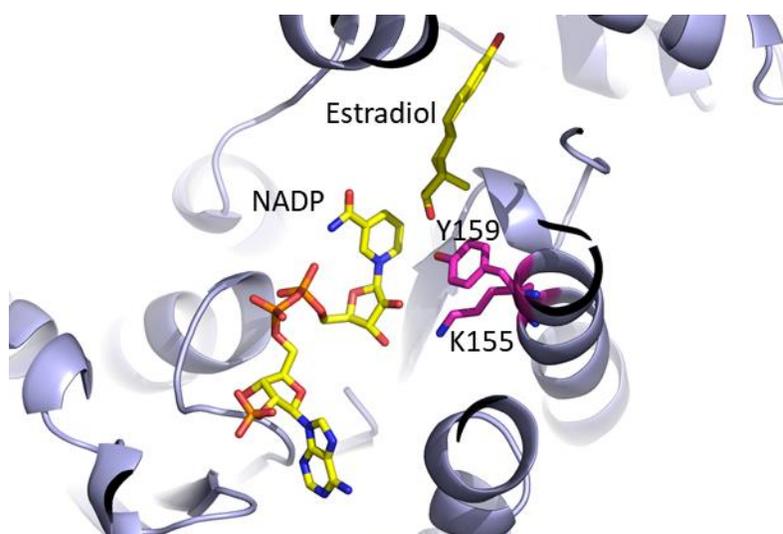


Fig 1.3: Active site of human 17β-hydroxysteroid dehydrogenase (PDB code 1A27). 17β-hydroxysteroid dehydrogenase belongs to the SDR family. The substrate is displayed in yellow, NADP⁺ is represented in yellow and residues essential for enzyme activity are labelled and colored magenta.

The SDR superfamily is active on a wide spectrum of substrates in both plants and animals. In humans, some of these SDRs are considered as promising therapeutic targets (17β-hydroxysteroid dehydrogenase) whereas, targeting others may lead to lethal outcomes (acyl-CoA dehydrogenase). Despite their essential roles in various biological processes, the physiological functions of 50% of SDR members are still unknown.

1.3.1.2 Medium chain alcohol dehydrogenases (MDR) superfamily:

The MDR proteins are typically zinc-dependent alcohol dehydrogenases¹⁸. There are over 15,000 members reported in database¹⁷. MDR proteins contains two domains, the C-terminal coenzyme-binding domain with conserved Rossmann fold and the N-terminal substrate binding domain with surface-positioned α -helices and a core of antiparallel β -strands. The coenzyme-binding domain contains α -helices surrounded by six-stranded parallel β -sheet on each side with These two domains are separated by a deep active site pocket present in the form of a cleft. MDRs required additional Zn^{2+} cofactor for their catalysis. Usually, many MDRs binds to two Zn^{2+} ions where one is catalytic Zn^{2+} and other is structural Zn^{2+} . The catalytic Zn^{2+} is coordinated by three conserved residues (two cysteine and a histidines) and a water molecule with tetrahedral geometry²⁴ (Fig 1.4). Upon substrate binding, the water molecule is displaced from the coordination shell of zinc. Subsequently, a proton is transferred to solvent from the substrate, forming an intermediate that is stabilized by Zn^{2+} . Further the intermediate transfers a hydride to $NADP^+$ thus, converting to an aldehyde/ketone product²⁴.

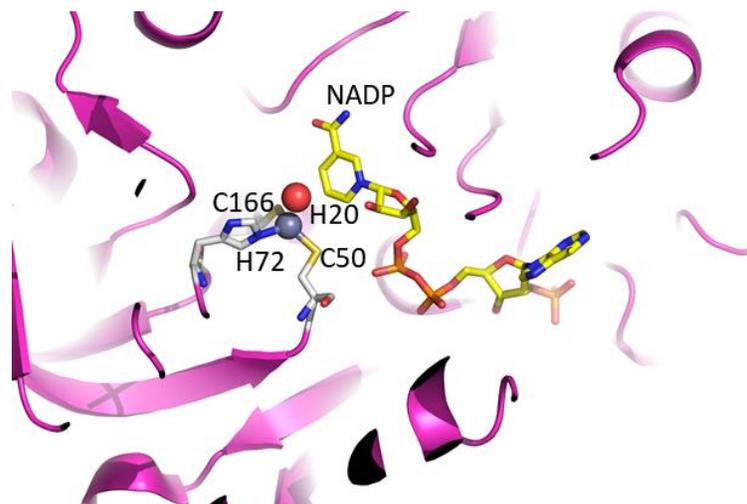


Fig 1.4: Active site of plant sinapyl alcohol dehydrogenase (PDB code 1YQD). Sinapyl alcohol dehydrogenase belongs to the MDR family. $NADP^+$ is displayed in yellow and residues that are essential for Zn^{2+} coordination are labelled and colored grey. Conserved water is colored red and the catalytic Zn^{2+} is colored blue.

1.3.1.3 Long chain alcohol dehydrogenases (LDR) superfamily:

LDRs are a diverse group of enzymes that differ in both, their catalytic mechanism and additional specificity for cofactors. Like SDRs and MDRs, LDRs possess similar topology. However, they employ a different catalytic mechanism compared to SDR and MDR members²⁵. The catalytic site contains a conserved Lys residue, which acts as a general base by receiving a proton from the substrate and transferring it to solvent. The role of lysine as a general base is enriched by hydrogen bond with an adjacent Asn residue (Fig 1.5). This hydride is now transferred to the oxidized cofactor (NADP⁺)²⁶.

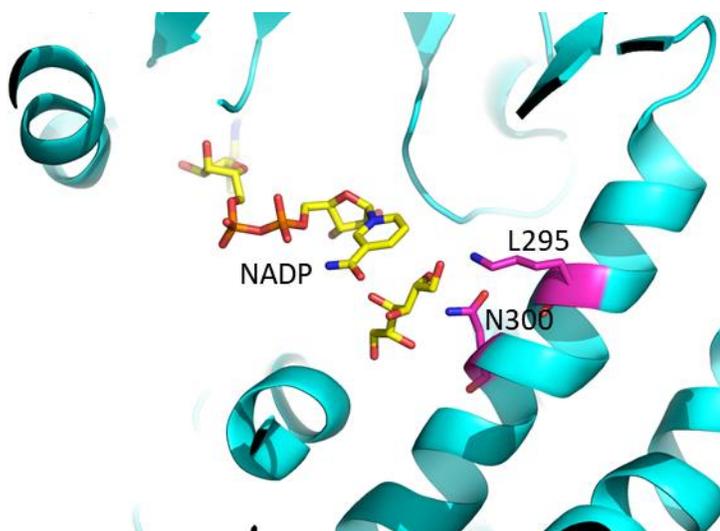


Fig 1.5: Active site of mannitol dehydrogenase (PDB code 1M2W). Mannitol dehydrogenase belongs to the LDR family. NADP⁺ and mannitol are displayed in yellow and residues that are essential for enzyme activity are colored magenta.

1.3.1.4 Aldo-keto reductases (AKRs) superfamily:

This family constitute about fourteen sub-families and are designated as AKR1-AKR14. Till now about 125 potential AKRs have been identified. AKRs adopt eight α/β barrel structural scaffold with a buried active site at the C-terminal end. The active site is composed of conserved Asp50, Tyr55, Lys84, His177 (as per PDB:1J96)²⁷ (Fig 1.6). The mode of NADP⁺ binding and fold is conserved in all AKRs. These enzymes are mainly involved in detoxification of aldehydes produced during lipid peroxidation, oxidative stress and drug metabolism²⁸. There is very little known about these enzymes, however, recent studies showed that AKRs are also involved in isoflavonoid and steroid metabolism²⁹.

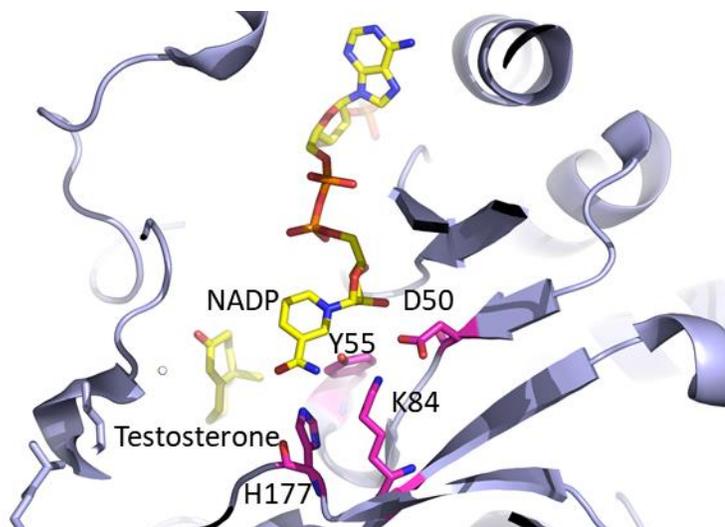


Fig 1.6: Active site of 3 α -hydroxysteroid dehydrogenase (PDB code 1J96). 3 α -hydroxysteroid dehydrogenase belongs to the AKR family. NADP⁺ and testosterone are displayed in yellow and residues that are essential for activity are labelled and colored magenta.

1.3.1.5 Iron-activated alcohol dehydrogenases (FeADH) superfamily:

FeADHs constitute an ancient protein family, comprising of 19 different subfamilies³⁰. FeADHs are activated primarily by Fe²⁺, but some require other divalent metal ions like Zn²⁺³¹. In eukaryotes, majority of FeADHs belongs to hydroxyacid oxoacid transhydrogenases (HOT) subfamily. The N-terminal domain of these enzymes consists of a Rossmann fold motif that mediates binding of coenzyme. Residues that are involved in coenzyme binding are conserved across subfamilies of FeADH. A conserved GGGS motif is present in the loop that links the α 4 helix and β 4 strand interacts with the pyrophosphate group of NADP⁺ and stabilizes its binding³⁰. Additionally, residue present at position 81 determines the specificity of the coenzyme. Those enzymes possess threonine or aspartate at 81st position, bind to NAD⁺, and the enzymes with glycine at the same position binds to NADP⁺. The Fe²⁺ of FeADH is coordinated *via*. an ion-dipole interaction with four conserved residues such as Asp, His, His, and His in tetrahedral geometry (Fig 1.7)³⁰. These enzymes are sensitive to oxygen and are majorly involved in various anaerobic metabolic processes.

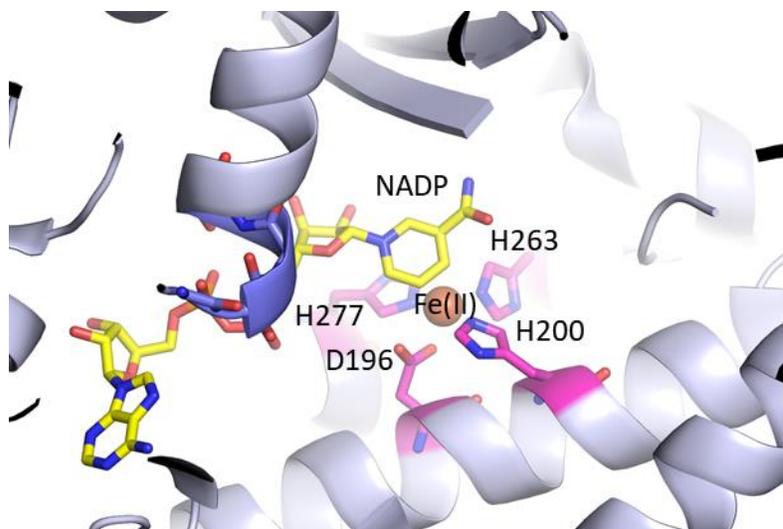


Fig 1.7: Active site of 1,2 propanediol oxidoreductase (PDB code 2BL4). 1,2 propanediol oxidoreductase belongs to the FeADH family. NADP⁺ is displayed in yellow and residues that are essential for cofactor stability are labelled and colored blue. Residues that are essential for Fe²⁺ coordination are labelled and colored magenta and Fe²⁺ labelled and colored in orange.

1.4 Applications of NADPH-dependent oxidoreductases:

In industrial point of view, NADPH-dependent oxidoreductases are on main focus as they perform crucial steps in the production of many compounds which are difficult to synthesize under amiable environment, for example, oxidation of primary alcohols. Although, several chemical methods have been developed to perform this reaction but they are generally laborious produce toxic end products. This can be avoided by employing enzymatic catalysis. Additionally, various properties of NADPH-dependent oxidoreductases make them attractive alternatives to chemical synthesis. Hence, NADPH-dependent oxidoreductases has been applied widely in synthetic biology for the production of various bioactive compounds such as artemisinin, taxol, ajmalicine, vinblastine and vincristine *etc* in a large scale. Amongst these, vinblastine and vincristine have grabbed a lot of attention due as they are widely used anticancer drugs and are majorly synthesized in plants through precursor molecules such as iridoids³². Furthermore, many of these enzymes are drug targets, as they are implicated in several diseases like cancer, cardiovascular dysfunction *etc*³³. Apart from this, these enzymes are also involved in the development of biosensors and diagnostic kits³⁴. In the previous decade, crucial progress have been made towards the use of oxidoreductases as biocatalysts to fulfil the complex pharmaceutical needs³⁴. Additionally, several of these enzymes are used in biodegradation processes like decomposition of polychlorinated biphenyls, phenolic

compounds and coal *etc*³⁵. This thesis reports structural and mechanistic studies on three atypical NADPH-dependent oxidoreductase systems which exhibit novel redox mechanism. Two of them govern the regulation of iridoid biosynthesis, a precursor for vincristine and vinblastine (Iridoid synthase & 10-hydroxygeraniol dehydrogenase) and the third WW domain containing oxidoreductases is a newly identified reductase shown to be involved in the migration of breast cancer cells.

1.5 Atypical oxidoreductases involved in plant iridoid biosynthesis:

Iridoids are structurally diverse class of bi-cyclic monoterpene indole alkaloids (MIAs), synthesized in plants as secondary metabolites³⁶. These are major biological source for production of biofuels, medicines and numerous chemical commodities³⁷. *C. roseus* is the source of iridoid derived MIAs and are known to produce about 2000 alkaloids in different tissues. Amongst, there are several important compounds such as ajmalicine, tabersonine and vindoline *etc*, which are widely used as anti-hypertensive and anti-inflammatory drugs. There are some dimeric indole alkaloids derived from MIA's such as vinblastine, vincristine that are being used as anticancer drugs³⁸. The preliminary step involved in the synthesis of these compounds is the condensation of tryptamine: a derivative of tryptophan and the iridoid monoterpene *i.e*, secologanin which gives rise to strictosidine intermediate. Strictosidine will further converted to various monomeric and dimeric indole alkaloids in the downstream process (Fig 1.8)

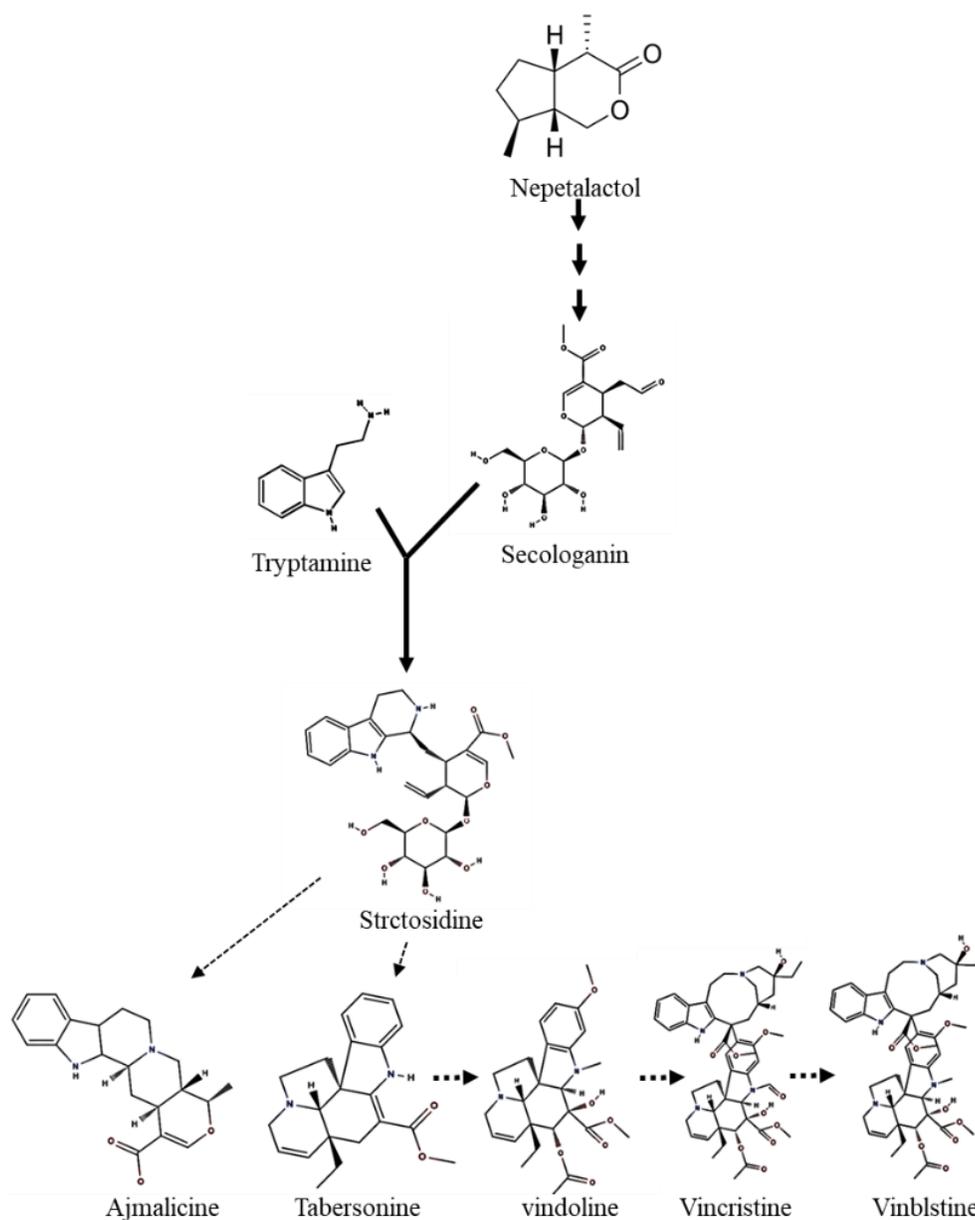


Fig 1.8 Biosynthesis pathway of ajmalicine, tabersonine, vindoline, vinblastine and vincristine

The MIAs biosynthesis diverges from the classical isoprenoid biosynthetic pathway at geranyl diphosphate (GPP), an intermediate formed from isopentenyl diphosphate (IPP) with dimethylallyl diphosphate (DMAPP), catalyzed by geranyl diphosphate synthase (GDS). Further, geraniol synthase (GS) hydrolyze GPP to geraniol, it undergoes hydroxylation at C10, forming 10-hydroxygeraniol in presence of geraniol 10-hydroxylase (G10H), a cytochrome P450 system. Later, 10-hydroxygeraniol is oxidized to 10-oxogeranial and nepetalactol by 10-hydroxygeraniol dehydrogenase (10-HGO) and iridoid synthase (IS), respectively (Fig 1.9). IS catalyzes reductive cyclization, unlike other canonical oxidoreductases, it combines reduction

with cyclization of linear 10-oxogeranial to bicyclic nepetalactol³⁶. The *cis-trans* nepetalactol, is a key precursor in the biosynthesis of vinblastine and vincristine. The cyclization of 10-oxogeranial and the stoichiometry between *cis-trans* nepetalactol and iridodials are majorly regulated by these two enzymes *i.e* 10-HGO and IS. IS is a 43 kDa protein, belongs to a family of progesterone reductases. Recent studies showed that IS exhibit firm substrate specificity. Furthermore, recent biochemical studies suggest that 10-HGO interacts with IS in order to shift the product equilibrium towards nepetalactol. In this thesis, the structural and mechanistic aspects of IS substrate specificity and 10-HGO-IS cascade interaction were discussed minutely.

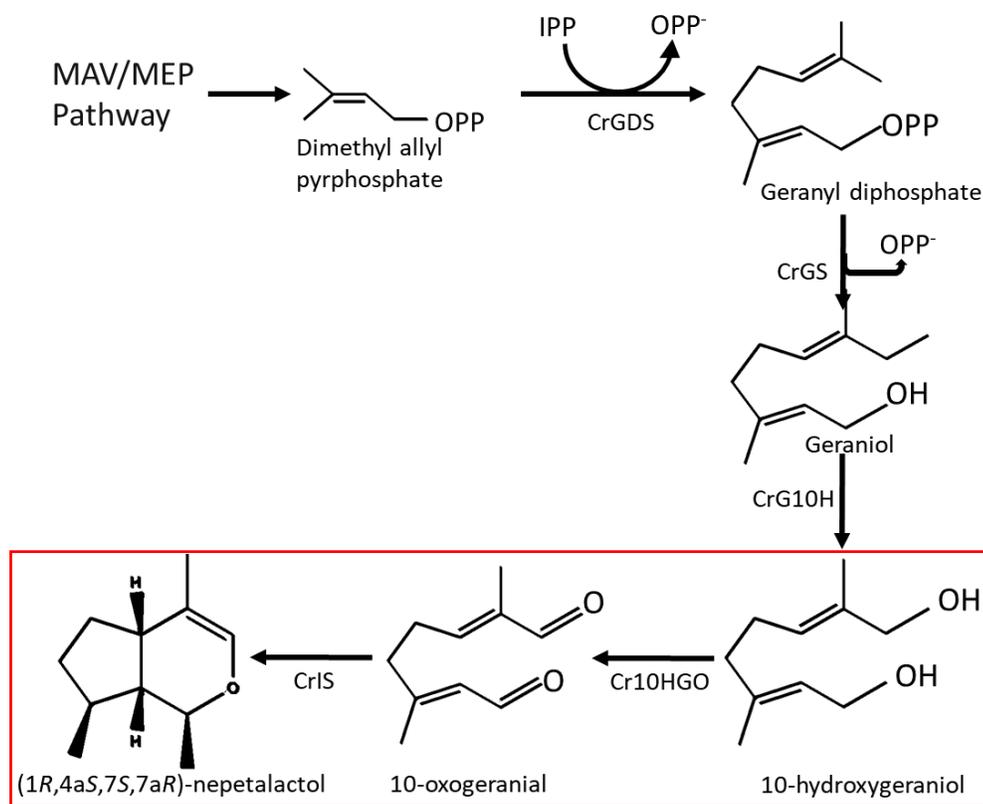


Fig 1.9: Iridoid biosynthesis pathway. Final two steps catalyzed by 10-HGO and IS showed in red box.

1.6 Atypical oxidoreductase involved in steroid biosynthesis:

WWOX (WW domain containing oxidoreductase) was recently identified in FRA16D, the highly affected chromosomal fragile site in several cancers³⁹. It is a member of NADP⁺/NADPH-dependent oxidoreductases, which plays an important role in various biological processes such as cell migration, cell differentiation, apoptosis and steroid

biogenesis⁴⁰. It is a 46 kDa protein consist of a C-terminal SDR domain and two N-terminal WW domains (Fig1.10A). The WW domains are shown to be involved in protein-protein interactions and thus regulate many cellular signaling pathways Fig (1.10B). The signaling pathways in which WWOX is implicated in, suggests that it could interact with other proteins via its first WW domain through proline-rich motifs-PPxY or PPPY of the interacting protein⁴¹. The WW domain is a small protein module consist of about 35 to 40 amino acids with two conserved Trp residues which are separated by 20-23 amino acids⁴². WW domains are known to occur with other structural and catalytically modules in a variety of signaling proteins such as adapter proteins, transcriptional co-activators (Yes-associated proteins (Yap): a transcription regulator of p73) and ubiquitin ligases (NEDD4 family proteins)^{43,44}. The foremost WW domain of WWOX categorized under class I WW domains family which are known to interact with proteins containing PP(P)(Y/Yp) motif (Yp indicates phosphorylated Tyr)⁴⁵ including p73, AP-2c, ErbB4, cJun, and dishevelled *etc*⁴¹. Among these p73 was the first identified WWOX partner. Recent reports showed that in cytoplasm WWOX binds p73 and suppresses its activation. However, the biochemical significance of WWOX-p73 interaction still remains elusive⁴⁶. Similarly, WWOX interacts with ErbB4 and protects it from degradation⁴⁷. WWOX is shown to interact with PPxY motif of dishevelled and acts as an inhibitor of the Wnt/ β -catenin pathway⁴⁸. Apart from these the first WW domain of WWOX is shown to be associated with cytoskeleton remodelling proteins including Ezrin a cytoskeleton membrane protein⁴⁹. This interaction has a role in the salient localization of WWOX towards the apex in parietal cells. Abrogation of WWOX-Ezrin synergy is shown to affect the remodeling of apical membrane cytoskeleton⁵⁰. Alterations in these interactions have been implicated in cancer metastasis, wherein WWOX plays a crucial role in cancer cell surface modulation, adhesion and migration. However, the complete role of WWOX in cytoskeleton regulation is yet to be explored. Till date, there are no interacting partners of the second WW domain of WWOX were reported. Interestingly the second WW domain contains Tyr in place of conserved tryptophan. This uncommon replacement of Tyr in place of Trp makes it an atypical WW domain. However the complete biological role this domain is not known. But some reports suggest that the secondary WW domain assists the first WW domain in protein-protein interaction⁵¹. In addition to two WW domains, WWOX contains C-terminal SDR domain. Presence of highly conserved nicotinamide binding motif and catalytic motifs

includes GxxxGxG motif and YxxxK motif respectively suggest that SDR domain of WWOX could be exhibit reductase activity similar to that found in canonical SDR superfamily enzymes⁵² Till date, the exact molecular function of this domain is not known. However, earlier studies have shown that the expression levels of WWOX are significantly higher in testis, prostate, and ovary tissues⁵³. Currently, WWOX is arousing much research interest because of its enigmatic role in cancers and several other disorders.

1.6.1 WWOX pathogenesis:

Several mutations in the *WWOX* gene have shown to result in reduction or complete loss of protein expression⁴⁰ as found in breast, ovarian, prostate, colon, and gastric carcinomas⁵⁴. In metastatic breast cancer tissues, reduced levels of WWOX alters the expression of integrin $\alpha 3$ and disturbs the interaction between cancer cells and extracellular matrix thus, resulting in cancer cells migration through the basal membrane⁵⁵. Additionally, loss of WWOX expression leads to severe neuronal diseases including ataxia, epilepsy, microcephaly and retinal degeneration etc. Loss of WWOX is also implicated in metabolic disorders associated with lipid and glucose metabolism. In neuronal cells down regulation of WWOX causes cascade of tau protein aggregation, which results in neuron degeneration and hence implicated in Alzheimer's diseases⁵⁶.

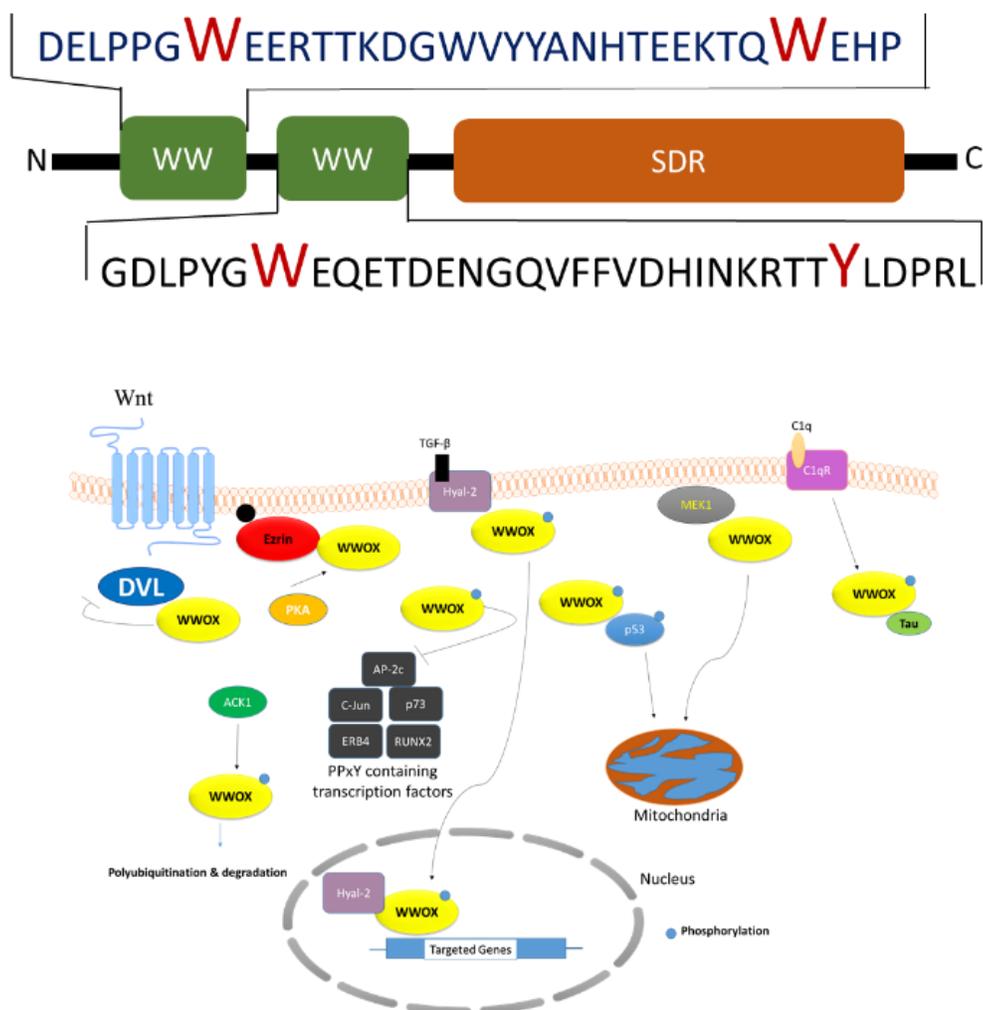


Fig 1.10: A) Domain organization of WWOX, Sequence of WW domains (1,2) Conserved Trp and Tyr shown in red color B) pathway showing the interactions of WWOX with other signaling proteins⁵⁷.

1.7 Statement of problem:

NADPH-dependent oxidoreductases have shown tremendous potential for a variety of biotechnological application. Furthermore, recent studies have shown that members of this family are involved in many pathologies including cancer progression. Hence, exploring these two aspects of NADPH-dependent oxidoreductases using structural biology tools could pave new ways for the development as well as large scale production of drugs.

IS, a non-canonical member of SDR superfamily, is the key enzyme involved in the synthesis of iridoids. IS catalyzes NADPH-dependent reduction of 10-oxogeranial to produce an enol or enolate intermediate, which further cyclizes *via* Michael addition to yield napetalactol⁵⁸. Recent structural studies on the enzyme from us and others have shown that the structure and

catalytic mechanism of IS is similar to that of progesterone 5 β -reductase (P5 β R)^{59,60}. However, these enzymes exhibit a remarkable difference in their substrate specificities; IS acts on 10-oxogeranial whereas P5 β R reduces progesterone to pregnane-3,20-dione. Previously, two loops L1 (residues from 156 to 167) and L2 (residues from 350 to 361) at the opening of the substrate binding pocket have been suggested to be involved in conferring ligand specificity of the enzyme⁶¹. However, the structural basis of specificity generation of these class of enzymes is still not clear.

Furthermore, in the iridoid biosynthesis pathway (Fig1.9), 10-HGO converts 10-hydroxygeraniol to 10-oxogeranial which is subsequently catalyzed by IS to produce *cis-trans* nepetalactol. Earlier studies have shown that when 10-HGO and IS act in presence of 10-hydroxygeraniol and NADP⁺ yields *cis-trans* nepetalactol as a major product³⁶. In contrast, when 10-HGO and IS act in tandem on 10-hydroxygeraniol and 10-oxogeranial, *cis-trans* nepetalactol is not the major product. This indicates that 10-HGO and IS could interact spatiotemporally to transfer the product of the former as the substrate of the later to produced *cis-trans* nepetalactol as the major product. However, the structural basis of 10-HGO-IS interaction and the mechanism of resultant single product formation is not known.

Similar to those found in plants, SDRs from eukaryotes regulate complex biological processes. The complexity gets compounded when SDR domain is part of multi-domain proteins. For example, WWOX a recently identified member of NADP⁺/NADPH-dependent oxidoreductases plays an important role in various biological processes like cell migration, cell differentiation, apoptosis and steroid biogenesis⁵⁰. In addition to the SDR domain, this protein contains two WW domains (Domain with two conserved Trp residues), which are shown to be involved in regulating diverse cellular signaling events. Alterations at the genomic level or expression levels of WWOX has been implicated in several cancers. WWOX's localization and presence of SDR domain suggest that it is also involved in steroid metabolism⁵⁰. However, the enzymatic mechanism of WWOX and its substrate/s specificity is still elusive. Thus elucidating the structure of WWOX and delineating its cognate substrates would provide insights on the functioning of this protein and also facilitate designing inhibitors that could potentially act as anti-metastatic drugs.

1.8 Objectives of the thesis:

- To elucidate the structure of IS in order to understand the enzyme catalytic mechanism and its substrate specificity.
- To elucidate the complex structure of 10-HGO and IS in order to understand the handshake interaction of two enzymes for the production of nepetalactol as a major product.
- To elucidate the structure of WWOX in order to understand its catalytic mechanism, substrate specificity, and the role of WW domains in regulating reductase function.

1.9 Bibliography:

1. Hsu, P. P. & Sabatini, D. M. Cancer cell metabolism: Warburg and beyond. *Cell* (2008). doi:10.1016/j.cell.2008.08.021
2. Aw, T. Y. Molecular and cellular responses to oxidative stress and changes in oxidation-reduction imbalance in the intestine. in *American Journal of Clinical Nutrition* (1999).
3. Berry, S. The chemical basis of membrane bioenergetics. *J. Mol. Evol.* (2002). doi:10.1007/s00239-001-0056-3
4. Müller, M. *et al.* Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol. Mol. Biol. Rev.* (2012). doi:10.1128/MMBR.05024-11
5. Joyce, S. A., Lango, L. & Clarke, D. J. in *Advances in Applied Microbiology* (2011). doi:10.1016/B978-0-12-387048-3.00001-5
6. Sellés Vidal, L., Kelly, C. L., Mordaka, P. M. & Heap, J. T. Review of NADPH-dependent oxidoreductases: Properties, engineering and application. *Biochimica et Biophysica Acta - Proteins and Proteomics* (2018). doi:10.1016/j.bbapap.2017.11.005
7. Lee, T. *et al.* Probing Enzyme Specificity. *Acta Chem. Scand.* (2008). doi:10.3891/acta.chem.scand.50-0697
8. Andreini, C., Bertini, I., Cavallaro, G., Holliday, G. L. & Thornton, J. M. Metal ions in biological catalysis: From enzyme databases to general principles. *J. Biol. Inorg. Chem.* (2008). doi:10.1007/s00775-008-0404-5
9. Paul, P. E. V., Sangeetha, V. & Deepika, R. G. in *Recent Developments in Applied Microbiology and Biochemistry* (2018). doi:10.1016/b978-0-12-816328-3.00009-x
10. Meijers, R. *et al.* On the Enzymatic Activation of NADH. *J. Biol. Chem.* (2001). doi:10.1074/jbc.M010870200
11. Stewart, J. D. in *Future Directions in Biocatalysis* (2007). doi:10.1016/B978-044453059-2/50012-1

12. Holm, A. K. *et al.* Metabolic and transcriptional response to cofactor perturbations in *Escherichia coli*. *J. Biol. Chem.* (2010). doi:10.1074/jbc.M109.095570
13. Adams, M. J. *et al.* Structure of lactate dehydrogenase at 2.8 Å resolution. *Nature* (1970). doi:10.1038/2271098a0
14. HANUKOGLU, I. & GUTFINGER, T. cDNA sequence of adrenodoxin reductase: Identification of NADP-binding sites in oxidoreductases. *Eur. J. Biochem.* (1989). doi:10.1111/j.1432-1033.1989.tb14671.x
15. Bottoms, C. A., Smith, P. E. & Tanner, J. J. A structurally conserved water molecule in Rossmann dinucleotide-binding domains. *Protein Sci.* (2009). doi:10.1110/ps.0213502
16. Hanukoglu, I. Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochem. Mol. Biol. Educ.* (2015). doi:10.1002/bmb.20849
17. Kavanagh, K. L., Jörnvall, H., Persson, B. & Oppermann, U. Medium- and short-chain dehydrogenase/reductase gene and protein families: The SDR superfamily: Functional and structural diversity within a family of metabolic and regulatory enzymes. *Cellular and Molecular Life Sciences* (2008). doi:10.1007/s00018-008-8588-y
18. Crichton, R. R. in *Biological Inorganic Chemistry* (2008). doi:10.1016/b978-044452740-0.50012-0
19. Oberleitner, N., Peters, C., Rudroff, F., Bornscheuer, U. T. & Mihovilovic, M. D. In vitro characterization of an enzymatic redox cascade composed of an alcohol dehydrogenase, an enoate reductases and a Baeyer-Villiger monooxygenase. *J. Biotechnol.* (2014). doi:10.1016/j.jbiotec.2014.04.008
20. Kallberg, Y., Oppermann, U. D. O. & Jörnvall, H. relationships : A large family with eight clusters common to human , animal , and plant genomes. 636–641 (2002). doi:10.1110/ps.26902.spite
21. Moummou, H., Kallberg, Y., Tonfack, L. B., Persson, B. & van der Rest, B. The Plant Short-Chain Dehydrogenase (SDR) superfamily: Genome-wide inventory and diversification patterns. *BMC Plant Biol.* (2012). doi:10.1186/1471-2229-12-219
22. Kavanagh, K., Jörnvall, H., Persson, B. & Oppermann, U. The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *FASEB J.* (2008).
23. Kirby, A. J. Enzymatic Reaction Mechanisms. Von Perry A. Frey und Adrian D. Hegeman. *Angew. Chemie* (2007). doi:10.1002/ange.200785514
24. Baker, P. J. *et al.* Active site dynamics in the zinc-dependent medium chain alcohol dehydrogenase superfamily. *Proc. Natl. Acad. Sci.* (2009). doi:10.1073/pnas.0807529106
25. Liu, X. *et al.* Two novel metal-independent long-chain alkyl alcohol dehydrogenases from *Geobacillus thermodenitrificans* NG80-2. *Microbiology* (2009).

doi:10.1099/mic.0.027201-0

26. Kavanagh, K. L., Klimacek, M., Nidetzky, B. & Wilson, D. K. Crystal structure of *Pseudomonas fluorescens* mannitol 2-dehydrogenase binary and ternary complexes: Specificity and catalytic mechanism. *J. Biol. Chem.* (2002). doi:10.1074/jbc.M206914200
27. Nahoum, V. *et al.* Structure of the Human 3 α -Hydroxysteroid Dehydrogenase Type 3 in Complex with Testosterone and NADP at 1.25-Å Resolution. *J. Biol. Chem.* (2001). doi:10.1074/jbc.M105610200
28. Grimshaw, C. E., Bohren, K. M., Lai, C. J. & Gabbay, K. H. Human Aldose Reductase: Pk of Tyrosine 48 Reveals the Preferred Ionization State for Catalysis and Inhibition. *Biochemistry* (1995). doi:10.1021/bi00044a014
29. Bruce, N. C., Willey, D. L., Coulson, A. F. & Jeffery, J. Bacterial morphine dehydrogenase further defines a distinct superfamily of oxidoreductases with diverse functional activities. *Biochem. J.* (2015). doi:10.1042/bj2990805
30. Gaona-López, C., Julián-Sánchez, A. & Riveros-Rosas, H. Diversity and evolutionary analysis of iron-containing (Type-III) alcohol dehydrogenases in eukaryotes. *PLoS One* (2016). doi:10.1371/journal.pone.0166851
31. Montella, C. *et al.* Crystal structure of an iron-dependent group III dehydrogenase that interconverts L-lactaldehyde and L-1,2-propanediol in *Escherichia coli*. *J. Bacteriol.* (2005). doi:10.1128/JB.187.14.4957-4966.2005
32. Geu-flores, F. *et al.* cyclization in iridoid biosynthesis. 2–8 (2012). doi:10.1038/nature11692
33. Roussel, E., Drolet, M. C., Lavigne, A. M., Arsenault, M. & Couet, J. Multiple short-chain dehydrogenases/reductases are regulated in pathological cardiac hypertrophy. *FEBS Open Bio* **8**, 1624–1635 (2018).
34. May, S. W. Applications of oxidoreductases. *Curr. Opin. Biotechnol.* **10**, 370–375 (1999).
35. Martínez, A. T. *et al.* Oxidoreductases on their way to industrial biotransformations. *Biotechnology Advances* (2017). doi:10.1016/j.biotechadv.2017.06.003
36. Krithika, R. *et al.* Characterization of 10-Hydroxygeraniol Dehydrogenase from *Catharanthus roseus* Reveals Cascaded Enzymatic Activity in Iridoid Biosynthesis. *Sci. Rep.* **5**, 1–6 (2015).
37. Tundis, R., Loizzo, M. R., Menichini, F., Statti, G. a & Menichini, F. Biological and pharmacological activities of iridoids: recent developments. *Mini Rev. Med. Chem.* **8**, 399–420 (2008).
38. Miettinen, K. *et al.* Erratum: The seco-iridoid pathway from *Catharanthus roseus* (Nature Communications (2014) 5:3606 DOI: 10.1038/ncomms4606). *Nature Communications* (2014). doi:10.1038/ncomms5175
39. Yang, J. & Zhang, W. WWOX tumor suppressor gene. *Histol. Histopathol.* **23**, 877–

882 (2008).

40. Paige, A. J. W. *et al.* WWOX: A candidate tumor suppressor gene involved in multiple tumor types. *Proc. Natl. Acad. Sci.* (2002). doi:10.1073/pnas.191175898
41. Ludes-Meyers, J. H. *et al.* WWOX binds the specific proline-rich ligand PPXY: Identification of candidate interacting proteins. *Oncogene* (2004). doi:10.1038/sj.onc.1207680
42. Ingham, R. J. *et al.* WW domains provide a platform for the assembly of multiprotein networks. *Mol. Cell. Biol.* (2005). doi:10.1128/MCB.25.16.7092-7106.2005
43. Chen, H. I. & Sudol, M. The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc. Natl. Acad. Sci.* (2006). doi:10.1073/pnas.92.17.7819
44. Ingham, R. J., Gish, G. & Pawson, T. The Nedd4 family of E3 ubiquitin ligases: Functional diversity within a common modular architecture. *Oncogene* (2004). doi:10.1038/sj.onc.1207436
45. Aasland, R. *et al.* Normalization of nomenclature for peptide motifs as ligands of modular protein domains. *FEBS Letters* (2002). doi:10.1016/S0014-5793(01)03295-1
46. Aqeilan, R. I. *et al.* Functional association between Wwox tumor suppressor protein and p73, a p53 homolog. *Proc. Natl. Acad. Sci.* (2004). doi:10.1073/pnas.0400805101
47. Schuchardt, B. J. *et al.* Molecular origin of the binding of WWOX tumor suppressor to erbb4 receptor tyrosine kinase. *Biochemistry* (2013). doi:10.1021/bi400987k
48. Bouteille, N. *et al.* Inhibition of the Wnt/B-catenin pathway by the WWOX tumor suppressor protein. *Oncogene* **28**, 2569–2580 (2009).
49. Jin, C. *et al.* PKA-mediated protein phosphorylation regulates ezrin-WWOX interaction. *Biochem. Biophys. Res. Commun.* (2006). doi:10.1016/j.bbrc.2006.01.023
50. Del Mare, S., Salah, Z. & Aqeilan, R. I. WWOX: Its genomics, partners, and functions. *Journal of Cellular Biochemistry* (2009). doi:10.1002/jcb.22298
51. Abu-Odeh, M. *et al.* Characterizing WW domain interactions of tumor suppressor WWOX reveals its association with multiprotein networks. *J. Biol. Chem.* (2014). doi:10.1074/jbc.M113.506790
52. Farooq, A. Structural insights into the functional versatility of WW domain-containing oxidoreductase tumor suppressor. *Exp. Biol. Med.* (2015). doi:10.1177/1535370214561586
53. Bednarek, A. K. *et al.* WWOX, a novel WW domain-containing protein mapping to human chromosome 16q23.3-24.1, a region frequently affected in breast cancer. *Cancer Res.* (2000).

54. Gao, G. & Smith, D. I. WWOX, large common fragile site genes, and cancer. *Exp. Biol. Med.* (2015). doi:10.1177/1535370214565992
55. Li, J., Liu, J., Ren, Y. & Liu, P. Roles of the WWOX in pathogenesis and endocrine therapy of breast cancer. *Exp. Biol. Med.* (2015). doi:10.1177/1535370214561587
56. Sudol, M., Sliwa, K. & Russo, T. Functions of WW domains in the nucleus. *FEBS Letters* (2001). doi:10.1016/S0014-5793(01)02122-6
57. Chiang, M.-F., Chou, P.-Y., Wang, W.-J., Sze, C.-I. & Chang, N.-S. Tumor Suppressor WWOX and p53 Alterations and Drug Resistance in Glioblastomas. *Front. Oncol.* (2013). doi:10.3389/fonc.2013.00043
58. Lindner, S., Geu-Flores, F., Bräse, S., Sherden, N. H. & O'Connor, S. E. Conversion of substrate analogs suggests a Michael cyclization in iridoid biosynthesis. *Chem. Biol.* **21**, 1452–6 (2014).
59. Munkert, J. *et al.* Iridoid synthase activity is common among the plant progesterone 5beta-reductase family. *Mol. Plant* **8**, 136–152 (2014).
60. Sandholu, A. S. *et al.* Dynamics of loops at the substrate entry channel determine the specificity of iridoid synthases. *FEBS Lett.* (2018). doi:10.1002/1873-3468.13174
61. Kries, H. *et al.* Structural determinants of reductive terpene cyclization in iridoid biosynthesis. **12**, 6–11 (2016).

Chapter2

"Would it not be better if one could really 'see' whether molecules...were just as experiments suggested?"

-Dorothy Hodgkin

Materials & methods

This chapter mainly deals with the methodologies used in structure determination and biochemical characterization of proteins to address the biological questions. The two popular methods for the structure determination, X-ray crystallography and Cryo-Electron microscopy (Cryo-EM), were employed for the current studies. Additionally, various other methods including molecular, analytical, biochemical and computational methods have also been discussed. The complete list of materials used in this study are listed in the *materials section*.

2.1 Materials:

Table 2.1 Sources of chemicals

Chemicals	Supplier
20x SYBR Green	Invitrogen, Germany
2 β -mercaptoethanol	Himedia
Acrylamide	Sigma-Aldrich, USA
Agarose	SRL, India
Ampicillin	MP Biomedicals, USA
Bromophenol blue	Hi-Media, India
Calcium chloride	MP Biomedical
Casaminoacids	MP Biomedicals
cOmplete EDTA-free protease	Roche
Coomassie Brilliant Blue R-250	SRL, India
Crystallization suits	Quiagen
Dithiothreitol	MP Biomedicals
Dimethylsulfoxide (DMSO)	MP Biomedicals
Ethanol	Merck Millipore, USA
Ethylenediamine Tetraacetic acid (EDTA)	MP Biomedicals, USA
Gel extraction and PCR purification kit	Qiagen, Germany
Glutathione reduced	MP Biomedicals, USA

Glycerol	SRL, India
Hydrochloric acid	Merck
His60 Ni gravity column	Roche, USA
Hexylene glycol	Sigma
Isopropanol	MP Biomedicals, USA
Isopropyl- β -D-thiogalactopyranoside	MP Biomedicals, USA
Kanamycin	Gibco
LB media	Hi-Media, India
LB agar	Hi-Media, India
Magnesium sulfate	Sigma
Minimal media	Molecular dimensions
Nicotinamide adenine dinucleotide phosphate	SRL, India
Noble agar	Sigma
Oligonucleotide primers	Eurofines, India
pET vector system	Clontech, USA
Poly Ethylene glycol	Sigma
Quick change lightning kit	Agilent
Restriction endonucleases	NEB, UK
Tert-butyl methyl ether	SRL
Thiamine	Sigma
Tris-HCl	MP Biomedicals, USA
Trimethoprim	Sigma
Tri-sodium citrate	Sigma
Selenomethionine	Molecular Dimensions
Silicon oil	Sigma
Sodium chloride	MP Biomedicals, USA
Sodium dodecyl sulfate	Sigma
Sodium hydroxide	Sigma
Sodium sulfate	Himedia
Sucrose	Sigma
T4-DNA ligase	Promega, USA
<i>Taq</i> DNA polymerase	NEB, UK

All other chemicals used were of analytical grade.

2.2 Methods:

2.2.1 Molecular biology methods:

2.2.1.1 Primers:

The complete nucleotide sequences of gene of interest were obtained from NCBI database, which was used for designing the primers. Short sequences that match the beginning and the end of gene were determined. The optimal primer length was considered to be 22bp-30bp. Other parameters like T_m and GC content of primer were determined using online Oligonucleotide Properties Calculator (Oligocal)¹. Designed short sequences were synthesized using oligo synthesizer. Author procured all the oligos from Sigma and Eurofins Bangalore.

2.2.1.2 Polymerase chain reaction (PCR):

PCR involves an enzymatic amplification of DNA, using DNA polymerase and a pair of primers; and synthesize a new strand of DNA which is complementary to the template strand². PCR reactions were performed in 30 µl reaction mixture containing 10X manufacture's polymerase buffer, 2 mM dNTPs, 30-50 ng of DNA template, 0.1 µmoles of primers, 1 U of DNA polymerase enzyme and RNase free water. PCR application was performed in thermal cycler (Veriti, Applied Biosystems) with the specific cycling profile as follows:

Step 1	Denaturation	95	5 min	Cycle 1
Step 2	Denaturation	95	30 Sec	Cycle 10
	Annealing	61	As per gene length	
Extension	68	(1kb/min)		
Extension	72	30 Sec		
Step 3	Denaturation	95	30 Sec	Cycle 15
	Annealing	58	As per gene length	
Extension	68	(1kb/min)		
Extension	72	30 Sec		
Step 4	Denaturation	95	30 Sec	Cycle 30
	Annealing	56	30 Sec	

	Extension Extension	68 72	As per gene length (1kb/min) 30 Sec	
Step 5	Extension	72	10 min	Cycle 1
Step 6	Hold	4	Infinite	

After PCR amplification samples were subjected to agarose gel electrophoresis. The amplified PCR products were extracted from agarose gel and subjected to restriction digestion followed by cloning.

2.2.1.3 Purification of PCR products:

Usually, in PCR reactions, along with expected size of amplicons some nonspecific amplicons and primer dimers are synthesized. The amplicons of expected size were purified by separating them on agarose gel and extracted using MiniGel extraction kit (Quiagen, Germany).

2.2.1.4 Ligation of vector and insert:

PCR amplified products and specific plasmid vectors are subjected to restriction digestion using set of restriction endonuclease (RE) enzymes prior to ligation. The reaction mixture for ligation contains RE digested vector and PCR products, manufactures ligation buffer and ligase enzyme (Promega) with the final reaction volume of 10 µl. The reaction was mixed and incubated at room temperature for 2 hours or overnight at 4°C. After incubation, the reaction mixture was used for transformation.

2.2.1.5 USER cloning:

It is a ligase-free uracil DNA glycosylase (UDG) mediated cloning, which is also referred as USER cloning (Uracil-Specific Excision Reagent)³. In this method, primers single standard flanking ends containing deoxyuridine nucleotide (dU) allows ligase-free directional cloning. The DNA amplification reactions should carry in presence of polymerase like PfuTurbo Cx Hotstart (Agilent) which can incorporate a deoxyadenine complimentary to dU present in the primers. The USER™ enzyme (mixture of UDG and endonuclease VIII)

excise the dU residues to generate 3' single strand overhangs on amplified DNA fragments which can incorporate into a linear vector with complementary overhangs (Fig 2.1)⁴. USER based ligation approach provides many advantages including (a) long single-stranded extensions can be produced (b) unique single-stranded extensions facilitate the linear vector cannot ligate to form circular DNA (c) unique single-stranded extension permits the directional assembly of multiple DNA fragments into the vector³.

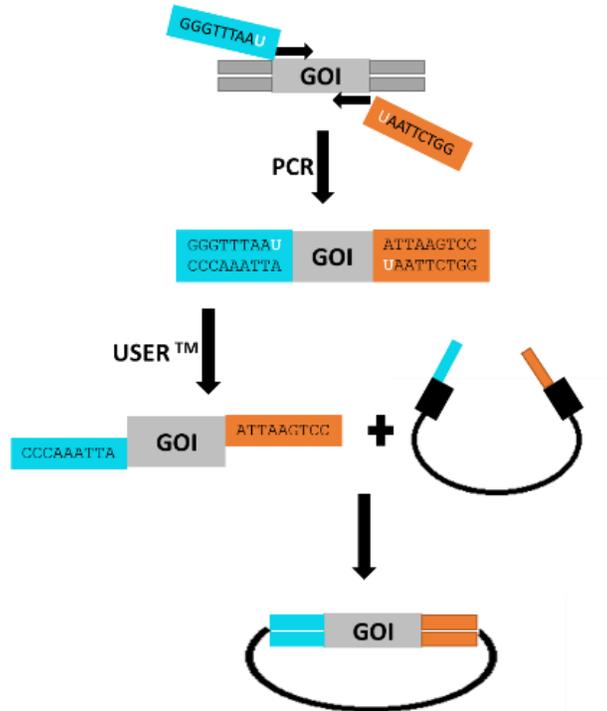


Fig2.1: Integration of a gene of interest (GOI) into the linear vector using USER cloning method

2.2.1.6 Transformation:

The ultra-competent cells of *E. coli* strains like DH5 α , B121* and B834 rare were prepared for transformation by Inoue method⁵. Competent cells were thawed on ice for 5 min and ligation reaction mixture was added. The mixture of cells and DNA were gently flicked and placed on ice for 30 min. Further, the cells were heat-shocked at 42°C in a water bath for 60 sec, followed by a 2 min incubation on ice. Later, 1000 μ l of LB broth was added to transformation reaction, which was then incubated on a shaker at 37°C for an hour. The transformation mix was then plated onto antibiotics containing agar plates. The plates were incubated at 37°C overnight. The success of transformation was determined by isolation of

plasmid from transformed *E. coli* and followed by restriction digestion and plasmid sequencing.

2.2.1.7 Purification of plasmid DNA:

The plasmid purification is based on alkaline lysis, after which there is binding of plasmid DNA to manufacturer's anion exchange matrix columns (Quiagen) under low-salt and pH conditions. Further, the matrix was washed with ethanol containing a medium salt buffer to remove protein, RNA and other impurities. Finally, elution is done in warm nuclease-free water.

2.2.1.8 Quantification of DNA:

The DNA was quantified by Nano-drop spectrophotometer (NanoVue GE Health Care). The concentration was determined by measuring the maximum absorption of double stranded DNA at 260 nm. As protein has an absorption peak at 280 nm, the ratio of 260:280 helps to determine the purity of the isolated DNA. The ratio between 1.8 and 2.0 indicates pure DNA⁶.

2.2.1.9 Site-directed mutagenesis:

Selective mutants were amplified by PCR using the QuikChange Lightning site-directed mutagenesis kit (Agilent) and mutation containing primer pair. All primers were synthesized by Eurofins Bangalore. The plasmid containing wild-type gene was used as a template. PCR was performed in thermal cycler with following programme: 1 cycle of 30 sec at 95°C denaturation followed by 18 cycles of 10 sec at 98°C denaturation; 30 sec at 55°C annealing; 5 min at 68°C extension followed by hold at 4°C. Amplified products were subjected to DPN1 digestion for 2 hrs at 37°C followed by the transformation in DH5 α competent cells.

2.2.1.10 Clone conformation and plasmid sequencing:

Before sequencing, purified plasmid samples were checked for the presence of gene of interest by restriction digestion with specific RE enzymes. The RE digestion reaction mixture contains 0.5 to 1 μ g of plasmid, 1X manufactures buffer and 1 U of RE enzyme. The reaction mixture was incubated at 37°C for 2-3 hrs. The digestion pattern was analyzed using 0.7% agarose gel. For those clones with correct RE digestion pattern will be given for

Sangers sequencing at Eurofins Bangalore. The sequencing chromatograms were analyzed using chromas 2.1 software.

2.2.2 Analytical methods:

2.2.2.1 Electrophoresis:

Electrophoresis describes the charged particles (ions) migration and separation under the influence of an electric field through the matrix of agarose or polyacrylamide⁷.

2.2.2.1.1 Agarose gel electrophoresis:

Agarose gels were prepared by melting agarose in TAE buffer (Tris-acetate, 40 mM; EDTA, 1 mM; pH 8.0). To cast and run the gel, Bio-Rad Mini electrophoresis cells were used (Bio-Rad, USA). The concentration of agarose varied according to the size of DNA being separated. The solidified gel was placed into the electrophoresis cell. TAE buffer was poured on to the gel up to approximately 5 mm over the gel. DNA samples were loaded into the wells after being mixed with gel loading dye (0.25% Xylene cyanol FF, 0.25% Bromophenol blue and 40% Sucrose). As molecular weight markers, 200 ng ladder (New England Biolabs, UK) was used. DNA samples were electrophoresed through the gel at 5 V/cm for 40 min. The DNA was visualized under ultraviolet light in a Gel Doc XR system (Bio-Rad, USA).

2.2.2.1.2 SDS-poly acrylamide gel electrophoresis:

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS PAGE) was carried out to fractionate denatured protein molecules according to their mass. All gels used in this study were lab made with either 10% or 12% resolving gel, and 5% stacking gel in 1.0 mm mini gel cassettes (Bio-Rad, USA). To perform SDS-PAGE, gels were first secured in a cassette with 1X SDS Running buffer filled in both cathode chamber and anode chamber. Before loading, protein samples were mixed with 1X SDS sample buffer and denatured at 95°C for 5 min. Electrophoresis was performed at 150 V for 2 hrs. Gels were then stained by InstantBlue Protein Stain (Lab made) to visualize protein bands.

2.2.2.2 Size-exclusion chromatography-multi angle light scattering (SEC-MALS):

Size-exclusion chromatography (SEC) coupled with multi-angle light scattering (MALS) detectors are the widely used and sensitive techniques for determining absolute molecular weight⁸. SEC is used to separate the molecules based on their hydrodynamic volume. MALS uses the scattered light intensity and the angular dependence to measure size of the molecules (root mean square radius, r_g) and absolute molar mass in solution⁹. Combining SEC and MALS allows more accurate mass determination compared to either method alone. SEC-MALS analysis of the purified protein was performed on a Superdex 200 10/300 GL column (GE Healthcare) connected to an Agilent HPLC system coupled with a refractive index detector (Wyatt Optilab TrEX) and an 18-angle light scattering detector (Wyatt Dawn HELIOS II). The molecular weight was calculated using ASTRA software (WyattTechnologies).

2.2.3 Biochemical methods:

2.2.3.1 Protein expression and purification:

2.2.3.1.1 Expression protocol:

The positive clones confirmed from plasmid sequencing were transformed into *E. coli* BL21* and B834 rare strains. Transformed colony was inoculated into 5 ml LB broth with 100 μ g/ml ampicillin antibiotic and incubated at 37°C at 180 rpm on rotary shaker overnight. From this starter culture, 1 ml was inoculated into 100 ml LB broth containing 100 μ g/ml ampicillin and incubated at 37°C on shaker till OD₆₀₀ reached to 0.6-0.8. Before induction, 1 ml culture aliquot from this was kept on ice as uninduced control. For induction, isopropyl 1-thio-D-galactopyranoside (IPTG) at the final concentration of 0.5 mM was added to the remaining culture and incubated at 37°C with shaking for 4-5 hrs. The samples were centrifuged at 3000g for 15 min at 4°C, supernatant containing media was discarded and the cell pellet was stored at -80°C till further use.

2.2.3.1.2 Expression analysis:

The pellet was resuspended in Tris-HCl buffer and was sonicated for 30 sec. The lysate was centrifuged at maximum speed. The obtained clear supernatant was loaded on 10% SDS-PAGE to check the expression of the protein.

2.2.3.1.3 Analysis of solubility:

The solubility of over expressed protein was determined by resuspending the pellet into 5 ml buffer and to this 0.75 mg/ml of lysozyme was added. The lysate was sonicated followed by centrifuged at 12000 g at 4°C for 20 min. The clear supernatant as well as settled pellet were examined for the presence of over expressed protein. If the protein is observed in clear supernatant indicates protein is soluble.

2.2.3.1.4 Large scale expression and purification:

Protein expression was found to be optimum expression in *E.coli* strain B834 (rare). For large-scale expression, cells were grown in 4L LB media at 37°C to an OD₆₀₀ of 0.5 and further induced with 0.5 mM IPTG and incubated for 14 hrs at 18°C. The culture was harvested by centrifugation. The cells were lysed using sonication after suspending the cell pellet in chilled lysis buffer containing one tablet of cOMplete EDTA-free protease inhibitor and a pinch of lysozyme. The lysate was filtered and loaded on to the GST affinity column (GE Healthcare Biosciences, USA) or Ni-NTA column for affinity binding. The column was washed with wash buffer and the protein was eluted with 20 mM glutathione reduced (Sigma) or 200 mM imidazole prepared in the wash buffer. To cut the tag, the elution fractions were dialyzed along with the PreScission protease or TEV protease (purified in our lab) overnight at 10°C. The tag and the protease (GST tag/6xHis tag & PreScission/TEV protease) were removed by the desalting the dialyzed fraction followed by a second affinity column. To improve the purity of the protein, size exclusion chromatography was performed using sephacryl or superdex columns (GE Healthcare Biosciences, USA). The peak fractions under specific molecular weight were collected and concentrated to perform crystallization and biochemical experiments.

2.2.3.2 Thermal shift assay:

Thermal shift assay measures the change in thermal denaturation temperature of a protein under varying conditions like buffer, salt concentration, pH, and the presence of various ligands. By using protein thermal shift dye (PTS) and real-time PCR system one can measure the melting temperature as a function of change in the fluorescence¹⁰. When the protein unfolds with an increase in temperature, the dye binds to exposed hydrophobic surfaces, resulting in an increase in fluorescence. This phenomenon can be used to confirm protein folding state like folded or aggregated. In case of aggregated proteins, there will be no fluorescence increase at all because protein is already in the aggregated form before heating¹¹. The reaction mixture consists of 5 μ M of protein, 8x PTS dye, 5 μ l manufacturer's assay buffer and buffer in which protein eluted to the final reaction volume of 20 μ l. The following program was used for the melt curve analysis: STEP1: 25 °C for 2 min with 100% ramp rate followed by 99°C with a ramp rate of 1% in continuous mode. The obtained melt curve data was analyzed using Protein Thermal Shift Software (Thermo Fisher).

2.2.3.3 GC-MS assay:

GC-MS based assays have been performed to analyze enzyme activity. The reaction mixture was consisting of 100-200 μ g of the enzyme (IS/10-HGO/WWOX), 200 μ M substrate and 200 μ M NADPH/NADP⁺ in 500 μ l of buffer containing 20 mM MOPS pH-7.0, 10% glycerol. Reaction mixtures were incubated at 25-30°C for 30 min followed by adding 1 ml of Tert-Butyl Methyl Ether (TBME) in order to terminate the reaction. Compounds were extracted in TBME, 1-5 μ L of extracted compounds were injected into gas chromatography (GC)–mass spectrometry (MS) for analysis using following methods: equilibration time for 2 min followed by 70°C for 0 min then 10°C/min to 230°C for 0 min then 2°C/min to 250°C for 0 min then 10°C/min to 280°C for 6 min and the overall run time is 39.5 min, equilibration time for 0.5 min. Oven program is as following; 60°C for 0 min then 20°C/min to 120°C for 0 min then 2.5°C/min to 160°C for 0 min then 20°C/min to 190°C for 2.5 min. The separation was performed on Rtx-5ms column (30 m x 0.25 mm x 0.25 μ m) using helium as carrier gas at 1 ml/min and with an injector temperature of 220°C.

2.2.3.4 Spectrophotometric assay:

The enzyme specific activity studies of *wild* type and mutant proteins were measured at the absorption edge (340 nm) of the cofactor NADPH. For the determination of Michaelis-Menten parameters with substrate, kinetics of NADPH/NADP⁺ consumption was determined spectrophotometrically at 340 nm. The reaction mixture contains 200 μM NADPH/NADP⁺, 10 μM to 100 μM of the substrate, 2.5 μM (IS/P5βR/WWOX) enzyme in a total volume of 1 ml buffer (20 mM MOPS pH 7.0, 10% glycerol). The reaction was kept at room temperature (24°C) and initiated by addition of enzyme. The K_{cat} and K_m value of the reaction plots were non-linearly fit to the Michaelis-Menten equation using GraphPad Prism software¹². Data were collected for 5 min at 1 sec interval. All the assays were performed in triplicates.

2.2.4 Macromolecular X-ray crystallography:

Macromolecular crystallography is a popular technique used to study the structure of biological molecules such as proteins, viruses and nucleic acids (RNA and DNA) at the atomic resolution. Since there is a high degree of correlation between structure and function of biomolecules, structural information helps to elucidate the detailed mechanism of these macromolecules. In this technique, proteins are crystallized under regulated conditions. The crystals consist of multiple copies of the molecule arranged periodically in the 3-dimensional lattice. When the X-rays are impinged on crystals, they are scattered by the atoms in equivalent positions in the crystal lattice and are projected as sharp intense spots on X-ray film. The macromolecular structure can be determined by the analysis of intensities and positions of diffraction spots¹³. The complete flow diagram illustrates a series of experimental procedures carried out during structure determination by X-ray crystallography is depicted in Fig 2.2¹⁴. The major limitations of this technique include the ability to get crystals with diffractable size and quality. Additionally, even after obtaining good diffraction data, getting phase information is another bottleneck in X-ray crystallography. Therefore, crystallization and phase determination are considered to be the rate-limiting steps in macro molecular crystallographic studies.

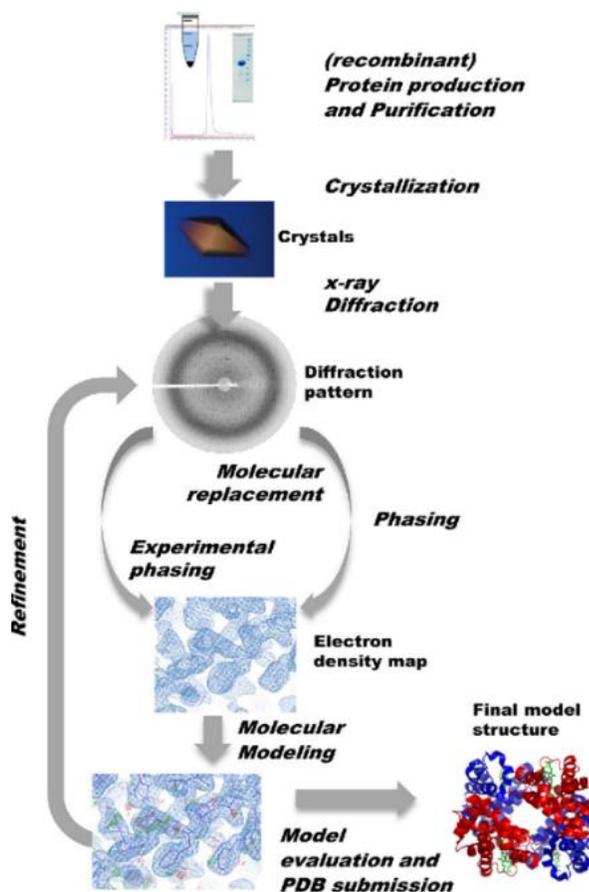


Fig 2.2: Work flow for structure determination of protein¹⁴.

2.2.4.1 Protein crystallization:

In protein crystallography, crystallization is the foremost, and often the rate-limiting step. The success of obtaining crystals depends on the quality of the protein. Over the years, a number of methods have been developed to grow protein crystals; vapour diffusion, free interface diffusion, batch and dialysis methods are most commonly used^{15,16}. The main aim of this experiment is to supersaturate the protein that triggers nucleation and crystal growth. Crystallization phase diagram explains the phenomenon of protein crystallization (Fig 2.3)¹⁷. When a protein is in below saturated condition, it is fully soluble and will never attain crystal. Precipitating agents like polyethylene glycol (PEG), ethylene glycol or a salt decrease the solubility of the protein. No crystal formation is observed if the supersaturation is very high and the protein precipitates. Hence, moderate supersaturation is considered to be a labile state for nucleation which drives the crystal formation. However, little below the liable state of supersaturation, where the nucleation occurs, is considered to be metastable. This

condition is optimum for the crystal growth¹⁷. Therefore, crystallization process demands a lot of optimization. This can be achieved by screening of temperature, pH, protein concentration and precipitant type which promote crystal growth. Several commercial screening kits are available with multiple variables that could lead to successful crystallization^{18,19}. This author has mainly employed micro batch method and vapour diffusion methods for crystallization studies. In micro batch method, crystallizing reagents were mixed with protein solution under the layer of silicon oil or paraffin oil in 72 well plates²⁰. In the vapour diffusion method, 0.5-1 ml of the precipitant solution is pipetted in 24-well plate. Then, 1-5 μ l of reservoir solution is mixed with the equal volume of a protein and allowed to equilibrate against precipitant solution in reservoir through vapour diffusion between them²¹. Hanging drop and sitting drop methods are most popular for crystal screening and optimization²².

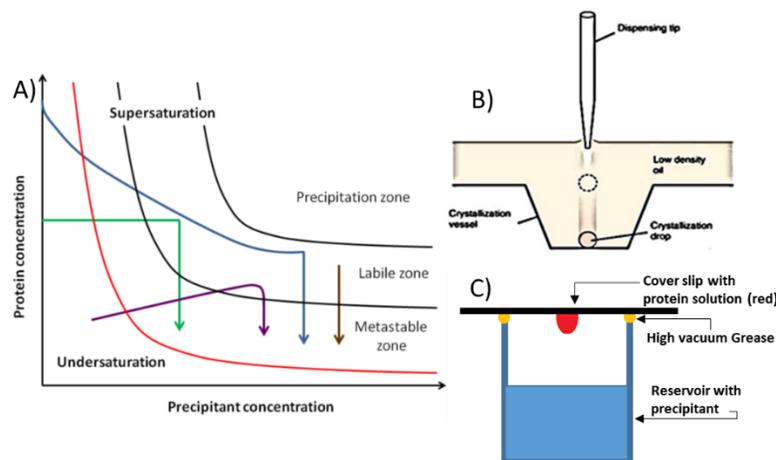


Fig 2.3: A) Phase diagram of crystallization²³ B) Micro batch under oil crystallization method²⁴ C) Hanging drop vapour diffusion method.

2.2.4.2 Data collection:

There are two main sources of X-ray radiation used for protein crystallography; laboratory X-rays or home source; and the synchrotron radiation. Typically these sources consist of three components, a radiation source, optical elements and a detector²⁵. The source may be a copper or molybdenum rotating anode in case of home source, bending magnet or an insertion device (undulator) in case of synchrotrons. Optical elements consists of monochromators for wavelengths selection, collimating slits and focusing mirrors²⁶. Home sources and most synchrotron beamlines are equipped with Hybrid Photon Counting (HPC)

detectors. More sensitivity and fast read out time of these detectors allows data collection within a short period of time. However, modern synchrotrons are well advanced and more efficient than conventional laboratory X-ray sources. It is primarily due to their high flux and tenability of wavelength. One of the major drawbacks of protein crystallography is the radiation damage of crystals under high intensity X-rays, which can be circumvented by collecting diffraction data under cryogenic temperatures (at about 100K)²⁷. Prior to placing the crystal at cryogen, it should be soaked with a cryoprotectant. It prevents ice formation in crystal during cryogenic cooling²⁸. The X-ray diffraction data collection process involves: 1) mounting of the crystal on the goniometer and aligning it with X-ray beam, 2) initial indexing to determine the strategy for data collection based on test data, and 3) implementing the optimum strategy for data collection. These steps may be performed manually at home source and automatically at synchrotrons. The second step includes collecting a few diffraction images and determining the crystal characteristics such as lattice type, unit cell dimensions and optimum crystal-detector distance for data collection. In case of a heavy atom containing crystals, fluorescence scan will be done to confirm the presence of heavy atom.

Data processing involves indexing, integration, scaling and merging of the diffraction data. Indexing provides the information about crystal orientation, unit cell dimensions, crystallographic symmetry and enables determination of Miller indices for the reflection present in different frames. Following this step, intensities of the spots will be measured. This measurement is done either by adding the photon counts in every pixel encompassing the spot or by creating an average 3D spot profile and then fitting observed photon counts against the average profile. The intensity of the spot obtained from the former technique is known as integrated intensity and from the later, it is known as profile fitted intensity. It has been shown that profile-fitted intensities are more accurate representation of the diffraction from the plane. The data obtained from integration is not on a common scale either due to changes in incident radiation intensity or radiation damage to the crystal. Those symmetry-equivalent reflections are combined through scaling and merging. One can determine the quality of diffraction data after scaling by looking at the parameters like resolution, signal-to-noise ratio, completeness, multiplicity and R_{merge} . The completeness of data corresponds to all the reflections have been recorded. Multiplicity is the average number of measurements for equivalent reflections and R_{merge} is a measure of the agreement between equivalent reflections²⁹ (Equation1).

$$R_{\text{merge}} = \sum_{hkl} \left| \frac{I_{hkl} - \bar{I}_{hkl}}{I_{hkl}} \right| \quad (1)$$

Where I_{hkl} is intensity of a reflection around their average \bar{I}_{hkl} . A good data set characterized by the R_{merge} value should be $\sim 10\%$ ²⁹. All the data presented in this work are indexed and integrated with XDS³⁰ and scaled with aimless. XDS is a data reduction program for single-crystal diffraction data collected by the rotation method. It contains three set of programs including XDS, XSCALE and XDSCONV. XDS program mainly performs indexing, integrating and scaling with the help of another set of eight programs including XYCORR, INIT, COLSPOT, IDXREF, DEFPIX, INTEGRATE and CORRECT. The unmerged scaled file obtained from XDS processed further using CCP4 program SCALA. This program scales and merges multiple observations of reflections into an average intensity.

2.2.4.3 Structure determination:

Once the scaling of diffraction data has been done, the further step is to calculate an electron density map. By employing the Fourier transformation it is possible to transform structure factors of reciprocal space to electron density map of real space or vice versa. The structure factor is a summation of the scattering intensities of each and every atom in the unit cell (Equation 2).

$$F_{hkl} = \sum_{j=1}^N F_j \exp 2\pi i (hx_j + ky_j + lz_j) \quad (2)$$

Where x_j, y_j, z_j are the fractional coordinates of the atom and F_j are the atomic scattering factor of j^{th} atom.

However, from the intensity of spots, we can obtain amplitude (Equation 3) of the wave scattered from the plane hkl but the phase information of the wave is lost in the diffraction experiment. This is referred as the "phase problem" in crystallography and determination of electron density map required for phase information³¹.

$$F_{hkl} = |F_{hkl}| e^{i\phi_{hkl}} \quad (3)$$

$$\rho(xyz) = 1/V \sum |F_{hkl}| e^{i\phi_{hkl}} \exp -2\pi i (hx + ky + lz) \quad (4)$$

Where F_{hkl} are structure factors, $|F_{hkl}|$ are the amplitudes, ϕ_{hkl} are the phase angles of hkl . The electron density map is referred as $\rho(xyz)$ (Equation 4). In case of small molecules, phases

of reflection can be calculated only from their diffraction data based on some strong reflections at certain areas of unit cell. These are direct methods of phase determination³¹. For large molecules like proteins, the phase problem can be accomplished by different ways like molecular replacement or experimental phasing.

2.2.4.4 Molecular replacement:

Molecular replacement (MR) is a popular method used in phasing the diffraction data from protein crystals. This method requires the availability of homologous structure of the protein, whose structure is being solved. MR involves the rotation and translation of the search model or homologous model which determine its orientation with respect to that of the unknown in a crystal unit cell. Typically, the homologous structure is expected to have at least 25% sequence identity for it to be used as search model³¹. Theoretically, the full MR search space is 6N dimensional, which includes three rotational dimensions and three translational dimensions. The 6N dimensional search becomes more complicated even if N in the single digits. The orientation of the search model in the target crystal is calculated by rotation function, which is an overlap function of Patterson functions (Equation 5) of the target protein and phasing model in different orientations. The rotation function can be expressed as (Equation 6)

$$P_{uvw} = \frac{1}{v} \sum_h \sum_k \sum_l \{ |F_{hkl}|^2 \cos[2\pi(hu + kv + lw)] \} \quad (5)$$

$$R(\emptyset, \varphi, x) = \int_{u,v,w} P^{target}(u, v, w) P^{model}\{(u, v, w) \times [\emptyset, \varphi, x]\} du dv dw \quad (6)$$

Where at each set of rotation angles \emptyset, φ and x the value of rotation function R is the integral of product of two Patterson functions; one is of the target molecule and other one is of the model. The maxima of this overlap will give rise to the best orientation of phasing model in the unit cell. Then the oriented model is translated to deduce its position relative to the position of unknown in the crystal unit cell. To determine the translational parameters translation function is used. The concept behind translation function is the same as for the rotation function. In this case, the correlation between the observed intensities and the Patterson cross-vectors of the symmetry-related molecules of the search model, as it is translated within the cell, is calculated. The translation function should have peaks at values corresponding to the translation vectors between the symmetry related molecules when the model is correctly positioned.

The translation [T] of the molecule X with an orientation [R] relative to the model M involves the maximization of the function (Equation 7)

$$T(\delta) = \int_{|u| < 2r} P_1(u + \delta) P_2([R]u - \delta) du \quad (7)$$

where P_1 and P_2 are Patterson functions and δ is a translation vector which is independent of the origin of the rotation axis and relates the "centres" of the two molecules. The three-dimensional structure solution can be obtained from a simple relation.

$$X = [R]M + [T] \quad (8)$$

where [R] is the appropriate rotation and [T] the required translation to correctly position the model in the unit cell.

There are several programs for searching molecular replacement solutions, that include Phaser³², AMoRe³³, MOLREP³⁴ *etc.* All the structures presented in this work were solved with the molecular replacement method using CCP4 PHASER module.

Phaser is a widely used program for molecular replacement. It mainly facilitates the automation of phase problem using maximum likelihood probability theory. In this theory, the consistency of the model with the data is assessed by the likelihood function which is defined as the probability of the data given the model³⁵. Phaser allows quick and sensitive rotation and translation likelihood searches. It also contains likelihood-based correction for anisotropy and a pruned tree search for multiple molecules allowed to solve complex MR problems.

2.2.4.5 Isomorphous replacement and anomalous dispersion:

Both isomorphous replacement and anomalous dispersion methods are *ab initio* phasing techniques and depend on the presence of heavy atoms in a crystal. Heavy atoms like Hg, Au, Pb and Ag *etc.* are incorporated into the protein crystal either by soaking them with heavy atom containing solution or by labelling the protein with selenium heavy atom by express protein in selenomethionine containing media. The differences in the intensities of isomorphs native crystals and heavy atom derivative crystals used to determine the positions of the heavy atoms³⁶. The contribution of the added heavy atom to the structure-factor amplitudes and phases is best represented by an Argand diagram (Fig 2.4). The amplitudes of a reflection are measured for the derivative crystal, $|F_{PH}|$ and the native crystal, $|F_P|$. For the

estimation of heavy atom, the isomorphous difference $|F_H|$ can be used (Equation 9). By employing Patterson or direct methods, one can determine the heavy atom's positions using structure-factor amplitudes.

$$|F_H| = |F_{PH}| - |F_P| \quad (9)$$

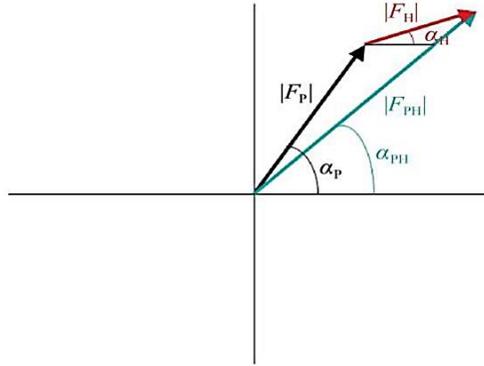


Fig 2.4 Argand diagram of isomorphous replacement

A second means of obtaining phases from heavy-atom derivatives takes advantage of the capability of heavy atom to absorb X-rays of a specified wavelength. Due to this, Friedel's law does not hold and the intensity of reflections of hkl and $-h-k-l$ are not same. This discrepancy in Friedel related reflections is called anomalous dispersion or anomalous scattering. When the wavelength of X-rays is near to the absorption edge of heavy-atom, it absorbs a fraction of the radiation and re-emit radiation with altered phase. The effect of anomalous scattering on a given structure factor F_{PH} in the heavy atom data can be showed in vector diagrams as one real (F_r) and the other imaginary (F_i) perpendicular contributions. In the Argand diagram, a structure factor for the heavy-atom derivative measured at wavelength λ_1 is represented as $F_{PH}^{\lambda_1}$ (green), where anomalous scattering does not occur. The same structure factor measured at a second X-ray wavelength λ_2 near the absorption edge of the heavy atom is represented as $F_{PH}^{\lambda_2}$ (red) which indicates that anomalous scattering alters the heavy-atom contribution to this structure factor. F_r (real, blue) and F_i (imaginary, cyan) are the vectors representing anomalous scattering contributions.

$$F_{PH}^{\lambda_1} = F_{PH}^{\lambda_2} + \Delta F_r + \Delta F_i \quad (10)$$

From the diagram (Fig 2.5), the magnitude of anomalous scattering contributions F_r and F_i for a given element are constant and independent of reflection angle θ . The phases of F_r and F_i

depend only upon the position of the heavy atom in the unit cell, so once the heavy atom is located by employing Patterson methods, the phases can be calculated.

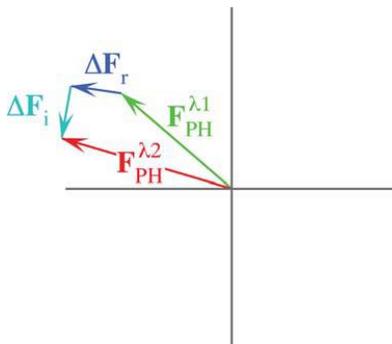


Fig 2.5 Argand diagram of anomalous dispersion

2.2.4.6 Model building and structure refinement:

Once an electron density map has been calculated, robust software is available for automatically building atomic models into the electron density. If this is not possible, programs like Coot help in manual model building by fitting amino acid residues into the electron density. However the resulting model may contain errors in geometry (backbone torsion angles, bond distances and bond angles), and also suffer from poor agreement with the observed structure factor amplitudes. Therefore, refinement process helps in maximizing the agreement of an atomic model with observed X-ray diffraction data. This in turn improves the phases and as well as electron density maps. The parameters that are accounted in the refinement process include B-factors, bulk solvent, constraints and restraints *etc* explained in below sections.

During refinement against X-ray data, refinement program search for the minimization of target function, which includes two components: one utilizes geometry and other utilizes experimental data (Equation 11)

$$f_{total} = f_{geom} + wf_{X-ray} \quad (11)$$

Where f_{total} is the total target function to be minimized consisting of functions which control the geometry of the model and the fit of the model parameters with that of experimental data, and w is a weight between the relative contributions of these two components. The fit of the model to the experimental data is monitored by ‘R-factor’. It is a measure of disagreement

between the amplitudes of the structure factors calculated from a crystallographic model (F_{cal}) and those from the X-ray diffraction data (F_{obs}) (Equation 12)

$$R = \frac{\sum \|F_{obs} - F_{cal}\|}{\sum F_{obs}} \quad (12)$$

In addition to R factor, free R-factor is used to monitor the progress of refinement and to cross-validate the model quality. In this process, the diffraction data set is divided into two sets: a test set (comprising 5-10% of randomly chosen reflections) and a complementary working set. The diffraction data in the working set are used in the normal crystallographic refinement process, whereas the test set data are not. During the course of successful refinement R_{free} is expected to decrease. Consequently, a high value of R_{free} and a concomitant low value of R may indicate an over fit model. Structures reported in this work were refined using Refmac5 refinement program³⁷.

2.2.4.6.1 Atomic displacement factor (ADP):

In crystal structures, Atomic Displacement Parameters (ADP) monitor the displacement of atoms which are frequently named as thermal factors or B factors³⁸. The atoms mean-square amplitude of displacements around their equilibrium positions (u^2) are,

$$B = 8\pi^2(u^2) \quad (13)$$

ADPs are determined from the atomic model refinement parameters against diffraction intensities, since the decrease of atomic form factors (f) are associated with diffraction angle (θ) is enhanced by an increase in ADP according to,

$$f = f_0 \times \exp(-B \times \sin^2\theta/\lambda^2) \quad (14)$$

Where f_0 is the atomic form factor at $B = 0 \text{ \AA}^2$ and λ is the X-ray wavelength. This implies that atoms with more ADPs contribute less to the diffraction intensities as compared to atoms with smaller ADPs. Consequently, it is possible to determine not only the positions of the atoms but, also their displacements. ADPs are usually refined isotropically or anisotropically depending upon the resolution of the data. In case of low resolution, one parameter is refined for all atoms in a group like, the side-chain atoms of a residue or the main-chain atoms of a residue by assuming that oscillation amplitudes are equal in all directions around the equilibrium position of the atom. In case of high resolution, the anisotropic B-factor refinement

can be employed which involves the refinement of six parameters that represent the B-factors for each atom³⁸.

2.2.4.6.2 Bulk solvent correction:

Protein crystals contain about 50% solvent of total crystal volume, most of which is disordered in the solvent channels present between the protein molecules of the crystal lattice³⁹. Thus, the electron densities of the protein molecules with typical values of $0.43 \text{ e}/\text{\AA}^3$, are surrounded by a continuous bulk solvent electron density ranging from $0.33 \text{ e}/\text{\AA}^3$ for pure water⁴⁰. The contribution of bulk solvent to structure factors is strong at low resolution and hence, disordered or bulk solvent correction is needed for better electron density map. One can correct it by volume that the solvent should fill that is identified by demarcating a solvent-accessible volume outside the van der Waals exclusion zone of protein. The optimal value for the average solvent density may be obtained by finding the minimum value of

$$\sum (|F_o| - |F_c(\text{total})|)^2 \quad (15)$$

$$\text{Where, } F_c(\text{total}) = F_c(\text{protein}) + K_{\text{sol}} F_c(\text{solvent}) \exp(-B_{\text{sol}}/4d^2) \quad (16)$$

K_{sol} is a scale factor, B_{sol} is an artificially large temperature factor applied to a flat solvent density, $F_c(\text{solvent})$ is a Fourier transform of a binary function M (solvent mask) whose value is 1 inside the solvent and 0 outside (protein region) and the summation is over low resolution reflections⁴⁰.

2.2.4.6.3 Constraints and restrains:

During the course of refinement, the data-to-parameter ratio will be improved by applying constraints (remove parameters) or restrains (adding data) to the atoms. Constraints are mathematical equations, which are related to two or more parameters or assign fixed numerical values to certain parameters, thus, reducing the number of independent parameters to be refined. In every structure site occupancy factors or constraints present. The sum of the occupancies is constrained to add up to 1.0 even for disordered atoms. Restraints include the assumptions which are used to introduce chemical or physical information into a refinement as additional experimental observations. Restraints are treated as data (with a standard uncertainty). If individual atoms coordinates represented by x , then each atom can be restrained as per the mathematical term:

$$E_{NCS} = w(x-x_{av})^2 \quad (17)$$

The corresponding B-factor restraints are given by:

$$B_{NCS} = (b-b_{av})^2/\sigma_{NCS}^2 \quad (18)$$

Where w is a weight function, b and b_{av} are the respective individual and average temperature factors of NCS related atoms and σ_{NCS} is the target deviation for B-factor restraints.

2.2.4.7 Structure validation:

The refinement programs analyze the geometrical parameters with respect to bond lengths, bond angles, dihedral angles and short contacts between symmetry related atoms *etc* of the refined structure. Post-refinement the obtained structures were validated for local errors like Ramachandran outliers, flipped branched protein side chains and incorrect sugar puckers *etc*, using molprobity⁴¹.

2.2.5 Cryo-electron microscopy:

Structure determination of macromolecules using electron microscopes is recently getting popularized due to the technical advancements in the field. In this technique images of the projections of individual molecules or complexes are recorded through a transmission electron microscope (TEM). The obtained images can be combined computationally to yield a 3D structural model. This approach is also called as single particle reconstruction. The common workflow of single particle reconstruction is depicted in Fig 2.6.

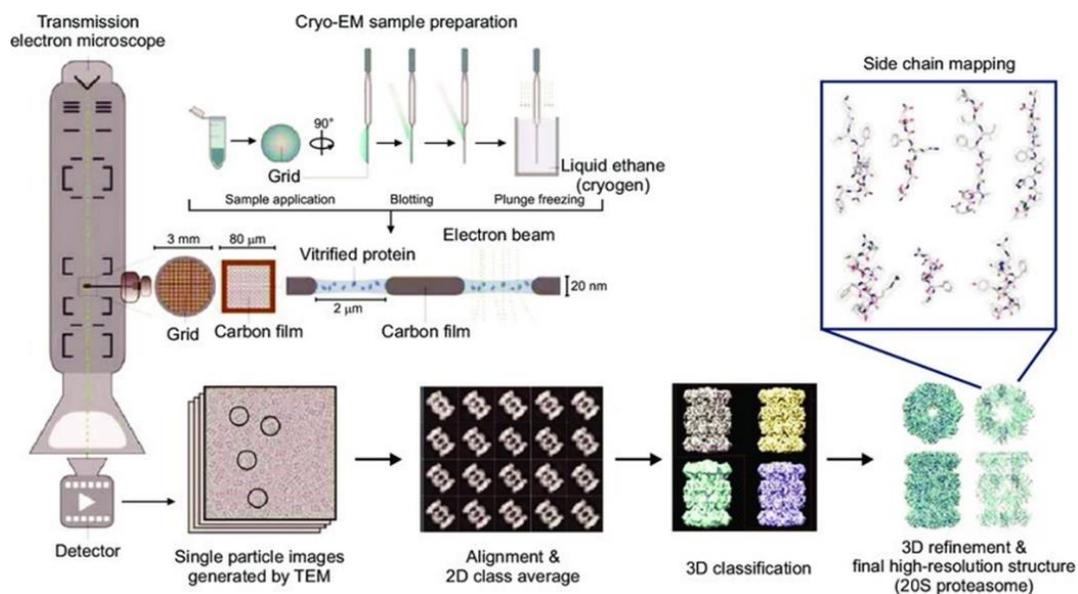


Fig 2.6: Workflow of single particle reconstruction⁴²

2.2.5.1 Cryo freezing:

Preparation of a good specimen is always a crucial step in the determination of structure by single particle Cryo-EM. The sample is flash-frozen in a process called vitrification, where the specimen containing grid is plunged into liquid ethane, to freeze the specimen fast enough, so that the specimen is preserved in a hydrated state and in its native conformation without forming ice crystals. Furthermore, flash-freezing also protects the sample from radiation damage as the samples are exposed to a high dose of electron beam⁴³.

2.2.5.2 Sample preparation:

Protein vitrification was achieved by the assistance of a FEI vitrobot. Firstly, Quantifoil 1.2/1.3, 300 mesh grids were glow discharged in the air for 90 sec at 25 mA and then placed on a pair of tweezers in the vitrobot blotting chamber. Chamber temperature was set to 18°C with 100% humidity to avoid evaporation after blotting. 3 µl of 1 µg/µl protein sample was applied on the grid for 30 sec to allow protein to absorb onto the grid. The grid was automatically blotted for 3 sec, followed by plunge-freezing in liquid ethane. Frozen grids were stored in cryo grid boxes in liquid nitrogen.

2.2.5.3 Data collection:

2.2.5.3.1 Source of electrons:

Recent days, the high-end microscopes comes with of field emission gun (FEG) with single pointed tungsten crystal coated with zirconium dioxide instead of large tungsten filament. Electrons generated with this technology are more coherent and brighter as compared to common electron source⁴³.

2.2.5.3.2 Spherical aberration and astigmatism:

In case of a conventional light microscope, the lens are considered to be perfect where overall beam on the lens of light microscope are converged to a single focal point. But, in case of electron microscope, focused beam on the lens converge closer to the optical axis in a different point, which is far away from the optical axis. This leads to several aberrations of a lens like astigmatism and spherical aberration. Astigmatism occurs when two waves propagate that perpendicular to each other has different foci. In case of spherical aberrations, as they are found in an electron microscope, are not able to focus all waves to single focal points. Ultimately leading to distortions of images, which need to be avoided. Using coma-free alignment and spherical aberration corrector one can minimize astigmatism and spherical aberration^{44,45}.

2.2.5.3.3 Detectors:

In recent times, the direct detection device (DDD) integrated with CMOS chips replaced photographic films and charge coupled devices (CCD). The signal to noise ratio is lesser compared to other two detectors. Considerably the resolution of Cryo-EM structures brought down with the help of direct detectors in combination with new data processing software gave rise to “resolution revolution” in the field of Cryo-EM⁴⁶.

2.2.5.4 Data processing:

Generally the projections collected for the particles of different orientations which are embedded in a thin layer of vitreous ice. The 2D projections of spatial orientation and position of a molecule are defined by six geometric parameters like three Euler angles, two in-plane positional parameters and the defocus which the z position along the direction of the electron

beam. The defocus for all particles in a micrograph is often assumed to be the same. After the correction of microscopic aberrations, a 3D reconstruction can be done by combining 2D projections of many molecules that are aligned to each other.

2.2.5.4.1 Particle Picking:

After the collection of datasets, the movies are aligned and averaged by beam-induced motion correction. Good micrographs are selected for particle picking. The quality of the picked particles is a limiting factor in the subsequent analysis, as inclusion of too many poor particles may affect the 3D reconstruction. Particles can be picked either by manual, semi-automated or fully automated process.

2.2.5.4.2 Contrast transfer function (CTF):

In Cryo-EM images contrast is primarily generated by setting up the imaging system to obtain interference between unscattered and scattered electrons and to a lesser extent by aberrations of the electromagnetic lenses. This type of contrast is called phase contrast because it is generated by the differences in the phase shifts of electrons when they traverse different parts of the sample. Usually, the phase contrast is generated by deliberately defocusing the objective to induce phase shifts in the scattered electrons and thus contrast in the images is described by contrast transfer function (Equation 19)⁴⁷. It gives information about the images being affected by imaging parameters and imaging lens system. Precise CTF estimation is important for evaluation of the quality of micrographs and subsequent structure determination. The parameters like acceleration voltage, defocus, spherical aberration, astigmatism and percentage of amplitude contrast are required for the CTF calculation.

$$\text{CTF} = -2 \sin \left[\pi \left(\Delta e \lambda q^2 - \frac{Cs \lambda^3 q^4}{2} \right) \right] \quad (19)$$

Where Δe is defocus, Cs is spherical aberration, q is spatial frequency and λ is electron wavelength.

2.2.5.4.3 Class averaging:

The first step in single-particle 3D reconstruction is the analysis of the 2D class averaging⁴⁸. In this process, all 2D projections are aligned and grouped into homogenous subsets. The main advantages of 2D class averaging to be done at first step include: 2D datasets

containing image artefacts, or empty fields that should be removed; if the set is dominated by a few orientations, the angular distribution of the particle views is unknown. In this case 3D analysis is unlikely to succeed. There are several strategies that have been proposed to deal with proper alignment and clustering 2D images which are embedded in K-means clustering algorithm. The alignment process is implemented through multi-reference alignment (MRA) in which the dataset represented as K seed templates. Once all images aligned, they can be compared with all templates and assigned them to the most resemble ones. It is an iterative process where the templates processed by averaging images based on the results from initial group⁴⁸. The obtained groups or class averages are 2D projections of a 3D map in different directions.

2.2.5.4.4 Angular reconstruction:

After sorting a good set of 2D projection images, one can reconstruct the 3D structure by calculating the orientational relationships between projection images. The angular reconstitution technique allows to find the relative orientations of the 2D projections of a 3D structure. The technique is based on the common line projection theorem stating that two different 2D projections of the same 3D object always have a one-dimensional (1D) line projection in common. By using the angles between such common line projections, the relative Euler-angle orientations of set projections can be determined. For an entirely asymmetric particle, at least three different projections are required to solve the orientation problem. The common line projection theorem is the real-space equivalent of the Fourier space common lines theorem. The similarities between line projections generated from 2D projection images are the central issue in the angular reconstitution approach. In the first step when a 2D projection image is given, an Euler-angle orientation of all line projections of that 2D projection image will be calculated using sinogram correlation function. The collection of all line projections of a 2D image is called a ‘sinogram’. At a position, where a pair of common lines matched, the sinogram correlation function has a maxima⁴⁴. If the molecule exhibit specific point group symmetry sinogram correlation function shows corresponding symmetry related peaks. Highly symmetric molecules were easy to process compared to asymmetric once. The common-lines approach was first used in IMAGIC as ‘angular reconstitution’. In

this method, orienting multiples of triplet class averages merged into a common framework, yielding the final 3D structure⁴³.

2.2.6 Computational methods:

2.2.6.1 Molecular dynamics simulations:

Molecular dynamics (MD) simulation technique is used to study the dynamics and molecular interactions of biological systems such as protein, nucleic acids and cell membranes at the atomic level⁴⁹. It enables the visualization of events such as protein folding, protein-drug interaction or functional dynamics associated with the biomolecules, which cannot be observed by experimental techniques. It is a technique in which the positions and velocities of atoms in the system are calculated by employing Newton's equation of motion. MD simulations use an empirical set of potential functions and parameters to describe the interactions between atoms and the energy of the system as a function of the coordinates of each atom. The accuracy of the simulations is directly related to the parameters and potential energy function used to describe the interatomic interactions. A number of force fields have been developed depending upon the energy function formulae and the strategy used for parameterization. The parameters are derived from higher level quantum mechanical calculations and experimental procedures such as NMR, X-ray diffraction, Raman and neutron spectroscopy. In this thesis, we have used charmm36 force field for protein simulations⁵⁰.

2.3 Bibliography:

1. Kibbe, W. A. OligoCalc: An online oligonucleotide properties calculator. *Nucleic Acids Res.* (2007). doi:10.1093/nar/gkm234
2. Mullis, K. B. The unusual origin of the polymerase chain reaction. *Sci. Am.* (1990). doi:10.1038/scientificamerican0490-56
3. Bitinaite, J. *et al.* USERTM friendly DNA engineering and cloning method by uracil excision. *Nucleic Acids Res.* (2007). doi:10.1093/nar/gkm041
4. Bitinaite, J. & Nichols, N. M. DNA cloning and engineering by uracil excision. *Current Protocols in Molecular Biology* (2009). doi:10.1002/0471142727.mb0321s86
5. Sambrook, J. & Russell, D. W. The inoue method for preparation and transformation of competent e. Coli: 'ultra-competent' cells. *CSH Protoc.* (2006). doi:10.1101/pdb.prot3944
6. Wilfinger, W. W., Mackey, K. & Chomczynski, P. Effect of pH and ionic strength on

- the spectrophotometric assessment of nucleic acid purity. *Biotechniques* (1997).
7. Styles, B. T. Encyclopedia of Food Sciences and Nutrition. *Encycl. Food Sci. Nutr.* (2003). doi:10.1016/B0-12-227055-X/00927-5
 8. Sahin, E. & Roberts, C. J. Size-exclusion chromatography with multi-angle light scattering for elucidating protein aggregation mechanisms. *Methods Mol. Biol.* (2012). doi:10.1007/978-1-61779-921-1_25
 9. Wyatt, P. in *Handbook Of Size Exclusion Chromatography And Related Techniques* (2010). doi:10.1201/9780203913321.ch21
 10. Huynh, K. & Partch, C. L. Thermofluor assay. *Curr Protoc Protein Sci.* (2016). doi:10.1002/0471140864.ps2809s79.Current
 11. Lo, M. C. *et al.* Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem.* (2004). doi:10.1016/j.ab.2004.04.031
 12. Motulsky, H. GraphPad Prism. *Clases* (2007). doi:10.1198/tas.2007.s86
 13. Crichton, R. R. & Louro, R. O. *Practical Approaches to Biological Inorganic Chemistry. Practical Approaches to Biological Inorganic Chemistry* (2013). doi:10.1016/C2011-0-00094-7
 14. Molekularnego, P. M., Biologiczno-chemicznych, C. N. & Warszawski, U. Oddziaływanie leków z celami molekularnymi Piśmiennictwo.
 15. Sauter, C. & Lorber, B. Crystallization – General Methods PART 4 . CRYSTALLIZATION. (2012).
 16. McPherson, A. [5]Crystallization of macromolecules: General principles. *Methods Enzymol.* (1985). doi:10.1016/0076-6879(85)14007-3
 17. Watts, A. Crystallization of nucleic acids and proteins. A practical approach. *FEBS Lett.* (1993). doi:10.1016/0014-5793(93)80565-c
 18. Brzozowski, A. M. & Walton, J. Clear strategy screens for macromolecular crystallization. *J. Appl. Crystallogr.* (2001). doi:10.1107/s0021889800017362
 19. Jancarik, J. & Kim, S. H. Sparse matrix sampling. A screening method for crystallization of proteins. *J. Appl. Crystallogr.* (1991). doi:10.1107/S0021889891004430
 20. Chayen, N. E. Comparative Studies of Protein Crystallization by Vapour-Diffusion and Microbatch Techniques. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (1998). doi:10.1107/s0907444997005374
 21. Chayen, N. E. A novel technique to control the rate of vapour diffusion, giving larger protein crystals. *J. Appl. Crystallogr.* (1997). doi:10.1107/S0021889896013532
 22. Chayen, N. E. Methods for separating nucleation and growth in protein crystallisation. *Progress in Biophysics and Molecular Biology* (2005). doi:10.1016/j.pbiomolbio.2004.07.007
 23. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta*

Crystallographica Section F: Structural Biology Communications (2014).
doi:10.1107/S2053230X13033141

24. Chayen, N. E. The role of oil in macromolecular crystallization. *Structure* (1997). doi:10.1016/S0969-2126(97)00279-7
25. Dauter, Z. in *Methods in Molecular Biology* (2017). doi:10.1007/978-1-4939-7000-1_7
26. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* (1991). doi:10.1126/science.1925561
27. Garman, E. Cool data: quantity AND quality. *Acta Crystallogr. D. Biol. Crystallogr.* (1999).
28. Garman, E. F. & Schneider, T. R. Macromolecular Cryocrystallography. *J. Appl. Crystallogr.* (1997). doi:10.1107/S0021889897002677
29. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* (80-.). (2012). doi:10.1126/science.1218231
30. Kabsch, W. *XDS*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010). doi:10.1107/S0907444909047337
31. Taylor, G. The phase problem. *Acta Crystallogr. D. Biol. Crystallogr.* (2003).
32. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
33. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. Sect. A* (1994). doi:10.1107/S0108767393007597
34. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010). doi:10.1107/s0907444909042589
35. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* (2007). doi:10.1107/s0021889807021206
36. Wang, B. C. Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* (1985). doi:10.1016/0076-6879(85)15009-3
37. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.* (2011). doi:10.1107/S0907444911001314
38. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).
39. Kantardjiev, K. A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci.* (2003). doi:10.1110/ps.0350503
40. Fokine, A. & Urzhumtsev, A. Flat bulk-solvent model: Obtaining optimal parameters. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2002). doi:10.1107/S0907444902010284

41. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography . *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010). doi:10.1107/s0907444909042073
42. Chung, J.-H. & Kim, H. M. The Nobel Prize in Chemistry 2017: High-Resolution Cryo-Electron Microscopy. *Appl. Microsc.* **47**, 218–222 (2018).
43. Thompson, R. F., Walker, M., Siebert, C. A., Muench, S. P. & Ranson, N. A. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* (2016). doi:10.1016/j.ymeth.2016.02.017
44. Zemlin, F., Weiss, K., Schiske, P., Kunath, W. & Herrmann, K. H. Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms. *Ultramicroscopy* (1978). doi:10.1016/S0304-3991(78)80006-0
45. Haider, M. *et al.* Electron microscopy image enhanced [7]. *Nature* (1998). doi:10.1038/33823
46. Kuhlbrandt, W. The resolution revolution. *Science* (2014). doi:10.1126/science.1251652
47. Zanetti, G., Riches, J. D., Fuller, S. D. & Briggs, J. A. G. Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging. *J. Struct. Biol.* (2009). doi:10.1016/j.jsb.2009.08.002
48. Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. S_A Primer to Single-Particle Cryo-Electron Microscopy. *Cell* (2015). doi:10.1016/j.cell.2015.03.050
49. Perilla, J. R. *et al.* Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.* (2015). doi:10.1016/j.sbi.2015.03.007
50. Huang, J. & Mackerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).

Chapter 3

Structural basis of substrate specificity of Iridoid Synthases

3.1 Background:

Iridoid synthase (IS), a key enzyme involved in the synthesis of Iridoids, belongs to short chain dehydrogenase/reductase (SDR) super family¹. Iridoids are major precursors for widely used anti-cancer drugs like vinblastine and vincristine². Unlike the canonical terpene synthases, IS catalyzes the NADPH-dependent reduction of 10-Oxogeraniol resulting in the formation of an *enol* or *enolate* intermediate, which further cyclizes *via* Michael addition to yield napetalactol³. Recent studies, by us and others, have revealed that the structure and catalytic mechanism of this enzyme are similar to the P5 β R^{4,5,6,7}. However, these two enzymes exhibit markedly different substrate specificities (Fig 3.1). Earlier studies have suggested that two loops L1 (residues from 156 to 167) and L2 (residues from 350 to 361) at the opening of the substrate binding pocket of IS are involved in ligand specificity⁵, however the determinants of substrate specificity and their mechanism is still evasive. The current chapter unravels the structural basis of substrate specificity of iridoid synthases by employing structural and sequence analysis, site-directed mutagenesis and molecular dynamics simulations.

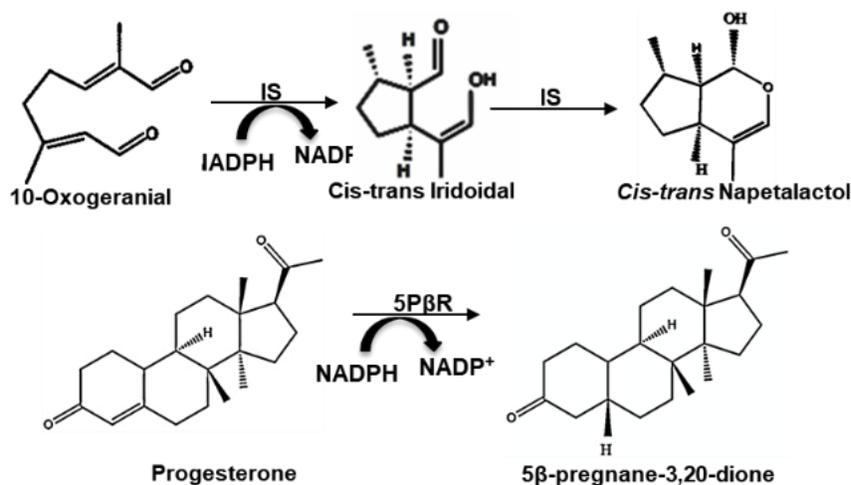


Fig 3.1 Reactions catalyzed by iridoid synthase and progesterone 5 β -reductase.

3.2 Methods:

3.2.1 Cloning, protein expression and purification:

To obtain high yields of homogenous recombinant IS, the gene (37-1167 bps) was cloned in pGEX-6P-1 expression vector, which contains a cleavable N-terminal GST tag followed by PreScission protease cleavage site. Optimum expression was achieved in *E.coli* strain B834 (rare). The cells were grown at 37°C to an OD₆₀₀ of 0.5 and further induced with 0.5 mM IPTG and incubated for 14 hrs at 18°C. The culture was harvested by centrifugation. The cells were lysed in chilled lysis buffer containing bufferA (50 mM Tris pH-8.0, 500 mM NaCl, 2 mM EDTA, 2 mM β-mercaptoethanol, and 5% glycerol), one tablet of cOmplete EDTA free protease inhibitor (Roche) and a pinch of lysozyme, employing sonication. The lysate was filtered and loaded on to the GST affinity column (GE Healthcare Biosciences, USA) for binding. The column was washed with buffer A and the protein was eluted with 20 mM glutathione reduced (Sigma) prepared in the same buffer. The elution fractions were dialyzed along with the PreScission protease (purified in our lab) overnight at 10°C. Glutathione, GST tag and PreScission protease were removed by the desalting followed by second GST affinity column. To improve purity of the protein, size exclusion chromatography was performed using Hiprep26/60 sephacryl S-300 (GE Healthcare Biosciences, USA) column in 20 mM Tris pH-8.0, 150 mM NaCl, 2 mM DTT. The peak fractions under specific molecular weight were collected and concentrated to perform crystallization experiments. Commercially synthesized genes encoding IS7M and P5βR were obtained from Life Technologies. The genes were cloned in pGEX-6P-1 expression vector. Expression and purification of P5βR and mutant-IS7M was done as per the method described for IS.

3.2.2 Site-directed mutagenesis

Selective mutants of IS were constructed by PCR using the QuikChange lightning site-directed mutagenesis kit (Agilent) with the appropriate primer pairs. All primers were purchased in a reverse phase-HPLC-purified quality from Eurofins Bangalore. The pGEX-6P expression plasmid containing wild-type IS was used as a template. PCR was performed in a program as follows: 30 sec at 95°C (denaturation), 18 cycles of 10 sec at 98°C (denaturation), 30 sec at 55°C (annealing), and 5 min at 68°C (extension). Mutated genes were sequenced, and the proteins were produced in *E. coli* (strain BL 21*) as described for the wild-type protein.

3.2.3 GC-MS assay:

GC-MS based assay has been performed to analyze enzyme activity. The reaction mixture contained 100-200 μg of enzyme, 200 μM 10-oxogeranial/progesterone and 200 μM NADPH in 500 μL of buffer containing 20 mM MOPS pH7.0, 10% glycerol. Reaction mixtures were incubated at 25-30°C and was terminated after 1 hrs by adding 1 ml of TBME. Compounds were extracted in TBME, 1-5 μL of extracted compounds were injected into GC-MS for analysis using following method for progesterone: equilibration time for 2 min followed by 70°C for 0 min then 10°C/min to 230°C for 0 min then 2°C/min to 250°C for 0 min then 10°C/min to 280°C for 6 min and the overall run time is 39.5 min. For 10-oxogeranial: equilibration time for 0.5 min, oven program is 60°C for 0 min then 20 °C/min to 120°C for 0 min then 2.5 °C/min to 160°C for 0 min then 20°C/min to 190°C for 2.5 min. The separation was performed on Rtx-5ms column (30m x 0.25mm x 0.25 μm) using helium as carrier gas at 1 ml min⁻¹ and with an injector temperature of 220°C.

3.2.4 Spectrophotometric assay:

The enzyme specific activity studies of wild type and mutant proteins were performed in presence of cofactor NADPH and for substrates, 10-oxogeranial and progesterone. For the determination of Michaelis-Menten parameters with 10-oxogeranial and progesterone, kinetics of NADPH consumption were determined spectrophotometrically at 340 nm. Reactions contained 200 μM NADPH, 10 μM to 100 μM of 10-oxogeranial/progesterone, 2.5 μM (IS/P5 β R) enzyme in a total volume of 1 ml buffer containing 20 mM MOPS pH 7.0, 10% glycerol. The reaction was equilibrated at room temperature (22°C) and started by addition of enzyme. To obtain K_{cat} and K_m value of the reaction plots of initial velocities versus substrate concentration were non-linearly fit to the Michaelis-Menten equation using GraphPad Prism⁸. Data were collected for 5 min at 1 sec intervals. All the assays were performed in triplicates.

3.2.5 MD simulations:

The crystal structure of IS was obtained from PDB (PDB ID code 5EMH)⁴. Since the loop region (residues 158 to 166) was not resolved in this structure, it was modeled using a template loop structure from another structure, PDB ID code 5COB⁶. Similarly the coordinates of the ligand (10-oxogeranial) and cofactor (NADP⁺) were obtained. Mutations in

the loop region E161-G162 to S161-Y162, corresponding to IS^{SY} were made using PyMol⁹. The force-field parameters for 10-oxogeranial and NADP⁺ were obtained using ParamChem web server¹⁰. Four systems were considered: the apo and 10-oxogeranial bound wild-type protein (IS) and the apo and 10-oxogeranial bound mutant protein (IS^{SY}), bound to NADP⁺ and solvated in TIP3P water¹¹. In order to ensure reproducibility of the observed loop dynamics, four independent simulation sets for the wild type and mutants (i.e. total 8 simulations) were performed. All atom simulations were performed using Charmm36 force-field and GROMACS 4.5.5 package^{12,13}. The systems were minimized using steepest descent algorithm. NVT equilibration was carried out for 100 ps at 300 K as 0.1 ps coupling constant for protein, ligand and cofactor, water and ions as groups using v-rescale thermostat¹⁴. Further, NPT equilibration was carried out for 100 ps using Parrinello-Rahman barostat¹⁵. Pressure was maintained isotropically at 1.0 bar. The electrostatic interactions were calculated using Particle-Mesh Ewald (PME)¹⁰. A time step of 2 fs was used. Finally, duplicate simulations were carried out for 300 ns for each of the four systems, corresponding to a total simulation time of 2.4 μ s. Simulations were visualized and snapshots were rendered using VMD¹⁶. Analysis was carried out with in-house and Gromacs utility tools.

To characterize the structural differences between IS and IS^{SY}, we computed a pairwise difference plot, $\Delta P = Pd (IS) - Pd (IS^{SY})$, where Pd is the probability of residue pairs to exist within a cut-off distance, d. A value of +1 for ΔP implies that the residue pair is closer in IS, whereas a value of -1 indicates that the residue pair is closer in IS^{SY}. ΔP was calculated for apo and ligand bound systems separately and averaged over simulation sets. A cut-off of 1.0 nm was used in the calculations.

3.2.6 Homology modelling:

The high throughput homology modelling was done using modeler 9.19¹⁷. Gene sequences of IS homologs were obtained from GenBank (AmISY-MF281392, SmIS1-MF044036, SmIS2-MF044037, OeISY-KT954038, Oe1,4-R1-KT954039, Oe1,4-R2-KT954043, Oe1,4-R3-KT954044, NmISY1-KY882235, NmISY2-KY882236, NcISY1-KY882233, NcISY2-KY882234). P5 β R (2V6G) and IS (5COA, 5COB) structures were used as the model templates, the coordinates were downloaded from PDB. A set of python script files executed for the sequence alignment and modelling. For each homolog five models were

generated. The best model out of five was selected based on the discrete optimized protein energy (DOEP) score which is generated from the evaluation.

3.3 Results:

3.3.1 Cloning, expression and purification of IS:

IS gene was cloned in pGEX-6P-1 bacterial expression vector (Fig 3.2A). It contains N-terminal GST-tag followed by PreScission protease cleavage site. The protein was expressed in B834 rare strain. The protein was soluble with GST-tag, as it appeared in supernatant. The post PreScission treatment followed by gel filtration, protein was eluted in expected fractions (Fig 3.2B). Purified protein was used for crystallization and other biochemical assays.

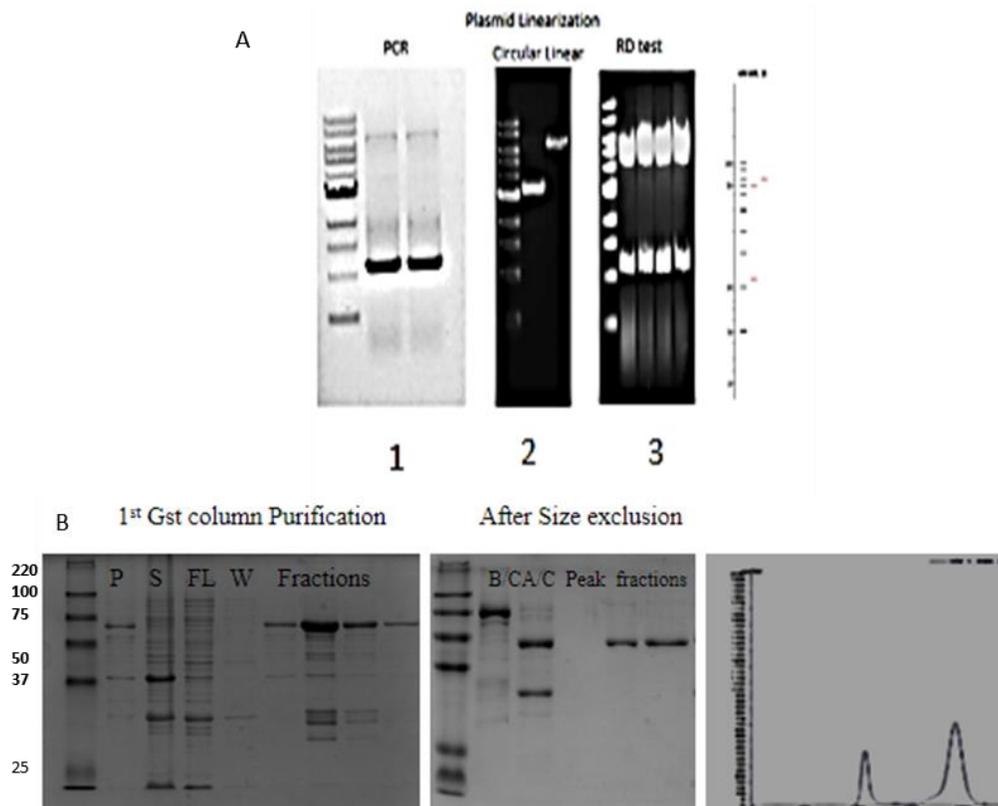


Fig 3.2 A) Agarose gel images of IS PCR (1), plasmid linearization (2) and restriction digestion test (3) B) SDS-PAGE images and gel filtration profile of IS purification.

3.3.2 Crystallization and structure determination of IS:

Crystallization was carried out with concentrated protein (10 mg/ml) along with co-factor and 10-oxogeraniol, each in 5 molar excess. 1 μ l of protein and 1 μ l of mother liquor (0.1

M tri sodium citrate pH 5.6, 15% MPD and 20% PEG4000) were mixed and allowed to incubate at 22°C in the vapor diffusion hanging drop method. The diffractable crystals were obtained within a day (Fig 3.3). The crystals were flash frozen in liquid nitrogen using 28% Glycerol in mother liquor as a cryoprotectant. The data sets were collected in the INDUS-2 synchrotron beamline BI-21, RRCAT, Indore. Data was integrated and scaled with XDS and SCALA respectively. The structure solution was obtained by molecular replacement with phaser¹⁸ using P5 β R as a search model. Iterative rounds of model building refinement were carried using COOT¹⁹ and REFMAC5²⁰ till the R and R_{free} of 0.182 and 0.215. Diffraction data and refinement statistics are summarized in Table 3.1.

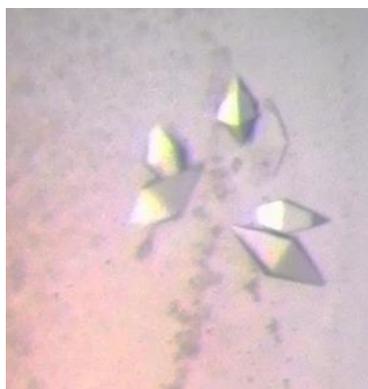


Fig3.3: Crystals of IS.

Table 3.1 Diffraction Data collection and refinement statistics

Data collection statistics	
Space Group	P4 ₃ 2 ₁ 2
a (Å)	66.86
b (Å)	66.86
c (Å)	173.1
α(°)	90
β(°)	90
γ(°)	90
Resolution limits (Å)	47.28-2.10
R_{merge}	0.119 (0.798)
I/s (I)	21.1 (4.0)
Number of reflections	281156 (22930)
Unique reflections	23849 (1910)
Completeness (%)	100 (100)
Multiplicity	11.8 (12.0)

Refinement statistics	
Resolution limits (Å)	62.36-2.10
Number of reflections	281156
Working set	23772
Test set	1219 (131)
R_{work}/R_{free}	0.1823/0.2156
Number of atoms	
Protein	2806
Ligand	48
Water	201
Mean B-factors	
Protein atoms (Å²)	30.20
Ligand	29.59
Water	38.87
RMSD from Ideal values	
Bond length (Å)	0.0174
Bond angles (°)	1.858
Ramachandran plot Statistics	
Preferred (%)	96.28
Allowed (%)	3.15

3.3.3 Structure of IS:

The overall structure of IS similar to its homolog P5βR (RMSD ~0.96 Å). As like other SDR family proteins, IS contains typical dinucleotide-binding “Rossmann” fold covered by a helical C-terminal extension. In the structure we have observed well defined electron density for NADP⁺ present in the central cleft of the protein (Fig 3.4A). It also harbors substrate and consist of conserved catalytic residues. The SEC results suggest that IS mainly exists as homodimers in solution, but in the structure we observed one molecule in the asymmetric unit. Usually SDR superfamily proteins contains conserved S_{X12}Y_{xxx}K catalytic motif. Surprisingly, IS contains conserved Tyr185 in the catalytic pocket. From the structure we found that the -OH group of Tyr185 essentially interacts with O-2' of nicotinamide ribose. Additionally in some of substrate bound structures, it has been shown that Tyr185 interacts with O1 of 8-oxogeraniol. It suggested that Tyr185 is crucial for the catalysis for stabilizing cofactor and substrate (Fig 3.4B). When substrate binds in active site, it immediately interacts with -OH group of Tyr185. The electron transfer to the substrate initiated by adjacent cofactor gives rise to *enol/enolate* intermediate in which a series of electron relay takes place and led to C4-C5 rotation. This enables cyclization of 10-oxogeraniol to nepetalactol (Fig 3.4C). In

case of P5 β R, Lys147 was suggested to play a role in reprotonating the Tyr²¹. However, IS indeed contain Lys at the identical position but the e-NH₃⁺ of Lys153 (equivalent of Lys147 of P5 β R) is too far from the Tyr (Fig 3.4C). Despite of having high structural similarity with P5 β R, IS shows different substrate specificity.

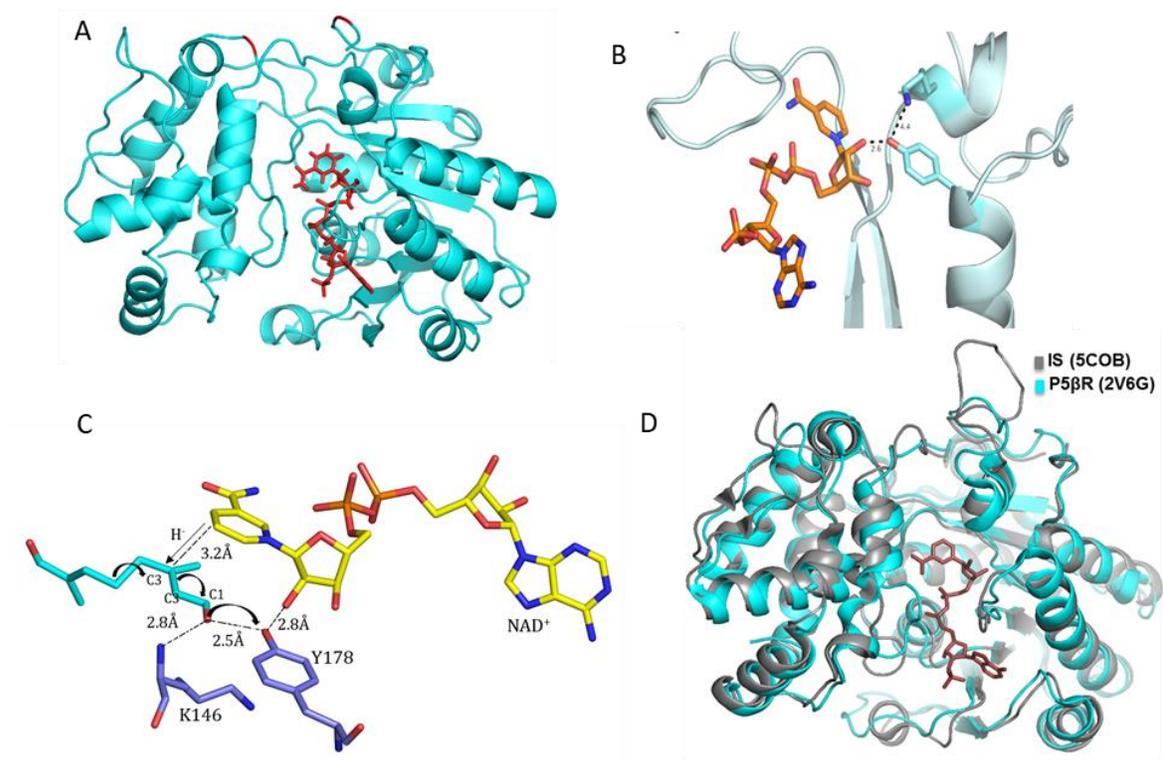


Fig 3.4: A) Structure of IS and NADP complex. IS colored in cyan and NADP colored in red. B) Active site of IS, Tyr185 and NADPH (orange) showed in sticks. C) Interaction of active-site residues Tyr178 and Lys146 with NADP and 8-oxogeraniol⁷. D) Superposition of IS (grey) (PDB ID code 5COB) and P5 β R (cyan) (PDB ID code 2V6G) (RMSD is 0.96), NADP+ (brown) shown in sticks.

3.3.4 Identification of sequence-signatures governing substrate specificity:

Selective site-directed mutagenesis of divergent residues lying within 5Å radius of the ligand has shown to modestly alter the activity of the enzymes towards their cognate substrate. However, significance of these mutants on the specificity of the enzymes is not clear. As a consequence, these studies suggest that it could be a sequence or structure motif, not individual residues that determine the substrate specificity. Therefore, to pin down such a motif we searched all annotated and unannotated protein sequences from Swiss-Prot and TrEMBL databases using the Ex-PASY ScanProsite tool (Fig 3.5)²². Initially, we searched with the query sequence motif “PWKVYGVARR”, found to be present in all annotated SDR family of

proteins. The sequence alignment obtained from this search was further optimized by introducing appropriate gaps and other conserved motifs, such as KHY. The result of this search gave around 113 hits that spanned all proteins from known SDR family, including IS and P5 β R. However, the alignments obtained from the search motif “X(40)PWKVYGVARRX(32,52)X(33)KHXYX(5)SYX(0,8)X(10)PRX(0,20)X95WKHX(100)” had hits with very high similarity to P5 β R. In fact almost all the annotated proteins in this cluster were progesterone reductases (Table 3.2, Fig 3.6.). Further, when “SY” motif of the search string was replaced with “EG”, which corresponds to the residues in IS, the search hits were found to be highly similar to IS. Like the hits with “SY” search, most of the proteins of the cluster obtained with “EG” search were IS or IS-like proteins (Table 3.3, Fig 3.7). Interestingly, “SY” or “EG” residues of the search motif belong to the L1 loop of the P5 β R and IS structures, which lies at the entrance of the substrate binding pocket (Fig 3.9A). Thus the sequence alignment study suggests that SY and EG might have important roles in function of these enzymes and thus could play significant role in conferring substrate specificity of P5 β R and IS.

[Home](#) | [ScanProsite](#) | [ProRule](#) | [Documents](#) | [Downloads](#) | [Links](#) | [Funding](#)



ScanProsite tool

This form allows you to scan proteins for matches against the [PROSITE collection of motifs](#) as well as against your own patterns.

- Option 1 - Submit PROTEIN sequences to scan them against the PROSITE collection of motifs.
- **Option 2 - Submit MOTIFS to scan them against a PROTEIN sequence database.**
- Option 3 - Submit PROTEIN sequences and MOTIFS to scan them against each other.

STEP 1 - Enter a MOTIF or a combination of MOTIFS [Examples](#) [\[help\]](#)

Enter a PROSITE accession or Identifier or your own pattern or a combination

Supported input:

- A PROSITE accession e.g. [PS50240](#) or identifier e.g. [TRYPSIN_DOM](#)
- Your own pattern e.g. [P-x\(2\)-G-E-S-G\(2\)-\[AS\]](#)

» More

» **Options** [\[help\]](#)

Fig 3.5: ScanProsite web server.

Table 3.2: Example of hits from ScanProsite with search string

{X(40)PWKVYGVARRX(32,52)X(33)KHXY(5)SYX(0,8)X(10)PRX(0,20)X(95)WKHX(100)}

Entry	Entry name	Protein names	Organism	Length
Q6PQJ9	5BPOR_DIGLA	3-oxo-Delta(4,5)-steroid 5-beta-reductase	Digitalis lanata (Grecian foxglove)	389
G9HWD8	G9HWD8_9LAMI I	Putative progesterone 5-beta-reductase	Digitalis cariensis	389
G9HWD9	G9HWD9_9LAMI I	Putative progesterone 5-beta-reductase	Digitalis lamarckii	389
Q00P19	Q00P19_9LAMI	Progesterone 5-beta-reductase	Digitalis cariensis	389
Q5XTQ6	Q5XTQ6_9LAMI	Progesterone 5-beta-reductase	Digitalis thapsi	389
Q3HLN0	Q3HLN0_9LAMI	Progesterone 5-beta-reductase	Digitalis viridiflora	389
Q3HLN1	Q3HLN1_9LAMI	Progesterone 5-beta-reductase	Digitalis laevigata	389
Q6PQK1	Q6PQK1_DIGGR	Progesterone 5-beta-reductase	Digitalis grandiflora (Yellow foxglove)	389
A0A075M2G6	A0A075M2G6_9LAMI	Progesterone 5-beta-reductase	Digitalis ferruginea subsp. ferruginea	389
Q2XSK7	Q2XSK7_9LAMI	Progesterone 5-beta-reductase	Digitalis nervosa	389
Q84RS3	Q84RS3_DIGPU	Progesterone 5-beta-reductase	Digitalis purpurea (Common foxglove)	389
Q6PQJ8	Q6PQJ8_DIGPU	Putative progesterone 5-beta-reductase	Digitalis purpurea (Common foxglove)	389
G9HWE0	G9HWE0_9LAMI	Putative progesterone 5-beta-reductase	Digitalis ferruginea subsp. schischkinii	389
Q078S3	Q078S3_9LAMI	Progesterone 5beta-reductase	Isoplexis sceptrum	389
Q3HLM8	Q3HLM8_9LAMI	Progesterone 5-beta-reductase	Digitalis davisiana	389
Q1HDK8	Q1HDK8_9LAMI	Progesterone 5-beta-reductase	Digitalis minor var. minor	389
G9HWE1	G9HWE1_9LAMI	Putative progesterone 5-beta-reductas	Digitalis trojana	389
A0A022RBG3	A0A022RBG3_E RYGU	Uncharacterized protein	Erythranthe guttata (Yellow monkey flower) (Mimulus guttatus)	389
Q3HLM7	Q3HLM7_9LAMI	Progesterone 5-beta-reductase	Digitalis lutea	389
A0A0U3SD11	A0A0U3SD11_O LEEU	1,4-reductase 3	Olea europaea subsp. europaea	389
Q5XLZ9	Q5XLZ9_9LAMI	Progesterone 5 beta reductase	Digitalis subalpina	389
Q078S5	Q078S5_9LAMI	Progesterone 5beta-reductase	Isoplexis chalcantha	389

Q6PQK0	Q6PQK0_9LAMI	Progesterone 5-beta-reductase	Digitalis parviflora	389
Q5XTQ7	Q5XTQ7_9LAMI	Progesterone 5-beta-reductase	Digitalis ferruginea	389
Q84UI3	Q84UI3_9LAMI	Putative progesterone 5-beta-reductase	Digitalis obscura	389
Q3HLM6	Q3HLM6_9LAMI	Progesterone 5-beta-reductase	Digitalis sibirica	389
Q5XTQ8	Q5XTQ8_DIGPU	Progesterone 5-beta-reductase	Digitalis purpurea subsp. heywoodii	389

```

tr Q5XTQ7 Q5XTQ7 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q5XLZ9 Q5XLZ9 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q3HLN0 Q3HLN0 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q00P19 Q00P19 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q1HDK8 Q1HDK8 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q2XSK7 Q2XSK7 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr G9HWE1 G9HWE1 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr A0A075M2G6|A0A075M2G6 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q84RS3 Q84RS3 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
sp Q6PQJ9 5BPOR_DIGLA pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q3HLM9 Q3HLM9 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q6PQK0 Q6PQK0 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppysedlPRlkymnfyydlediml
tr G9HWE0 G9HWE0 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q3HLM7 Q3HLM7 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q078S5 Q078S5 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr G9HWD9 G9HWD9 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q3HLM6 Q3HLM6 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q078S3 Q078S3 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr A0A022RBG3|A0A022RBG3 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppysedlPRlkymnfyydlediml
tr Q3HLN1 Q3HLN1 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q6PQK1 Q6PQK1 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr G9HWD8 G9HWD8 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q078S4 Q078S4 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q6PQJ8 Q6PQJ8 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q1HN30 Q1HN30 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q3HLM8 Q3HLM8 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q5XTQ8 Q5XTQ8 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q5XTQ6 Q5XTQ6 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr Q84UI3 Q84UI3 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppytedlPRlkymnfyydlediml
tr A0A0U3SD11|A0A0U3SD11 pncpnlkhislqtgrKHYmgpfeSYgkie...shdppysedlPRlkymnfyydlediml

```

Fig 3.6: Sequence alignment of P5βR hits from ScanProsite, conserved motifs showed in orange box.

Table 3.3: Example of hits from ScanProsite with search string:

{X(40)PWKVYGVARRX(32,52)X(33)KHYYX(5)EGX(0,8)X(10)PRX(0,20)X(95)WKHX(100)}

Entry	Entry name	Protein names	Organism	Length
K7WDL7	IRIS_CATRO	Iridoid synthase (EC 1.3.1.99)	Catharanthus roseus	388
A0A1B2LUC8	A0A1B2LUC8_CATRO	Mutant iridoid synthase	Catharanthus roseus	386
A0A1B2LUC3	A0A1B2LUC3_CATRO	Mutant iridoid synthase	Catharanthus roseus	386
A0A1B2LUD1	A0A1B2LUD1_CATRO	Mutant iridoid synthase	Catharanthus roseus	386
A0A1B2LUD0	A0A1B2LUD0_CATRO	Mutant iridoid synthase	Catharanthus roseus	386
A0A0A0S6P1	A0A0A0S6P1_CATRO	Progesterone 5-beta-reductase 5	Catharanthus roseus	388
A0A1B2LUC4	A0A1B2LUC4_CATRO	Mutant iridoid synthase	Catharanthus roseus	386
A0A1B2LUC9	A0A1B2LUC9_CATRO	Mutant iridoid synthase	Catharanthus roseus	386
A0A067YE54	A0A067YE54_CATRO	Iridoid synthase	Catharanthus roseus	386
A0A1B2LUE1	A0A1B2LUE1_CATRO	Mutant iridoid synthase	Catharanthus roseus	386
A0A1B2LUC7	A0A1B2LUC7_CATRO	Mutant iridoid synthase	Catharanthus roseus	386
A0A1B2LUC5	A0A1B2LUC5_CATRO	Mutant iridoid synthase	Catharanthus roseus	386
A0A126KD71	A0A126KD71_LONJA	Iridoid synthase	Catharanthus roseus	388

```

tr A0A1B2LUC8|A0A1B2LUC8|pnasnlqhvclqtgiKHYfgifeEGskvv...phdspftedlPRlnvpnfyhledily
sp K7WDL7|IRIS_CATRO|pnasnlqhvclqtgiKHYfgifeEGskvv...phdspftedlPRlnvpnfyhledily
tr A0A1B2LUD0|A0A1B2LUD0|pnasnlqhvclqtgiKHYfgifeEGskvv...phdspftedlPRlnvpnfyhledily
tr A0A1B2LUC9|A0A1B2LUC9|pnasnlqhvclqtgiKHYfgifeEGskvv...phdspftedlPRlnvpnfyhledily
tr A0A126KD71|A0A126KD71|pnasnlqhvclqtgiKHYfgifeEGskvv...phdspftedlPRlnvpnfyhledily
tr A0A1B2LUC4|A0A1B2LUC4|pnasnlqhvclqtgiKHYfgifeEGskvv...phdspftedlPRlnvpnfyhledily
tr A0A1B2LUC5|A0A1B2LUC5|pnasnlqhvclqtgiKHYfgifeEGskvv...phdspftedlPRlnvpnfyhledily
tr A0A1B2LUC3|A0A1B2LUC3|pnasnlqhvclqtgiKHYfgifeEGskvv...phdspftedlPRlnvpnfyhledily
tr A0A1B2LUC7|A0A1B2LUC7|pnasnlqhvclqtgiKHYfgifeEGskvv...phdspftedlPRlnvpnfyhledily
tr A0A067YE54|A0A067YE54|pnasnlqhvclqtgiKHYfgifeEGskvv...phdspftedlPRlnvpnfyhledily
tr A0A1B2LUE1|A0A1B2LUE1|pnasnlqhvclqtgiKHYfgifeEGskvv...phdspftedlPRlnvpnfyhledily
tr A0A0A0S6P1|A0A0A0S6P1|pnasnlqhvclqtgiKHYfgifeEGskvv...phdspftedlPRlnvpnfyhledily
tr A0A1B2LUD1|A0A1B2LUD1|pnasnlqhvclqtgiKHYfgifeEGskvv...phdspftedlPRlnvpnfyhledily

```

Fig 3.7: Sequence alignment of IS hits from ScanProsite conserved motif showed in box.

3.3.5 Site-directed mutagenesis studies to validate “SY”/“EG” as substrate specificity defining residues:

Initially, to explain the structural determinant of the substrate specificity, we identified 7 divergent residues between IS and P5 β R, that lie within a radius of 5Å from the substrate (Fig 3.8A, Fig 3.9A). In order to switch the substrate specificity of IS from 10-oxogeranial to progesterone, we mutated these divergent residues with the corresponding residues from P5 β R. However, none of these mutants showed activity towards progesterone, when detected through GC-MS analysis. To test the synergistic role of these residues in generating substrate specificity, a mutant-IS7M, carrying all these mutations together, was generated. The catalytic efficacies (K_{cat}/K_m) (Table 3.4 and Table 3.5) of the mutant (IS^{7M}) was measured for the progesterone and 10-oxogeranial (Fig 3.8B). Surprisingly, even this mutant did not exhibit activity towards progesterone, suggesting that the specificity defining factors might involve complex interplay of structure and sequence. These observations were further confirmed with GC-MS analysis (Fig 3.10). Subsequently, our sequence alignment studies showed that sequence motif “SY”/ “EG”, from the L1 loop of the P5 β R/IS, could be crucial in conferring the substrate specificity. To further verify this we mutated EG (residue number 161 and 162) in IS with SY and vice-versa in P5 β R. Interestingly, both the mutants, IS^{SY} and P5 β R^{EG}, gained activity towards their non-cognate substrates, without impairing activity towards their cognate substrates (Fig 3.9B, Table 3.4 and Table 3.5). GC-MS analysis of the products further confirmed that IS^{SY} and P5 β R^{EG} have indeed acquired activity to reduce progesterone and 10-oxogeranial, respectively (Fig 3.10). Thus the study clearly indicates that “SY”/ “EG” motif plays a critical role in substrate selectivity.

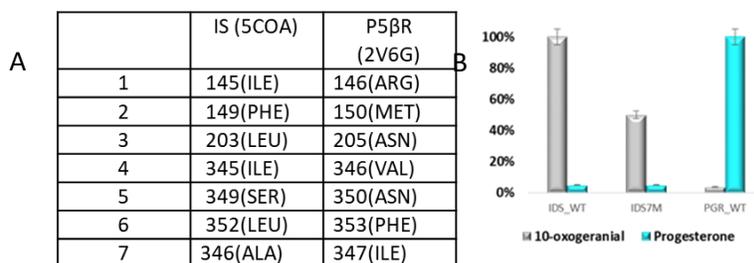


Fig 3.8 A) Divergent residues between IS and P5 β R, lying within a radius of 5Å from the substrate B) Relative activities of IS, IS7M and P5 β R towards 10-oxogeranial and Progesterone.

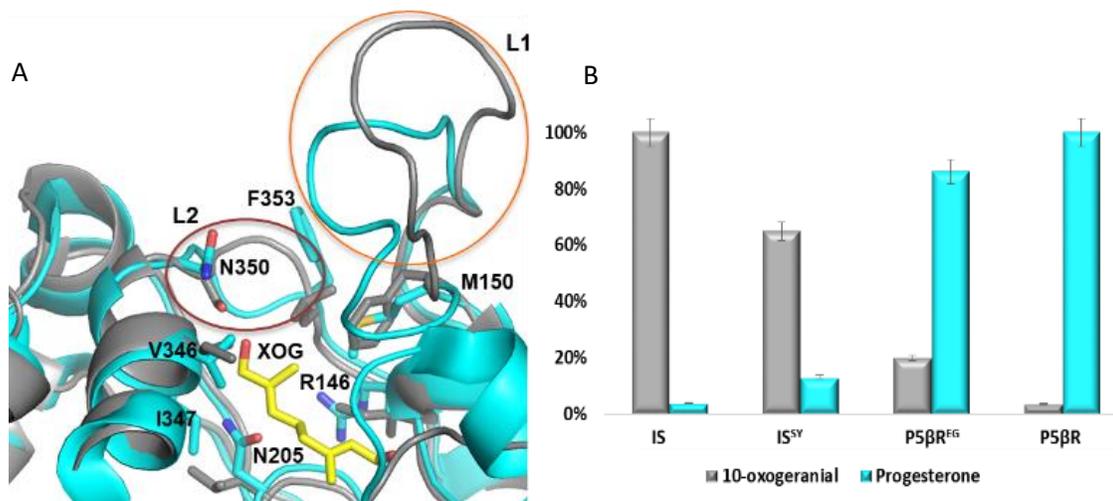


Fig 3.9 A) Superimposition of IS (PDB ID code 5COB) (grey) and P5βR (PDB ID code 2V6G) (cyan) structures, showing the difference in conformation of L1. The divergent residues at the binding pocket, lying within 5 Å radius from 10-oxogeranial (XOG), are shown in sticks. B) Relative activity of wild-type IS & P5βR, IS^{SY} and P5βR^{EG} with respect to 10-oxogeranial and progesterone.

Table 3.4: Kinetic parameters of NADPH consumption with 10-oxogeranial

Protein	Substrate K_m (μM) (10-oxogeranial)	Substrate V_{max}	Substrate K_{cat}	K_{cat}/K_m
IS	5.04± 1.7	0.021	10.87±0.67	2.15
IS ^{SY}	6.27±2.3	0.017	8.59±0.58	1.37
IS ^{7M}	7.7±1.5	0.016	8.2±0.39	1.06
P5βR ^{EG}	214±98	0.174	87.4±30.2	0.4
P5βR	438±307	0.078	39±9.6	0.08

Table 3.5: Kinetic parameters of NADPH consumption with Progesterone

Protein	Substrate K_m (μM) (Progesterone)	Substrate V_{max}	Substrate K_{cat}	K_{cat}/K_m
IS	342±294	0.04	21.16±8.9	0.06
IS ^{SY}	243±119	0.086	43.37±11.2	0.176
IS ^{7M}	296±118	0.042	21.01±8.6	0.07
P5βR ^{EG}	102±35.5	0.177	88.7±21.5	0.86
P5βR	85±12.2	0.170	85.16±21.08	1.002

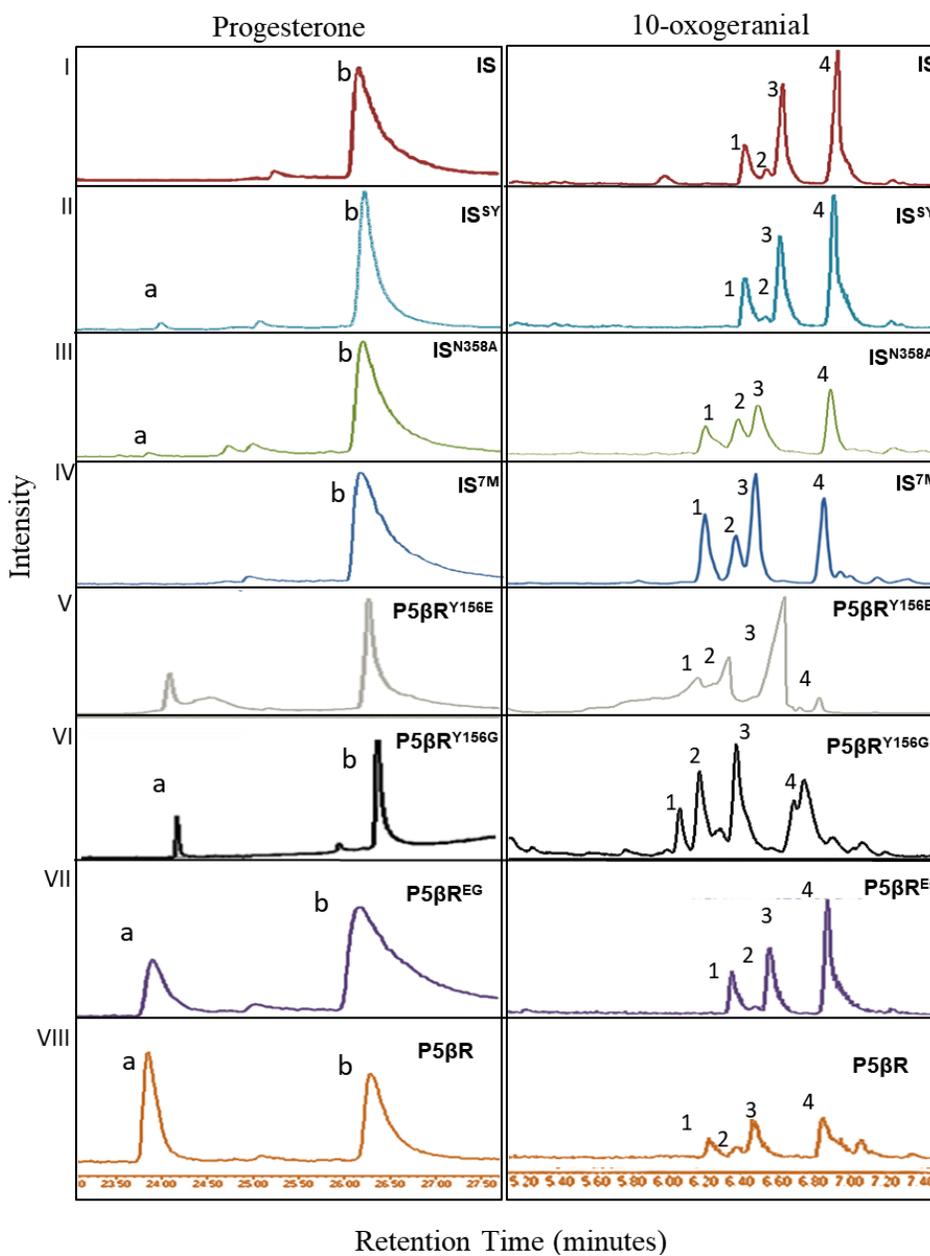


Fig 3.10: Product profiles of GC-MS analysis with 10-oxogeraniol and progesterone. Reactions were performed for IS (red), IS^{SY} (cyan), IS^{N358A} (green), IS^{7M} (black), P5 β R^{EG} (blue) and P5 β R (orange) with respect to progesterone and 10-oxogeraniol. In the chromatograms, peak a: 5 β -pregnane-3,20-dione, b: Progesterone, 1,2,3: Isomeric mix of iridoidals, 4: Napetalactol.

3.3.6 Structural basis of substrate selectivity of SDR family of reductases:

Our observations clearly indicate that the loops L1 and L2 at the orifice of substrate binding pocket define the substrate specificity of the enzymes. Superposition of the available structures of IS show L1, harboring “SY”/“EG”, to be highly flexible with different conformations (Fig 3.11). In fact in 5 out of the 10 available crystal structures the electron density for this loop was not observed. Further, previous studies have attributed these conformational differences of the loop to the possible crystal contacts. Hence it was inferred that this particular loop might not have any role in conferring the substrate specificity. On the contrary, Kries *et al.* noticed the conformational difference of L1 and L2, with respect to the corresponding loops in P5 β R and surmised them to have role in substrate selectivity⁵. In IS, when these loops were successively swapped with the corresponding loops in P5 β R the mutants (ISL1 and ISL2) the enzyme gained activity towards progesterone. But, the activity of ISL2 towards 10-oxogeranial diminished by about 2000 folds. However, our sequence and mutational analysis of IS and P5 β R has narrowed down the residues from only L1, Glu161 and Gly162 of IS and corresponding Ser155, Tyr156 of P5 β R, as the ones involved in conferring the substrate selectivity of these enzymes.

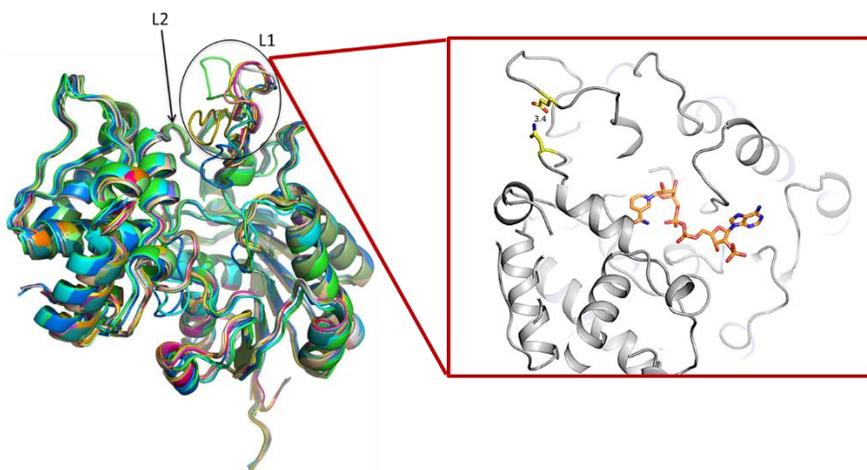


Fig 3.11: Difference in L1 conformation among different crystal structures of IS 5coa_A(Green),5coa_B(cyan),5cob_A(magenta),5cob_B(yellow),5cob_C(wheat),5cob_D(bluewhite),5dbf_A(blue),5dbf_B(orange),5dbg_A(lightgreen),5dbg_B(forest),5dbi_A(pink),5dbi_B(yelloworange),5dcu_A(purple),5dcu_B(grey),5dcw_A(marine),5dcy_A(olive),5dcy_B(smudge),5df1_A(teal),5df1(brown),5emh_A(lightorange)(RMSD: 0.74-1.30), Red box showed the interaction of residues E151 and N358 colored in yellow.

3.3.7 Determinants of IS specificities:

To delineate the structural basis of the substrate specificity of IS we examined the conformation and interaction of these residues in the substrate bound and free forms of IS crystal structures. In the substrate bound forms these residues neither interact with the ligand nor with the neighboring residues, thus facilitating L1 to remain in “open” state. However, in one of the substrate free structures (PDB ID Code 5COA) Glu161 interacts with Asn358 of L2, enabling L1 to adopt a “closed” state conformation by locking it with L2. These open and closed state analysis of L1 also revealed that IS has two channels, “top-channel” and “side-channel”, for the substrate entry (Fig 3.12A). Given the dimensions of the channel, it is clear that progesterone cannot enter IS through its top-channel (Fig 3.12A and Fig 3.12B). Furthermore, in the mutant IS^{SY} the Glu161- Asn358 interaction gets abolished, which would shift the equilibrium of IS towards the open state, making it amiable for the progesterone to enter the catalytic pocket and orient itself to facilitate the catalytic activity. It is worth noting that this particular mutation did not disrupt the activity of the enzyme towards 10-oxo- Δ^4 -steroid, suggesting that this ligand could equally use the side-channel to enter the catalytic site in the event of perturbations in the top-channel. Interestingly, disruption of Glu161-Asn358 interaction through the N358A mutation also generates activity of IS towards progesterone. However, modest reduction in the activity for 10-oxo- Δ^4 -steroid was observed, indicating that this residue might be involved in the substrate stabilization. Furthermore, glycine at the 162nd position provides required flexibility to L1 to shuttle between the open and closed states. Mutation of this glycine with alanine has shown L1 to adopt a rigid open state confirmation. In addition to the open state of L1, reduction of progesterone requires proper orientation of its cholesterol ring, which is achieved in P5 β R structure through the interaction of its C20 acyl-oxygen with the tyrosine at 156th position. That is why P5 β RY156E mutant shows reduced activity towards progesterone (Fig 3.10). Thus E161S & G162Y mutations in IS is sufficient to make it active towards progesterone.

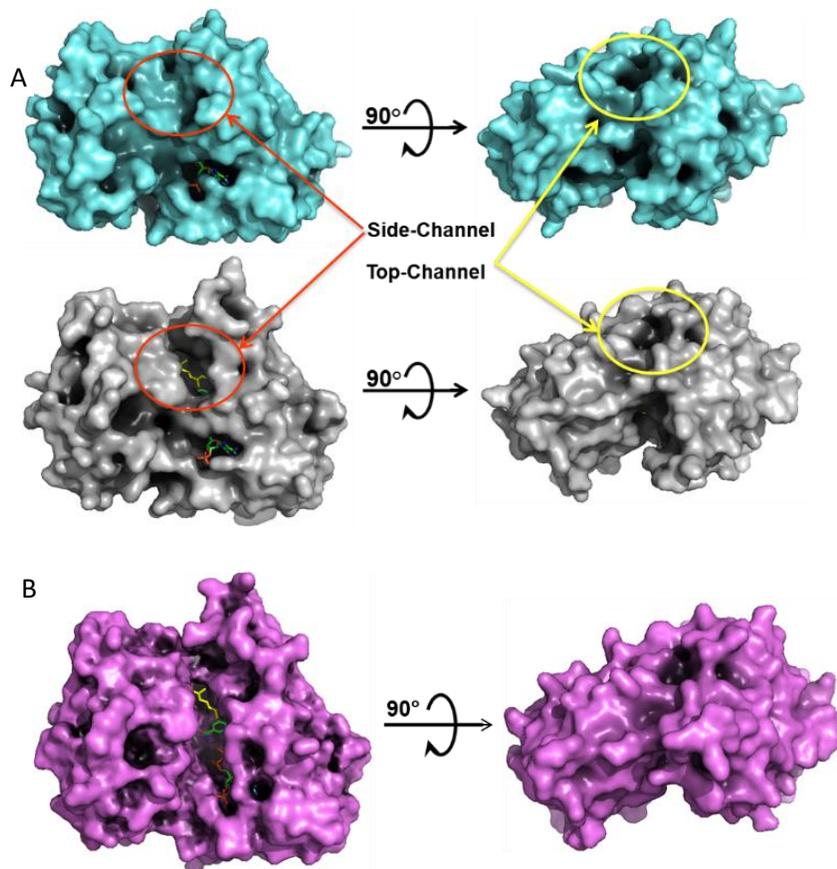


Fig 3.12 A) Surface representation of IS (PDB ID code 5COB) (grey) and P5βR (PDB ID code 2V6G) (cyan) structures showing channels in the proteins for substrate movement. 10-oxogeraniol (yellow) and NADPH (green) are shown in sticks B) Surface representation of IS (PDB ID code 5COA) with closed state conformation. Structures show channels in the protein for substrate movement.

3.3.8 Determinants of P5βR specificities:

Residues present at the side-channel of IS might make it more favorable than the top-channel for 10-oxogeraniol to enter the catalytic site. Perhaps, that is why the side-channel in the substrate free structures of IS is found to be occupied with linear extended ligand like hexaethylene glycol. In contrast with IS structure, P5βR structure has only a narrow top channel, which compels the substrate to enter the catalytic site only through this. Petersen et.al have shown that, even in P5βR subfamily, L1 exhibits gating mechanism, exerted by the phenylalanines present at the rim of catalytic pocket²³. They also reported that the gating mechanism might influence the enzyme specificity for smaller substrates whereas it has marginal effect on progesterone selectivity. Clearly, gating of L1 in P5βR is markedly different from that observed in IS, which might be largely due to the absence of polar interaction between residues at 358th and 161st positions that can modulate the L1 dynamics. Perhaps, the

entry of 10-oxogeranial might also be regulated by the L1 gating, owing to its smaller size and extended conformation. Consequently, difference in the gating mechanism could affect the selectivity of 10-oxogeranial. Additionally, from the superposition of structures of substrate bound IS and P5 β R, it appears that the tyrosine at the 162nd position of the IS (residue numbering as per the IS structure) causes steric clash with the methyl group of 10-oxogeranial, if L1 adopts the conformation found in P5 β R (Fig 3.13). The side-chain oxygen of serine at 155th position interacts with the backbone amide and carboxylic groups of Pro152 residue, thus making L1 less flexible. Therefore, replacing the Tyr156 residue with a short-side-chained polar amino acid and the preceding residue with an amino acid (such as Glycine) that can enhance the flexibility of L1 could make P5 β R active towards 10-oxogeranial. Thus, these observations clearly show that a complex interplay of location of substrate entry channel, gating mechanism of L1 and orientation and stabilization of the substrate are the desideratum for substrate specificity of SDR family of reductases. In case of IS and P5 β R this complex interplay is orchestrated by residues at 161st and 162nd position of the enzyme.

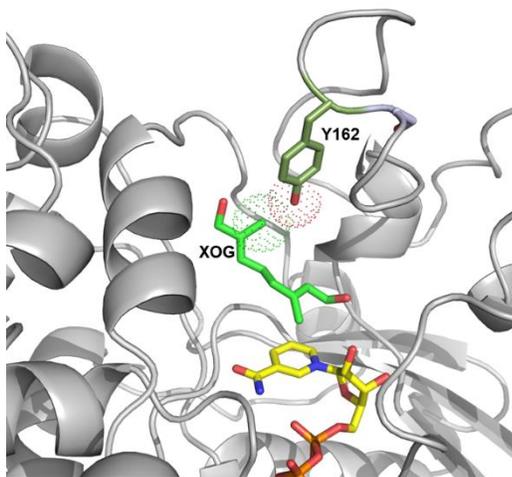


Fig 3.13: Modeled structure of IS^{SY} with confirmation of L1 as in P5 β R. The η O of Tyr162 introduces steric clash with the methyl group of 10-oxogeranial (10-XOG).

3.3.9 Molecular dynamics simulation of IS reveal altered top-channel dynamics:

To further validate the mechanism involved in substrate selectivity, we performed four sets of MD simulations, including the apo and 10-oxogeranial bound IS (PDB ID code 5EMH) and IS^{SY} structures. The simulations were performed for 300 ns in duplicate for each system. The L1 loop in the open state structure (PDB ID code 5EMH) was not resolved and was modelled using coordinates from the ligand bound IS structure (PDB ID code 5COB). Similarly, the IS^{SY} structure was modelled based on the template structure. The structural and dynamic differences between the wild-type and the mutant are mainly confined to the L1-L2 dynamics (Fig 3.14A). In the simulations, a large dynamics was observed at L1 in all four systems. To analyze the differential dynamics caused due to the EG-SY mutation, we calculated a difference plot of pairwise interactions within the protein (Fig 3.14B). In the plot, a value of +1 (colored magenta) implies that the residue pair is closer in IS, whereas a value of -1 (colored green) indicates that the residue pair is closer in IS^{SY} (see methods for details). The analysis shows that the replacement of EG with SY in IS alters both inter and intra loop interactions of L1 and L2). It is interesting to note that intra-loop interactions, especially for L1, alters the flexibility of the loop whereas the inter-loop interaction results in the changes of geometry and nature of the top and side channels. In IS the residues of L1, 153-155 interact with 169-175. However, in the mutant (IS^{SY}) instead of 169-175 stretch, residues 161-165 interact with residue. Thus the change in interaction results in altered geometry of the substrate channels (Fig 3.14C). In IS, the residues 161-169 form L1 with interact with the residues 348-358 of L2 residues, which could act as lid to close the top channel. However, in IS^{SY} interactions within the top channel (L2 and L1 region) are altered. The main interactions of L2 are now formed with the proximal region of the L1 (residues 155-159). The loss in interactions of L2 with the distal region of L1 in IS^{SY} appears to result in a more open top channel (Fig 3.14C). Taken together, the changes in residue-pair contacts indicate a different conformational dynamics of the loop in IS^{SY}, leading to a more open top-channel that could accommodate large ligands like progesterone.

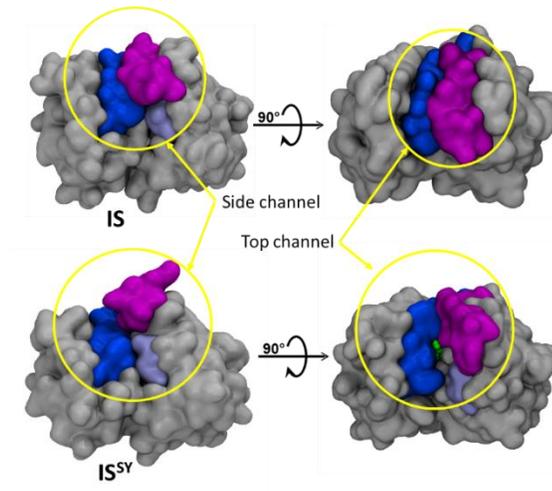
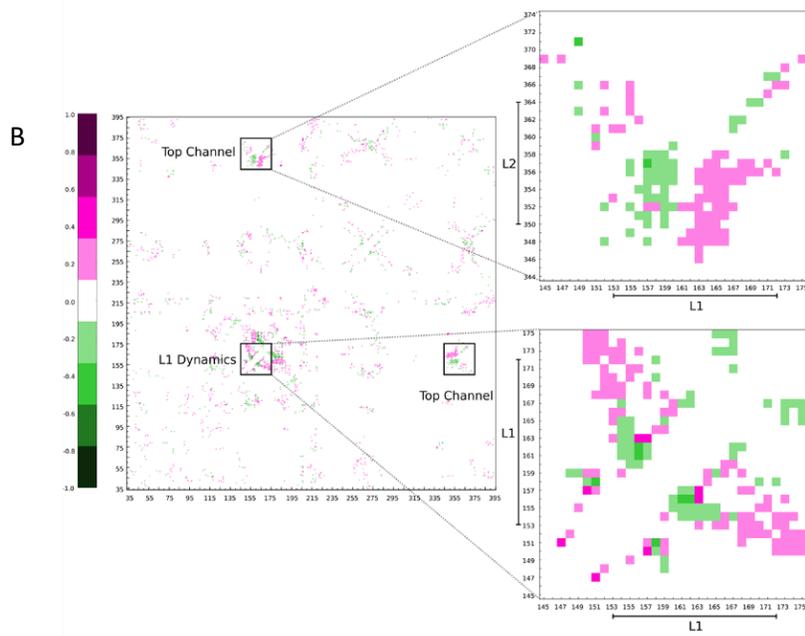
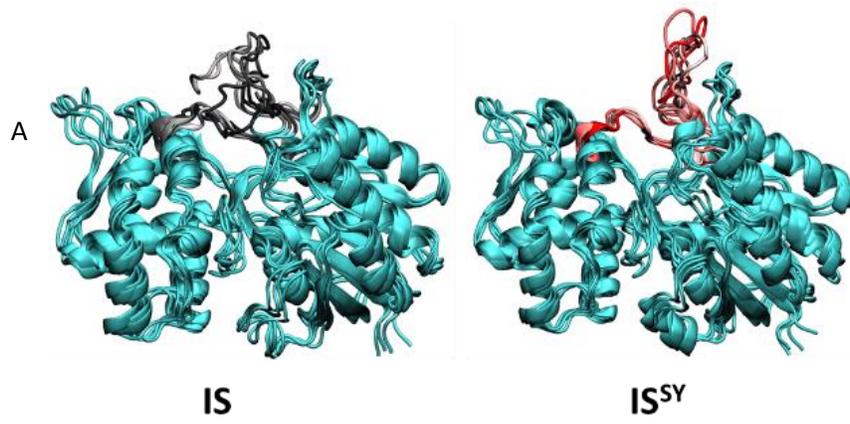


Fig: 3.14 A) Snapshots of the L1 loop conformational dynamics in the IS (Black-gray) and IS^{SY} (Red) simulation systems. B) Difference plot of probability of pair-wise contacts between all protein residues. A value of +1 (magenta) implies that the residue pair is closer in IS, whereas a value of -1 (green) indicates that the residue pair is closer in IS^{SY}. The zoom in plots show the difference in probability of contacts between L1-L1, L1-L1* and L1-L2 regions. C) The surface representation of the wild-type and mutant protein as observed in the simulations. The IS (top) and IS^{SY} (bottom) system shows side channel and a top channel. In IS L1 (purple) and L2 (blue) are closely interacting and ligand is not visible, in IS^{SY} the ligand (green) is clearly visible at the binding site from top channel.

Furthermore, in P5 β R specificity defining residues (SY) are part of a Type II β -turn, whereas the corresponding residues in IS form random coil (Fig 3.15). To understand implications of these structural changes in L1 dynamics we calculated the secondary structure content of this loop. Across 8 sets of simulations, we observe that in the mutant systems, the L1 loop adopts varied secondary structures while the wild-type systems prefer to be in a beta bend or random coil conformation (Fig 3.16). We believe the ability of the mutant to adopt different secondary structural content is governed by the SY mutation and results in the differential loop interactions seen in the difference plots. This could mean that the flexibility of the loop is restricted due to these secondary structural changes, and thus the interactions with the L2 top channel residues are abolished. Overall, the analysis of intra-protein contacts support our structural analysis that the loop dynamics is involved in cavity opening and closing, which are asserted by the residues at 161st and 162nd position of L1. Thus, the substrate cavity dynamics, primarily governed by the two residues of L1 (161 and 162) and a residue at L2 (358), confer the substrate specificity to this class of reductases.

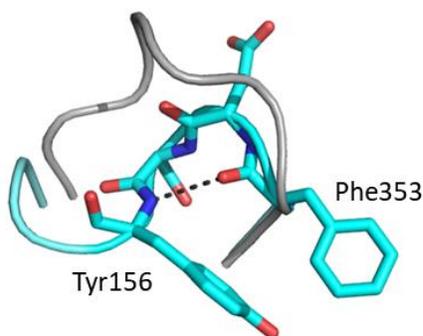


Fig 3.15: Segment of L1 in P5 β R (cyan) carrying the specificity determinants, form type-II β turn. i to $i+3$ residues are shown in sticks. Corresponding segment in IS (grey) adopts random loop conformation.

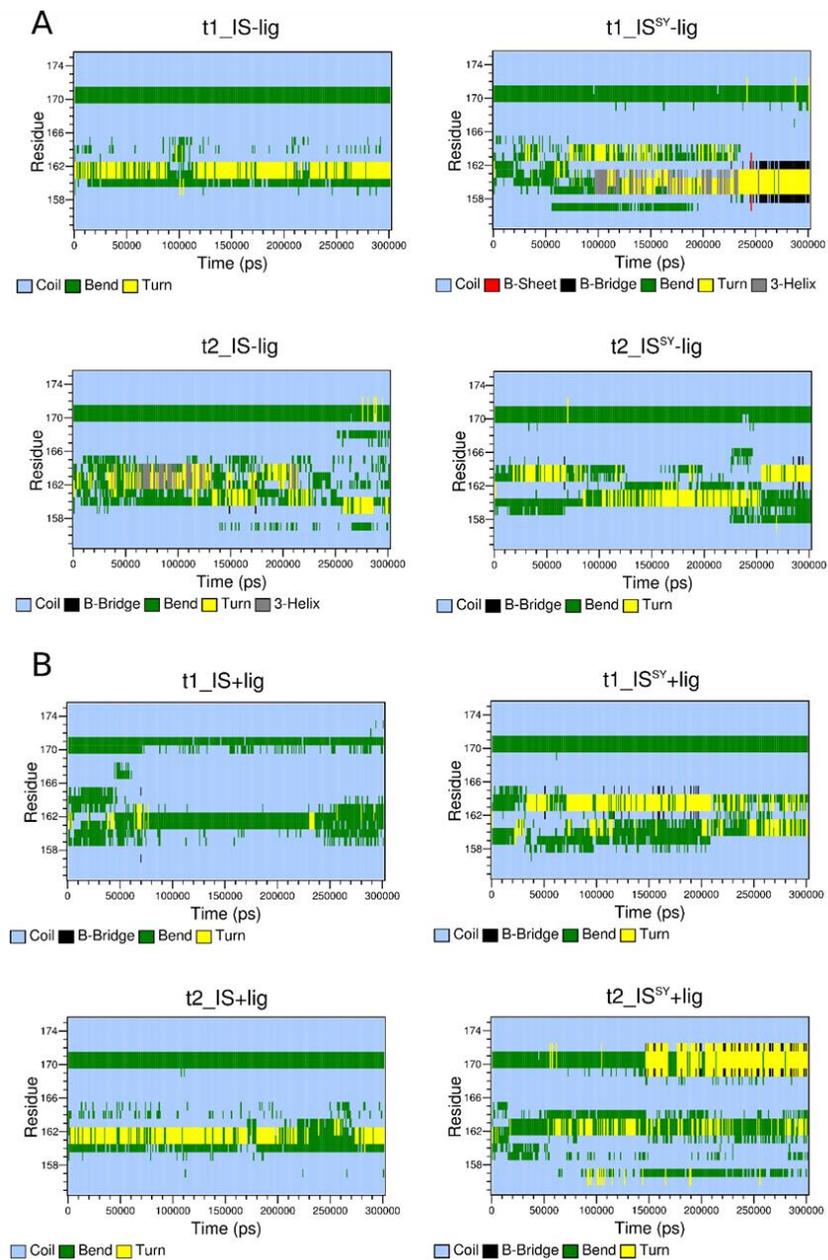


Fig 3.16: Secondary structure content for apo (A) and 10-oxogeraniol bound (B) IS and IS^{SY}. Calculations are carried out for each replicate over the entire simulation time using do dssp utility.

3.4 Discussion:

SDRs have shown tremendous potential for a variety of biotechnological application. Due to its unusual biochemical properties, the recent member of this family, IS has opened up new avenues for the technological exploitation. However, the limited knowledge of specificity determinants of the enzyme has posed serious lacuna in our understanding of its function. In this study, our sequence and structural analysis, site-directed mutagenesis and MD simulations (based on the available structures of IS & P5 β R) have shown that two residue, Glu161 and Gly162, of the loop L1 and Asn358 residue at L2 determine the substrate specificity of the enzymes. Recently, iridoid synthases and their isoforms from other plant sources such as *Swertia mussotii* (SmSI1 & SmSI2)²⁴, *Olea europaea L.* (OeISY, Oe1,4-R1, Oe1,4-R2, Oe1,4-R3 & Oe1,4-R4)²⁵ and *Antirrhinum majus* (AmISY)²⁶ have been characterized for their reductase activity. Interestingly, some of these enzymes exhibit dual specificity, although not with the same extent (Table 3.6). Sequence alignment of these reductases unambiguously show that when 161st, 162nd and 358th position are occupied with Glu, Gly and polar residues, respectively, the enzymes exhibit higher selectivity towards 10-oxogeranial (Fig 3.18 and Table 3.6). However, when Glu161 is replaced with other polar residue, like in AmISY, the enzyme still retains activity towards 10-oxogeranial with modest reduction in its catalytic efficiency. Perhaps, the relative change of catalytic efficiency of the enzymes, compared with CrIS, could also be due to presence of a bulkier residue (Phe/Tyr) at 162nd position. In SmSI2, similar trend in reduction of catalytic efficiency towards 10-oxogeranial is observed, where the polar amino acid at 358th position is occupied by a cysteine. Interestingly, Oe1,4-R3 & Oe1,4-R4 mimic our IS^{SY} mutant with possible dual specificity and indeed kinetic data on Oe1,4-R3 confirms the enzyme is active towards 10-oxogeranial and progesterone. It would be worth to note from our MD simulations that it is not just the geometric and/or electrostatic aspects of the residues at these positions responsible for conferring substrate specificity but the dynamics of L1 and L2 loop steered by them. These observations are further corroborated by the recent report on four iridoid synthases, NcISY1, NcISY2, NmISY1 and NmISY2, from *N. cataria* and *N. mussinii*, respectively²⁷. Although, all the four of them possess identical or similar residues at 161st (Asp), 162nd (Leu/Trp) and 358th (Cys) positions, only one of the homologs (NcISY2 and NmISY2) exhibit robust activity towards 10-oxogeranial. However, in other set of homologs (NcISY1 and NmISY1) L1 is shorter by one residue which could result in altered

dynamics of the loop and hence the binding efficiency of the enzymes for 10-oxogeraniol. The loop conformation of the loops L1 and L2 in the modelled structures of AmISY, SmIS1, SmIS2, OeISY, Oe1,4-R1.1a, Oe1,4-R2, Oe1,4-R3, NmISY1, NmISY2, NcISY1, NcISY2 is similar to that in either IS or P5 β R, depending on their differential affinity towards either 10-oxogeraniol or progesterone (Fig 3.17). Thus, all of these studies substantiate our observations on the substrate specificity of SDR family of reductases. The outcome of this study would help in engineering this class of enzyme for optimal product synthesis.

Table: 3.6 Kinetic parameters of IS (*Catharanthus roseus*) and P5 β R (*Digitalis lanata*) with other ISs such as SmSI1 & SmSI2 (*Swertia mussotii*), OeISY, Oe1,4-R1, Oe1,4-R2, Oe1,4-R3, Oe1,4-R4 (*Olea europaea L.*) and AmISY (*Antirrhinum majus*).

Protein	Substrate <i>K_m</i> (μM)	Substrate <i>K_{cat}</i> (sec⁻¹)
CrIS	1.9 \pm 0.3	4.2 \pm 0.2
SmIs1	7.29 \pm 1.59	0.11 \pm 0.007
AmISY	1.1 \pm 0.1	0.72 \pm 0.02
OeISY	0.6 \pm 0.1	3.8 \pm 0.2
SmIs2	54.37 \pm 4.29	0.034 \pm 0.001
Oe1,4-R1	N.A	
Oe1,4-R2	N.A	
Oe1,4-R3	2900 \pm 400	0.9 \pm 0.08
P5 β R	438 \pm 37	39 \pm 9.6

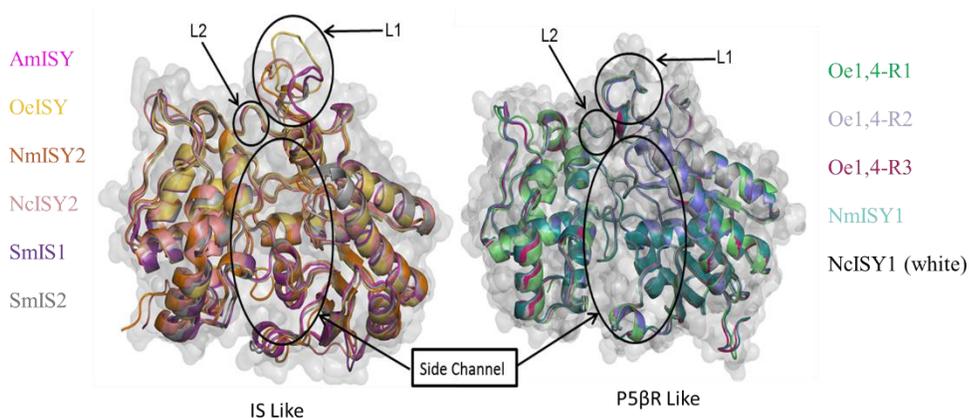


Fig 3.17 Superimposition of modeled IS homologs, AmISY(magenta), OeISY(paleyellow) NmISY2 (Orange) NcISY2 (wheat), SmIS1(Violet), SmIS2(grey) showing similar loop conformation like IS, are known to be active towards 10-Oxogeranial (left) and modeled IS homologs showing similar loop conformation like P5βR Oe1,4-R1 (Green), Oe1,4-R2 (marine), Oe1,4-R3 (hotpink), NmISY1(tblue), NcISY1 (white) are inactive /less active (Oe1,4-R3) towards 10-Oxogeranial (right).

However, when Glu161 is replaced with other polar residue, like in AmISY, the enzyme still retains activity towards 10-oxogeranial with modest reduction in its catalytic efficiency. Perhaps, the relative change of catalytic efficiency of the enzymes, compared with CrIDS, could also be due to presence of a bulkier residue (Phe/Tyr) at 162nd position. In SmSI2, similar trend in reduction of catalytic efficiency towards 10-oxogeranial is observed, where the polar amino acid at 358th position is occupied by a cysteine. Interestingly, Oe1,4-R3 and Oe1,4-R4 mimic our IS^{SY} mutant with possible dual specificity and indeed kinetic data on Oe1,4-R3 confirms the enzyme is active towards 10-oxogeranial and progesterone. It would be worth to note from our MD simulations that it is not just the geometric and/or electrostatic aspects of the residues at these positions responsible for conferring substrate specificity but the dynamics of L1 and L2 loop steered by them. These observations are further corroborated by the recent report on four iridoid synthases, NcISY1, NcISY2, NmISY1 and NmISY2, from *N. cataria* and *N. mussinii*, respectively. Although, all the four of them possess identical or similar residues at 161st (Asp), 162nd (Leu/Trp) and 358th (Cys) positions, only one of the homologs (NcISY2 and NmISY2) exhibit robust activity towards 10-oxogeranial. However, in other set of homologs (NcISY1 and NmISY1) L1 is shorter by one residue which could result in altered dynamics of the loop and hence the binding efficiency of the enzymes for 10-oxogeranial (Fig 3.18). Thus, all of these studies substantiate our observations on the substrate specificity of SDR family of reductases. The outcome of this study would help in engineering this class of enzyme for optimal product synthesis.

```

CrISY      --SEDCQTNATMFKNILNSVIPNASNLQHVCLQTGIKHYFGIFEEG--SKVVPHDSPTED 168
SmIs1     --SEDCETNAVMEFKNILYNIIPVAPNLKHISLQTGSKHYFGIFEEG--SEAVPHESPATED 170
AmISY     --SEDVALNTLMFKNILDSVIPNAPNLKHVSLQTGIKYYWGNMAEMDSTNQPECFYEN 167
OeISY     --NEDCSMNAVMEFQNILNSVIPNAPNLQHICLQTGSKHYIGLFETDT--PESHDTFYSYED 172
SmIs2     TEAECQVNGAMLQNVDAVLPNAPNLSHICIQTGHKHYIGPFETFG--KIKPHETPFIED 168
Oe1,4-R1 TEAECCKVNGGMLQNVLDVAMPNAPNLSHICLQTGKHKHYIGPFETFG--KIKPHETPFVED 169
Oe1,4-R2 TEIENCEANGKMFKNVLDVLIIPNCPNLQHICLQTGRKHKHYIGPFESYG--KIKTNDTPFHED 169
Oe1,4-R3 TEAENCEANGKMFKNVLDALIPNCPNLKHICLQTGTGKHKHYIGPFESYG--KIKTHDPPFHED 169
P5βR     TEQENCEANSKMFERNVLDVAVIPNCPNLKHISLQTGRKHKHYMGPFESYG--KIESHDPPTED 169

CrISY      CSENLISSMNSKELGFLGFRNSMKSEFVSCIDKMRDYRFIP 388
SmIs1     SCVDLVMVSVTKSREFGFLGFRDSMKSEFKSCIHKMRDFKFI 390
AmISY     RNKETLCSMNKNKEFGFMGFRD'TTKSEFVSSINKMRDFKFI 387
OeISY     CLENVLSSTHKNRLHGFMGFRNTYTSFVSCIDKMRAYRFIP 392
SmIs2     SGECMLDSMSKSKDGHGFLGFRD'TTKSLLSVVGKMKAHKIVP 388
Oe1,4-R1 GFECMLDSMNKSKEHGFLGFRNTKYSLISVIDKMKSHKIVP 390
Oe1,4-R2 GNECPLDTMNSKKEHGFLGFRNSKNSFISWIDKVKANKIVP 389
Oe1,4-R3 GSECYLSNMNSKKEHGFLGFRNSKNSFMSWIDKVKAYKIVP 389
P5βR     GNECFLDSMNKSKEHGFLGFRNSKNAFISWIDKAKAYKIVP 389

```

Fig 3.18: Multiple sequence alignment of IS (*Catharanthus roseus*) and P5βR (*Digitalis lanata*) with other ISs such as SmSI1 & SmSI2 (*Swertia Mussotii*), OeISY, Oe1,4-R1, Oe1,4-R2, Oe1,4-R3 & Oe1,4-R4 (*Olea europaea* L.) and AmISY (*Antirrhinum majus*). 161st, 162nd and 358th residues are highlighted in color. Sequences highlighted in yellow color are specific to 10-oxogeranial, sequences highlighted in red are specific to Progesterone.

3.5 Bibliography:

1. Munkert, J. *et al.* Iridoid Synthase Activity Is Common among the Plant Progesterone 5β-Reductase Family. *Mol. Plant* **8**, 136–52 (2015).
2. Tundis, R., Loizzo, M. R., Menichini, F., Statti, G. a & Menichini, F. Biological and pharmacological activities of iridoids: recent developments. *Mini Rev. Med. Chem.* **8**, 399–420 (2008).
3. Lindner, S., Geu-Flores, F., Bräse, S., Sherden, N. H. & O'Connor, S. E. Conversion of substrate analogs suggests a Michael cyclization in iridoid biosynthesis. *Chem. Biol.* **21**, 1452–6 (2014).
4. Sandholu, A. S. *et al.* Dynamics of loops at the substrate entry channel determine the specificity of iridoid synthases. *FEBS Lett.* (2018). doi:10.1002/1873-3468.13174
5. Kries, H. *et al.* Structural determinants of reductive terpene cyclization in iridoid biosynthesis. **12**, 6–11 (2016).
6. Qin, L. *et al.* Structure of iridoid synthase in complex with NADP + / 8-oxogeranial reveals the structural basis of its substrate specificity. *J. Struct. Biol.* (2016). doi:10.1016/j.jsb.2016.02.010
7. Hu, Y. *et al.* Structures of Iridoid Synthase from *Cantharanthus roseus* with Bound NAD⁺, NADPH, or NAD⁺ /10-Oxogeranial: Reaction Mechanisms. *Angew. Chemie Int. Ed.* n/a-n/a (2015). doi:10.1002/anie.201508310
8. Motulsky, H. GraphPad Prism. *Clases* (2007). doi:10.1198/tas.2007.s86
9. DeLano, W. L. The PyMOL Molecular Graphics System, Version 0.99 Schrödinger, LLC. *Schrödinger LLC Version 1.*, <http://www.pymol.org> (2002).
10. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–

- 8593 (1995).
11. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
 12. Huang, J. & Mackerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
 13. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **26**, 1701–1718 (2005).
 14. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, (2007).
 15. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
 16. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
 17. Andrej Šali. MODELLER A Program for Protein Structure Modeling. *Comp. protein Model. by Satisf. Spat. restraints.* (1993). doi:10.1006/jmbi.1993.1626
 18. McCoy, A. J. *et al.* Phaser crystallographic software . *J. Appl. Crystallogr.* (2007). doi:10.1107/s0021889807021206
 19. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
 20. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).
 21. Thorn, A. *et al.* The Crystal Structure of Progesterone 5 α -Reductase from *Digitalis lanata* Defines a Novel Class of Short Chain Dehydrogenases / Reductases * □. *J. Biol. Chem.* **283**, 17260–17269 (2008).
 22. de Castro, E. *et al.* ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**, (2006).
 23. Petersen, J. *et al.* Progesterone 5 β -reductases/iridoid synthases (PRISE): gatekeeper role of highly conserved phenylalanines in substrate preference and trapping is supported by molecular dynamics simulations. *J. Biomol. Struct. Dyn.* **34**, 1667–1680 (2016).
 24. Xiang, B. *et al.* Cloning and characterization of two iridoid synthase homologs from *swertia mussotii*. *Molecules* **22**, 1–14 (2017).
 25. Alagna, F. *et al.* Identification and characterization of the iridoid synthase involved in oleuropein biosynthesis in olive (*Olea europaea*) fruits. *J. Biol. Chem.* **291**, 5542–5554 (2016).
 26. Kries, H., Kellner, F., Kamileen, M. O. & O'Connor, S. E. Inverted stereocontrol of iridoid synthase in snapdragon. *J. Biol. Chem.* **292**, 14659–14667 (2017).

27. Sherden, N. H. *et al.* Identification of iridoid synthases from *Nepeta* species: Iridoid cyclization does not determine nepetalactone stereochemistry. *Phytochemistry* **145**, 48–56 (2018).

Chapter 4

Structural insights into cascade mechanism of IS and 10-HGO

4.1 Background:

As explained in the introduction, in plants the biosynthesis of iridoids involves the oxidation of 10-hydroxygeraniol to 10-oxogeraniol, which subsequently undergoes reductive cyclization to yield an equilibrium mixture of *cis-trans* nepetalactol and iridoidals¹ (Fig 4.1). These two crucial steps are catalyzed by 10-hydroxygeraniol dehydrogenase (10-HGO) and iridoid synthase (IS), supposedly in tandem.

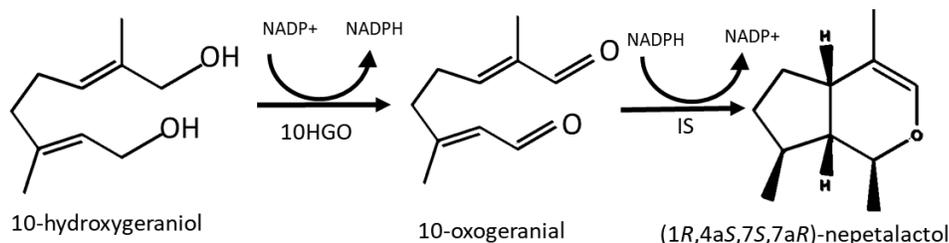


Fig 4.1: Reactions catalyzed by 10-HGO and IS

Amongst these two enzymes IS plays a unique role by catalyzing both reduction and subsequent cyclization of 10-oxogeraniol². Recent structural studies on IS have shown that dynamic rearrangement of loop at this opening of substrate binding pocket, away from the catalytic site, play crucial role in determining the substrate specificity of the enzyme³. Perhaps that is why mutational studies on the residues present at the catalytic site pocket did not reveal much about the factors governing the specificity aspects. Even the product profile of some of the catalytic site mutants of IS, by and large, remained the same. Taking information from these mutational studies as a lead, it was proposed that IS does not catalyze the cyclization of the *enol* or *enolate* intermediate to yield nepetalactol and there could be another enzyme involved in this pathway that catalyzes the cyclization step. Extending this argument, a class of nepetalactol-related short-chain dehydrogenase enzymes (NEPS) from catmint (*Nepeta mussinii*) was suggested to catalyze the stereoselective cyclisation into distinct nepetalactol stereoisomers⁴. Although the class of cyclases present in *Nepeta* species were supposed to be

involved in the cyclization of nepetalactol biosynthetic pathway, a cross correlation between their expression in other plants like *Cataranthus roseus* for the synthesis of iridoids is remains unestablished. Furthermore, the structural and mutational studies on these classes of enzymes fell short of raveling their catalytic mechanism. However, Krithika *et al.*, have shown that when 10-HGO and IS from *Cataranthus roseus* act on 10-hydroxygeraniol in the presence of NADP⁺ (a cofactor for 10-HGO) results in *cis-trans* nepetalactol as a major product. This is in contrast to the observation that when the two catalytic systems carried out sequentially cyclic *cis-trans* nepetalactol is not a major product⁵.

In first part of this chapter we probe the physical interaction of 10-HGO & IS and explore whether such an interaction leads to a concreted chemical synthesis of 10-hydroxygeraniol. Using a surrogate system of 10-HGO-IS Chimera we delineate the structural basis of their interaction. Appendix of this chapter outlines the structural and catalytic mechanism of 10-HGO.

4.2 Methods:

4.2.1 Cloning, protein expression and purification:

To obtain high yields of homogenous recombinant chimera of IS-10-HGO, the genes were cloned in pOPIN-SUMO* expression vector using USER cloning method⁶. Optimum expression was achieved in *E. coli* strain B834 (rare). The cells were grown at 37°C to an OD₆₀₀ of 0.5 and further induced with 0.5 mM IPTG (isopropyl 1-thio-D-galactopyranoside) and incubated for 14 hrs at 18°C. The culture was harvested by centrifugation. The cells were lysed in chilled lysis buffer containing bufferA (50 mM Tris pH-8.0, 500 mM NaCl, 5 mM Imidazole, 2 mM β-mercaptoethanol, and 5% glycerol), one tablet of cComplete EDTA free protease inhibitor (Roche) and a pinch of lysozyme, employing sonication. The lysate was filtered and loaded on to the Ni-NTA affinity column (GE Healthcare Biosciences, USA) for binding. The column was washed with buffer A and the protein was eluted with buffer B (50 mM Tris pH-8.0, 500 mM NaCl, 200 mM imidazole, 2 mM β-mercaptoethanol, and 5% glycerol). The elution fractions were dialyzed along with the TEV protease (purified in our lab) overnight at 10°C. Sumo tag, His tag and TEV protease were removed by the desalting followed by second Ni-NTA affinity column. To improve the purity of the protein, size exclusion chromatography was performed using Hiprep26/60 sephacryl S-300 (GE Healthcare

Biosciences, USA) column in 20 mM Tris pH-8.0, 150 mM NaCl, 2 mM DTT. The peak fractions under specific molecular weight were collected and concentrated to perform activity assays, Cryo-EM experiments. For the protein complementation assay (PCA) DHFR1,2 and DHFR3 genes were cloned in pET22b and pET28A respectively. Further 10-HGO and IS genes were cloned in pET22b-DHFR1,2 and pET28A-DHFR3 vectors respectively.

4.2.2 Activity assay:

GC-MS based assay has been performed to analyze enzyme activity. The reaction mixture contained 100-200 µg of enzyme, 200 µM 10-hydroxygeraniol and 200 µM NADP⁺ in 500 µL of buffer containing 20 mM MOPS pH 7.0, 10% glycerol. Reaction mixtures were incubated at 30°C and was terminated after 30 min by adding 1 ml of TBME (Tert-Butyl Methyl Ether). Compounds were extracted in TBME, 1-5 µl of extracted compounds were injected into gas chromatography (GC)–mass spectrometry (MS) for analysis using following method equilibration time for 0.5 min, oven program is 60°C for 0 min then 20 °C/min to 120°C for 0 min then 2.5 °C/min to 160°C for 0 min then 20°C/min to 190°C for 2.5 min. The separation was performed on Rtx-5ms column (30 m x 0.25 mm x 0.25 µm) using helium as carrier gas at 1 ml min⁻¹ and with an injector temperature of 220°C.

4.2.3 Bacterial DHFR PCA survival selection assay:

10-HGO-DHFR1,2-pET22b and IS-DHFR3-pET28A plasmids were co-transformed into B121* cells. Transformed cells were washed with sterile PBS for three times in order to remove LB media, resuspended in fresh PBS. Resuspended cells were plated on DHFR PCA selective medium which contains 2% noble agar, 1x M9 salts solution, 0.4% glucose, 2 mM MgSO₄, 0.1 mM CaCl₂, 100 mg/ml ampicillin (for pET22b-10-HGO-DHFR1,2 vector selection), 25 mg/ml kanamycin (for pET28A-IS-DHFR3 vector selection), 1 mM IPTG, 10 mg/ml trimethoprim (Sigma), 800 mg/ml casamino acids (MP biomedical) and 10 mg/ml thiamine (Sigma). The plates were incubated at 37°C for 48-72 hrs. The appearance of colonies indicates that were IS and 10-HGO interacting. The obtained colonies were subjected to protein expression as described above for further validation using western blot.

4.2.4 Western blot:

Protein expression was done from the colonies obtained from DHFR PCA survival assay. The cells were harvested and lysed in chilled lysis buffer. The lysate was centrifuged at 14000 rpm for 30 min the supernatant and pellet samples were fractionated by SDS-PAGE and transferred to a polyvinylidene difluoride (PVDF) membrane using a transfer apparatus according to the manufacturer's protocols (Invitrogen). After incubation with 5% nonfat milk in TBST (10 mM Tris, pH 8.0, 150 mM NaCl, 0.5% Tween 20) for 60 min, the membrane was washed once with TBST and incubated with antibodies against anti-His (1:3000) (Thermo) at RT for 1 hr. Membranes were washed three times for 10 min and incubated with a 1:3000 dilution of horseradish peroxidase-conjugated anti-rabbit antibodies for 30 min. Blots were washed with TBST three times and developed with the diaminobenzoic acid (DAB) tablets (Sigma) according to the manufacturer's protocols.

4.2.5 SEC-MALS:

In order to measure the accurate molecular weight and oligomeric state of the purified chimera SEC-MALS analysis was performed. For SEC we used Superdex 200 10/300 GL column. (GE Healthcare) SEC column connected to an Agilent HPLC system coupled with the 18-angle light scattering detector (Wyatt Dawn HELIOS II) and a refractive index detector (Wyatt Optilab TrEX). Molecular weight was calculated using ASTRA software (Wyatt Technologies).

4.2.6 Cryo-EM:

4.2.6.1 Sample preparation:

Protein vitrification was achieved by the assistance of a FEI vitrobot. Firstly, Quantifoil 1.2/1.3, 300 mesh grids were glow discharged in air for 90 seconds at 25 mA, and then placed on a pair of tweezers in the vitrobot blotting chamber. Chamber temperature was set to 18°C with 100% humidity to avoid evaporation after blotting. 3 μ l of 1 μ g/ μ l protein sample was applied on the grid for 30 sec to allow protein to absorb onto the grid holes. The grid was automatically blotted for 3 sec, followed by plunge-freezing in liquid ethane. Frozen grids were stored in cryo grid boxes in liquid nitrogen.

4.2.6.2 Data collection and processing:

Grids were transferred to Krios TEM and images were collected at 59000 X (corresponding to a pixel size of 1.38 Å/pixel) with a Falcon 3 direct detector which is operating in counting mode. The total exposure given was 1 min and the dose was ~16.4 electrons. The movies recorded in 25 fractions. The defocus range used for automated data collection was 2.4 to 3.6 µm with intervals of 0.3 µm. The obtained micrographs were subjected to drift correction, beam-induced motion and dose-weighting with MotionCor235⁷. Contrast transfer function (CTF) fitting and phase-shift estimation were performed on motion-corrected sums without dose weighting with Gctf-v.1.1836⁸. Micrographs were manually checked and a total of 102 micrographs used for all other image processing. An initial round of manual picking on small subset of micrographs was performed in RELION 2.0⁷. Picked particles were subjected to reference-free 2D classification and good classes used for auto-picking on the entire dataset. Auto-picked particles were manually inspected and false positives removed, resulting in a total of ~106500 particles.

4.3 Results and discussion:

4.3.1 10-HGO and IS physical interaction to produce cyclic nepetalactol:

Previous studies have shown that when 10-HGO and IS from *Cataranthus roseus* act on 10-hydroxygeraniol in the presence of NADP⁺ (a cofactor for 10-HGO) results in *cis-trans* nepetalactol as a major product, contrary to their interaction in tandem. This observation clearly hints that 10-HGO and IS could physically interact to transfer the product of the former as a substrate of the later in a hand-shake manner to produce *cis-trans* nepetalactol. Thus from this observation we hypothesize that 10-HGO and IS are sufficient to produce *cis-trans* nepetalactol *in vivo* without the involvement of any other cyclases. To test whether these two enzymes physically interact in the hypothesized manner we performed protein complementation assay. 10-HGO and IS were tagged with two fragments of murine DHFR (dihydrofolate reductase) (Fig 4.2A). DHFR is an enzyme essential for prokaryotic cells for the synthesis of thymidylate, purines, methionine, serine and pantothenate and folds to its functional shape only when the two fragments of it are physically brought together⁹. *E. coli* were transformed with plasmids of different antibiotic resistance, carrying fragments of DHFR with 10-HGO and IS (see methods section). These cells were plated on trimethoprim (potent

inhibitor of *E. coli* DHFR but not of murine DHFR) containing agar. Transformed *E. coli* cells resulted in the formation of colonies, indicating that 10-HGO and IS indeed physically interact (Fig 4.2A). Thus the complementation assay supports the physical interaction of 10-HGO and IS. As a negative control we used arf RhoGTPase protein cloned with DHFR1,2. This plasmid transformation along with IS-DHFR3 resulted in no colonies.

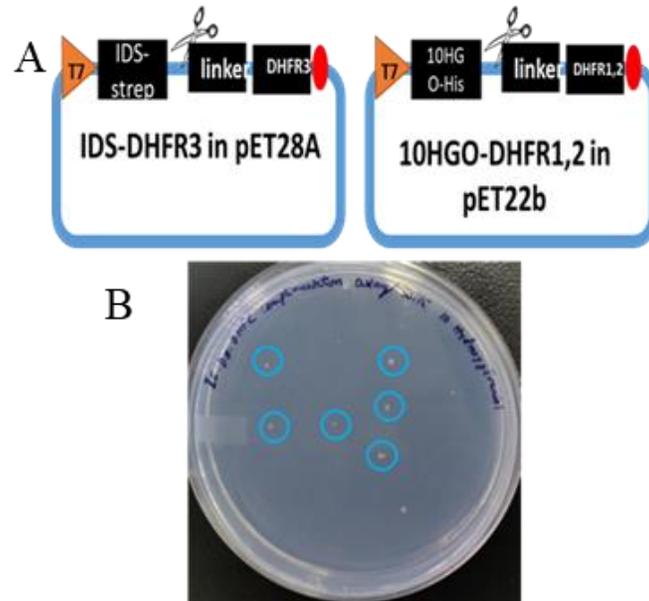


Fig 4.2: A) schematic representation of IS-DHFR3 and 10-HGO-DHFR2 plasmids B) Agar plate showing colonies of IS-DHFR3 and 10-HGO-DHFR2 co transformation.

4.3.2 Chimera of 10-HGO and IS as surrogate system:

To further characterize the 10-HGO-IS interaction an *in vitro*, complexation of the two proteins on size-exclusion column was tried. However, the proteins co-eluted rather forming a complex (Fig 4.3A) with higher molecular weight indicating the interaction to be a weak or highly transient. In order to obtain stable complex, we generated a chimera of 10-HGO and IS separated by 10 amino acids linker using genetic engineering (Fig 4.3B). The SEC-MALS of recombinant chimera showed it be a dimer with a molecular weight of ~160 kDa, which is in line with the fact that both 10-HGO and IS exist as natural dimers (Fig 4.3C) having an individual molecular weight of ~80 kDa. The GC-MS analysis revealed that the chimera has activity similar to that of a physical mixture of 10-HGO-IS and thus could potentially act like a surrogate system for further characterization of the 10-HGO-IS interaction (Fig 4.3D). To understand the structural basis of 10-HGO and IS interaction attempts were made to crystalize

the chimera but with no success. The lack of crystallizability of the chimera could be due to the transient interaction of 10-HGO and IS

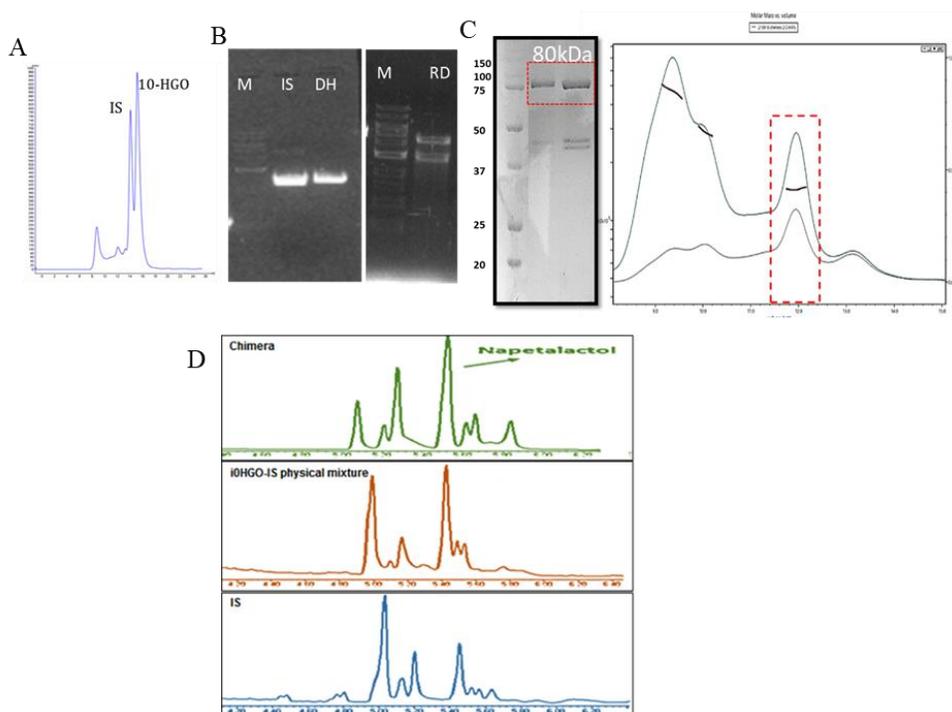


Fig 4.3: A) SEC profile of 10-HGO and IS in vitro complex. B) Agarose gel of chimera PCR and RD. C) SDS-PAGE of chimera and MALS results fitting of chimera. Peak corresponds to tetramer showed in red dotted box. D) GCMS product profile of chimera, 10-HGO-IS physical mixture and only IS.

4.3.3 Structural basis of 10-HGO-IS interaction:

Therefore, to analyze the structures, we collected 550 movies under cryo conditions using Titan Krios TEM fitted with a CMOS Falcon 3 direct detector. Further we performed 2D classification using RELION 2.1 with ~106500 auto picked particles and obtained 2D class averages. Analysis of the data showed that the chimera particles on the grids suffered from preferred orientation problem. Furthermore, due to lack of stability of the chimera the particles found on the grids were heterogeneous. Hence, the complete *ab initio* 3D reconstruction was not achieved. However, the 2D classes provided invaluable insights on the nature of interaction between 10-HGO and IS. 2D class based manual docking of crystal structures of 10-HGO (Appendix) and IS (PDB code: 5EMH) revealed the substrate binding channel of IS monomer getting aligned with the catalytic pocket of 10-HGO, in a handshake manner. Thus direct transfer of 10-oxogeranial, product of 10-HGO to the substrate binding pocket of IS. It is

interesting to note that an arrangement of this nature is possible only if the interaction between 10-HGO and IS is mediated by their monomers rather than their natural oligomers (dimers). Thus, the overall study suggests that the interaction between IS and 10-HGO is stoichiometric dependent. Furthermore, the available 2D classes clearly indicate that monomer to monomer interaction is more favourable because the product exit channel of 10-HGO and the substrate entry channel of IS could align with each other to transfer the product of former (10-oxogeranial) as the substrate to the later, in a “hand shake” manner. To evaluate our model we generated 2D classes from the coordinates. From visual inspection, the docked structure mimics the plausible interaction between IS and 10-HGO.

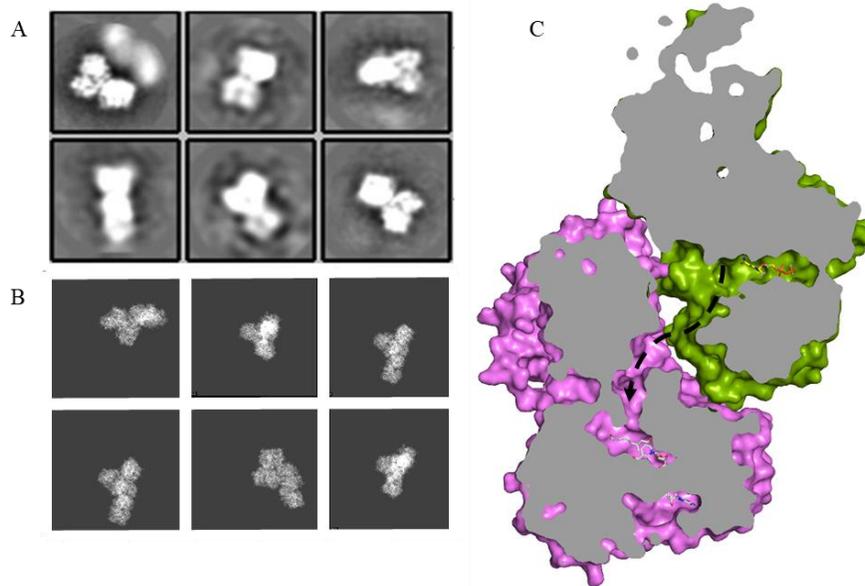


Fig 4.4 A) 2D class averages of chimera. B) 2D projection of docked IS-10-HGO structure map. C) Putative structure model of chimera showing the cavity through which substrate transfer can occur from 10-HGO (green) active site to IS (magenta).

Further, we asked whether the IS and 10-HGO interaction has any evolutionary basis. The co-evolution is used to understand the coordinated changes that occur in pairs of biomolecules. The protein-protein co-evolution study clearly showed the correlation coefficient of 85% between two protein families. This indicates that these proteins indeed interact with each other and coevolved together during the evolution.

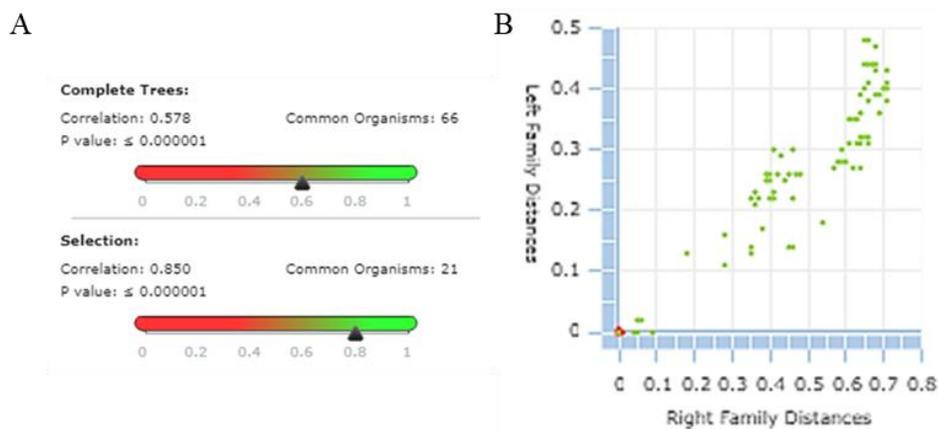


Fig4.5: A) Coevolution analysis for 10-HGO and IS showing correlation with selected similar species. B) Distance plot of selected sequences from phylogeny.

4.4 Conclusion:

Thus, the overall study suggests that interaction between IS and 10-HGO is stoichiometric dependent. Furthermore, the orientations of 2D classes clearly indicate that monomer to monomer interaction is more favorable because the product exit channel of 10-HGO and the substrate entry channel of IS could align with each other to transfer the product of former (10-oxogeraniol) as the substrate to the later, in a “hand shake” manner.

4.5 Bibliography:

1. Geu-flores, F. *et al.* cyclization in iridoid biosynthesis. 2–8 (2012). doi:10.1038/nature11692
2. Lindner, S., Geu-Flores, F., Bräse, S., Sherden, N. H. & O'Connor, S. E. Conversion of substrate analogs suggests a Michael cyclization in iridoid biosynthesis. *Chem. Biol.* **21**, 1452–6 (2014).
3. Sandholu, A. S. *et al.* Dynamics of loops at the substrate entry channel determine the specificity of iridoid synthases. *FEBS Lett.* (2018). doi:10.1002/1873-3468.13174
4. Lichman, B. R. *et al.* Uncoupled activation and cyclization in catmint reductive terpenoid biosynthesis. *Nat. Chem. Biol.* **15**, 71–79 (2019).
5. Krithika, R. *et al.* Characterization of 10-Hydroxygeraniol Dehydrogenase from *Catharanthus roseus* Reveals Cascaded Enzymatic Activity in Iridoid Biosynthesis. *Sci. Rep.* **5**, 1–6 (2015).
6. Bitinaite, J. *et al.* USERTM friendly DNA engineering and cloning method by uracil excision. *Nucleic Acids Res.* (2007). doi:10.1093/nar/gkm041
7. Scheres, S. H. W. Single-particle processing in RELION. *Manuals* (2018).

8. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* (2016). doi:10.1016/j.jsb.2015.11.003
9. Remy, I., Campbell-Valois, F. X. & Michnick, S. W. Detection of protein-protein interactions using a simple survival protein-fragment complementation assay based on the enzyme dihydrofolate reductase. *Nat. Protoc.* **2**, 2120–2125 (2007).

Structural studies of 10-hydroxygeraniol dehydrogenase from *C. roseus*

4.6 Background:

The initial committed step for the synthesis of secologanin is oxidation of 10-hydroxygeraniol to 10-oxogeraniol and subsequent reduction followed by cyclization of 10-oxogeraniol to nepetalactol¹. The first step is catalyzed by a zinc-dependent alcohol dehydrogenase enzyme called 10-hydroxygeraniol dehydrogenase, belongs to a family of medium chain dehydrogenase/reductase (MDR). MDR superfamily alcohol dehydrogenases share high sequence conservation and adopt Rossmann fold². Despite of this, they exhibit divergent substrate specificity and some of them are also promiscuous in terms of their substrates. The substrates of these alcohol dehydrogenases are classified into primary, secondary and tertiary alcohols³. Recent studies have shown that enzymes like geraniol dehydrogenases exhibit regiospecificity towards various primary alcohols^{3,4}. Interestingly these enzymes recognize an allelic site on C2 of primary alcohols which is essential for the enzyme catalysis⁴. This was verified by the previous studies, where 10-HGO failed to catalyze the oxidation of, β -citronellol which lacks a double bond at C2 carbon. These observations suggest that 10-HGO enzyme in *C. roseus* might have evolved as a *bona fide* member of iridoid biosynthesis pathway to catalyze the oxidation of 10-hydroxygeraniol (a C10 primary alcohol substrate derived from geraniol by geraniol 10-hydroxylase) to 10-oxogeraniol.

4.7 Methods:

4.7.1 Cloning, protein expression and purification

To obtain high yields of homogenous recombinant 10-HGO, the genes were cloned in pRSET expression vector. Optimum expression was achieved in *E. coli* strain BL21*. The cells were grown at 37°C to an OD₆₀₀ of 0.5 and further induced with 0.5 mM IPTG (isopropyl 1-thio-D-galactopyranoside) and incubated for 14 hrs at 18°C. The culture was harvested by centrifugation. The cells were lysed in chilled lysis buffer containing bufferA (50 mM Tris pH-8.0, 500 mM NaCl, 5 mM Imidazole, 2 mM β -mercaptoethanol, and 5% glycerol), one tablet

of cOmplete EDTA free protease inhibitor (Roche) and a pinch of lysozyme, employing sonication. The lysate was filtered and loaded on to the Ni-NTA affinity column (GE Healthcare Biosciences, USA) for binding. The column was washed with buffer A and the protein was eluted with buffer B (50 mM Tris pH-8.0, 500 mM NaCl, 200 mM imidazole, 2 mM β -mercaptoethanol, and 5% glycerol). The elution fractions were dialyzed along with the TEV protease (purified in our lab) overnight at 10°C. The His-tag and TEV protease were removed by the desalting followed by second Ni-NTA affinity column. To improve the purity of the protein, size exclusion chromatography was performed using Hiprep26/60 sephacryl S-300 (GE Healthcare Biosciences, USA) column in 20 mM Tris pH-8.0, 150 mM NaCl, 2 mM DTT. The peak fractions under specific molecular weight were collected and concentrated to perform crystallization.

4.7.2 Substrate docking:

The molecular interaction of the 10-HGO with its ligand 10-hydroxygeraniol was predicted using Glide software⁵. The ligand coordinates were obtained from PubChem server. In first step grid box was generated for receptor using receptor grid generation module. Ligands were docked to the protein by using Glide module. The standard precision (SP) mode was used for docking. The docked ligand conformations were evaluated based on glide score⁶.

4.8 Results and discussion:

4.8.1 Cloning, expression and purification of 10-HGO:

10-HGO gene was cloned in pRSET bacterial expression vector with N-terminal 6xHis-tag followed by TEV protease cleavage site. The protein was expressed in B121* strain. The protein was soluble with His-tag, as apparent from SDS-PAGE with lane belonging to the supernatant fraction (Fig 4.6A). Post TEV treatment followed by gel filtration, protein eluted in expected fractions. Purified protein was used for crystallization.

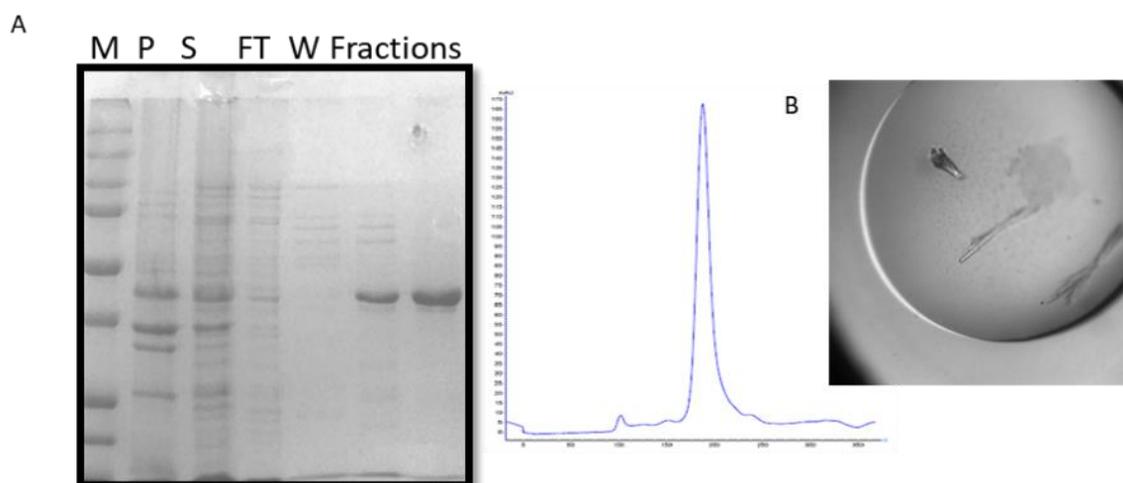


Fig 4.6 A) SDS-PAGE images and gel filtration profile of 10-HGO purification. B) Crystals of 10-HGO.

4.8.2 Crystallization and structure determination of 10-HGO:

Crystallization was carried out with concentrated protein (10 mg/ml) along with co-factor in 5 molar excess. 1 μ l of protein and 1 μ l of mother liquor (0.1 M MES pH 6.5, 15% MPD and 15% PEG4000, 2 M NaCl) were mixed and allowed to incubate at 22°C in the under oil micro batch method. The diffractable crystals were obtained within a day. The crystals were flash frozen in liquid nitrogen using silicon oil as a cryoprotectant. The data sets were collected in the Elettra synchrotron beamline XRD-2, Trieste, Italy. Data was integrated and scaled with XDS⁷ & SCALA respectively⁸. The structure solution was obtained by molecular replacement with phaser using sinapyl alcohol dehydrogenase (PDB CODE: 1YQD) as a search model. Iterative rounds of model building refinement were carried using COOT⁹ and REFMAC5¹⁰ till the R and R_{free} of 0.21 and 0.25. Diffraction data and refinement statistics are summarized in Table 4.1.

4.8.3 Overall structure of 10-HGO:

The recombinant 10-HGO was crystallized in trigonal space group P3₂12 and diffracted up to 2.4Å. The structure of the protein was solved using molecular replacement (Table 4.1). The asymmetric unit of 10-HGO (PDB CODE: 6K3G) crystal lattice was composed of one molecule. However, SEC experiment suggested that 10-HGO exists as a dimer in solution. Therefore, we submitted 10-HGO coordinates to PISA online server which showed the dimeric conformation of two monomeric subunits with two-fold symmetry¹¹. The dimeric assembly of

10-HGO stabilized through a hydrogen bonding between anti-parallel β -strands of each monomer. The overall structure was found to be similar to its homologue sinapyl alcohol dehydrogenase (SDR) (RMSD \sim 0.51). There is well defined electron density for NADP⁺ and two Zn²⁺ atoms were observed (Fig 4.7A). Structural superposition of 10-HGO and SAD showed that the positions of NADP⁺ and Zn²⁺ atoms of 10-HGO are identical to SAD. Like other Zn²⁺ dependent MDR proteins, each subunit of 10-HGO consist of two domains, one is C-terminal cofactor-binding domain with conserved Rossmann fold. This composed of six-stranded parallel β -sheet sandwiched between α -helices on each side. The second one is N-terminal substrate binding domain with a core of antiparallel β -strands and surface-positioned α -helices. These two domains are separated by a cleft containing a deep active site pocket. Similar to other MDRs, 10-HGO composed of two Zn²⁺ ions where one is catalytic Zn²⁺ and other is a structural Zn²⁺ (Fig 4.7 B, C) Unlike in tetrahydroalstonine (THAS)¹², the catalytic Zn²⁺ in 10-HGO is highly conserved and could play role in stabilizing the alkoxide intermediate. In the 10-HGO structure, catalytic Zn is coordinated with three conserved residues including Cys50, His72, Cys166 and a conserved water molecule in a tetrahedral geometry (Fig4.7 B). During the catalysis, binding of substrate facilitates the displacement of the conserved water from the Zn²⁺ and release a proton to the solvent. This substrate intermediate is stabilized by the catalytic Zn²⁺. Further the hydride transfer takes place from substrate intermediate to the oxidized cofactor (NADPH) and converted to an aldehyde/ketone product. The enzyme product and reduced NADPH exit the active site to complete the ordered bi-bi mechanism. In addition to catalytic Zn²⁺, we observed another Zn²⁺ coordinated with conserved residues like Cys103, Cys106, Cys109, and Cys117 (Fig 4.7C). However, the role of this structural Zn is not known, but some reports suggest that it could regulate the substrate entry and exit¹³.

Table 4.1 Diffraction Data collection and refinement statistics

Data collection statistics	
Space Group	P3212
a (Å)	53.18
b (Å)	53.18
c (Å)	252.77
α(°)	90
β(°)	90
γ(°)	120
Resolution limits (Å)	42.16-2.41
R_{merge}	0.12 (0.849)
I/s (I)	10.9 (2.7)
Number of reflections	110176 (11597)
Unique reflections	16295 (1645)
Completeness (%)	99.5 (99)
Multiplicity	6.8 (7.0)
Refinement statistics	
Resolution limits (Å)	42.17-2.41
Number of reflections	16307
Working set	15375
Test set	849
R_{work}/R_{free}	0.21/0.26
Number of atoms	
Protein	2767
Ligand (NAP)	48
Ligand (Zn)	2
Water	31
Mean B-factors	
Protein atoms (Å²)	62.120
Ligand	66.86
Water	45.922
RMSD from Ideal values	
Bond length (Å)	0.0114
Bond angles (°)	1.873
Ramachandran plot Statistics	
Preferred (%)	92
Allowed (%)	8

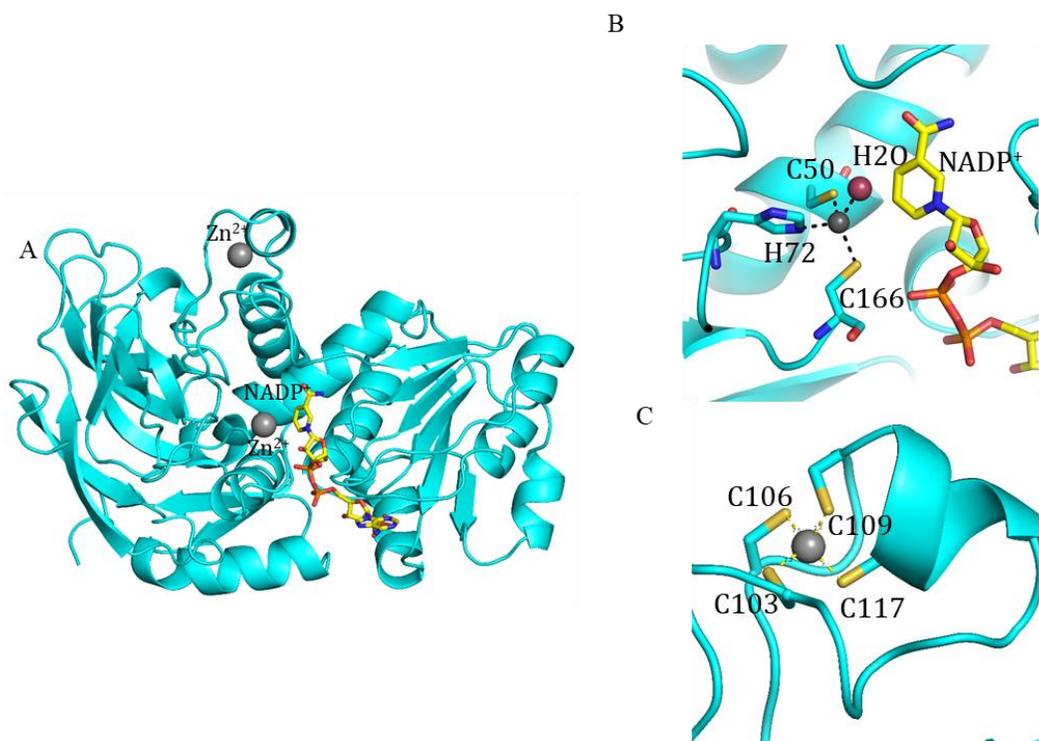


Fig 4.7: A) Overall structure of 10-HGO, NADP shown in yellow sticks and Zn^{2+} shown in grey spheres. B) Active site of 10-HGO contains one catalytic Zn^{2+} (grey sphere) coordinated by C50, H72, C166 and conserved H_2O (Red sphere) with tetrahedral geometry. C) Position of structural Zn coordinated by C103, C106, C109, C117.

4.8.4 Active site and catalytic mechanism:

The deep cleft between the catalytic and cofactor-binding domains contains one NADP⁺ molecule and a catalytic Zn^{2+} . The catalytic Zn^{2+} is found to be important for the catalytic mechanism of MDR superfamily enzymes. The close observation of the structure showed that the residues involved in the binding of NADP⁺ is facilitated by both catalytic domain and cofactor binding domain. The side chains of Ser52 and His55 form H-bond respectively with 2'-OH and 3'-OH group of ribose sugar of NADP⁺. Furthermore, the side chains of His51 and Arg348 forms electrostatic interaction with pyrophosphate groups of NADP⁺. A part from these interactions Gly194 and Leu195 backbone amide groups forms H-bonds with pyrophosphate groups of NADP⁺. On the other hand the side chain amide group of Lys219 and Ser216 interact with the phosphate group of adenosine moiety. It is worth to note that Ser216 is responsible for the specificity of cofactor. Additionally, O7 and N7 of nicotinamide ring establish H-bond with main chain amide group of Ile303 and side chain OH of Ser301, respectively and stabilize the NADP⁺ interaction. These interactions also facilitate

its dissociation once it reduced to NADPH. In order to understand the mode of substrate binding and complete catalytic mechanism we docked 10-hydroxygeraniol in the NADP bound 10-HGO structure, as the protein did not crystallize with bound substrate. From the docked structure and superimposition with other homolog structures of 10-HGO, we found that the complete active site of a MDR family proteins covered by four loops (58-65, 111-120, 277-283, 300-305) which are shown to be regulate the binding pocket volume and also play major role in specificity of each enzyme. The substrate binding pocket of 10-HGO lined with majorly hydrophobic amino acids including Trp61, Phe63, Cys98, Leu122, Val277, Ala279 and Ile303 which are distributed in all four loops and most of them conserved among the alcohol dehydrogenases (Fig 4.8A, B). The hydrophobic environment helps in stabilization and proper orientation of the substrate. As seen in other Zn^{2+} dependent MDRs, the docked the hydroxyl group of C1 in 10-hydroxygeraniol within the distance of coordination bond to Zn^{2+} . Similarly, this OH group is within hydrogen bond distance to hydroxyl group of Ser52 side chain. On substrate binding, the water coordinating with Zn^{2+} is replaced by the OH group of the substrate. The substrate loses one proton into the solvent and form coordination bond with Zn^{2+} for its intermediate state stability. Further, there will be relay of hydride transfer takes place from Ser52 to alkoxide intermediate and then to C4 of $NADP^+$ (Fig 4.8C). This facilitates the formation of 10-oxogeraniol or 10-hydroxygeranial intermediate depending upon the close proximity of C1 or C8 of substrate with Zn^{2+} and cofactor. These intermediates further oxidized and form 10-oxogeranial. The previous studies suggest that C1 of 10-hydroxygeraniol is more susceptible to oxidation compared to C8³. The preference of C1 and C8 alcohol groups determined by the interaction of its methyl groups at C2 and C6. Furthermore, the preference of this enzyme towards the C2 allelic primary alcohol substrates like 10-hydroxygeraniol, geraniol and nerol compared to non-allelic primary alcohol substrate like citronellol is still elusive.

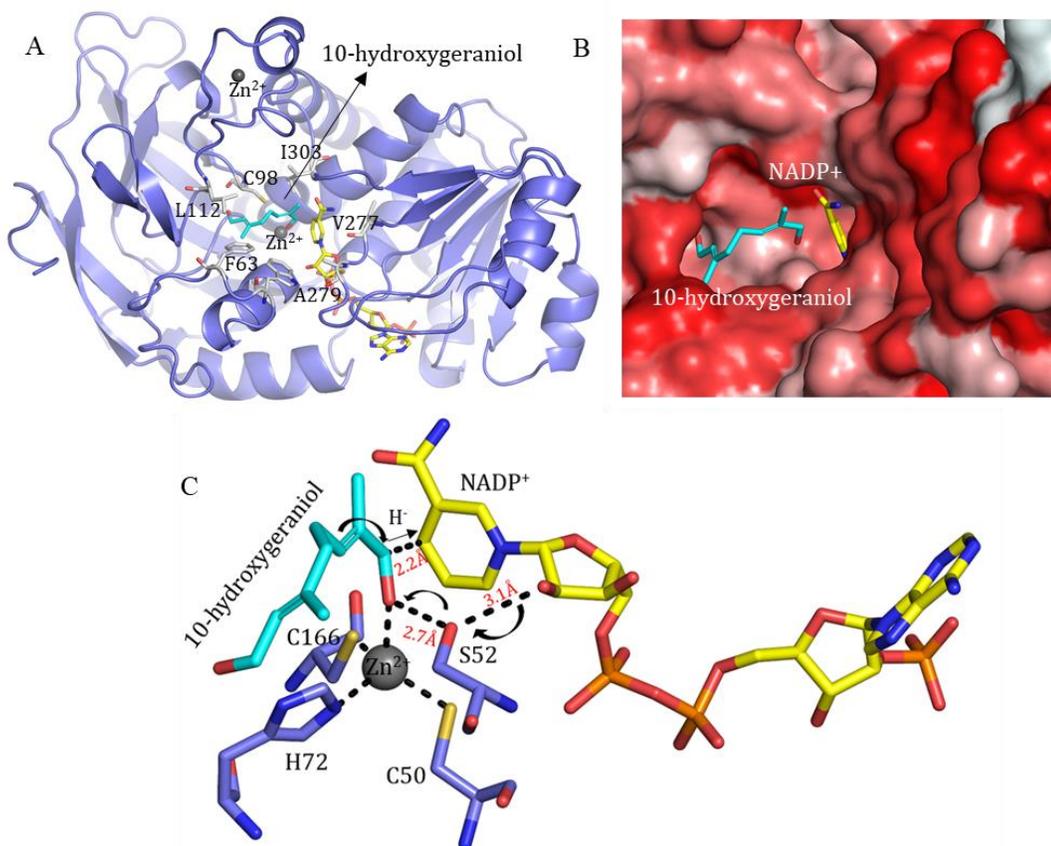


Fig 4.8. A) Docked structure of 10-HGO with 10-hydroxygeraniol, hydrophobic residues around binding pocket shown in white sticks, 10-hydroxygeraniol shown in cyan. B) Hydrophobic cavity of 10-HGO, Red surface indicates hydrophobicity. C) Proton shuttling mechanism during the reduction process in the active site of 10-HGO. NADP⁺ shown in yellow sticks, 10-hydroxygeraniol shown in cyan, Zn²⁺ shown in grey sphere, catalytic and Zn²⁺ co-ordination residues shown in blue sticks.

4.9 Conclusion:

This is the first structure of geraniol dehydrogenases so far. Our structural studies revealed that the overall structure of 10-HGO is similar to MDR family alcohol dehydrogenases. The substrate docking studies helped in determining the enzyme catalytic mechanism. The conserved Ser52 play an important role in proton relay. The hydrophobic substrate binding pocket is crucial for the substrate binding and the residues laid in the binding pocket determines the enzyme specificity.

4.10 Bibliography:

1. Krithika, R. *et al.* Characterization of 10-Hydroxygeraniol Dehydrogenase from *Catharanthus roseus* Reveals Cascaded Enzymatic Activity in Iridoid Biosynthesis. *Sci. Rep.* **5**, 1–6 (2015).

2. Baker, P. J. *et al.* Active site dynamics in the zinc-dependent medium chain alcohol dehydrogenase superfamily. *Proc. Natl. Acad. Sci.* (2009). doi:10.1073/pnas.0807529106
3. Ikeda, H. *et al.* Acyclic monoterpene primary alcohol: NADP + oxidoreductase cells : The key enzyme in biosynthesis of monoterpene alcohols. *J. Biochem.* **109**, 341–347 (1991).
4. Tan, C. S. *et al.* Structural and kinetic studies of a novel nerol dehydrogenase from *Persicaria minor*, a nerol-specific enzyme for citral biosynthesis. *Plant Physiol. Biochem.* (2018). doi:10.1016/j.plaphy.2017.12.033
5. Suites, S. Glide 5.5 User Manual. *Schrödinger Press* (2009).
6. Friesner RA, Banks JL, M. R. *et al et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* (2004). doi:10.1021/jm0306430
7. Kabsch, W. *XDS. Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010). doi:10.1107/S0907444909047337
8. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D. Biol. Crystallogr.* (2006). doi:10.1107/S0907444905036693
9. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
10. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).
11. Krissinel, E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res.* (2015). doi:10.1093/nar/gkv314
12. Stavrinides, A. *et al.* Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity. *Nat. Commun.* (2016). doi:10.1038/ncomms12116
13. Bomati, E. K. Structural and Kinetic Basis for Substrate Selectivity in *Populus tremuloides* Sinapyl Alcohol Dehydrogenase. *PLANT CELL ONLINE* (2005). doi:10.1105/tpc.104.029983

Chapter 5

Structural studies of WW domain containing oxidoreductase (WWOX)

5.1 Background:

SDRs are ubiquitous and present in all three domains of life. Although they have low sequence identity within the domains of life, SDR domain structures display a highly similar folding pattern¹. Similar to plant SDRs animal SDRs regulate complex metabolic processes. The complexity of these enzymes gets compounded when SDR domain is part of other structural domains like WW domain. For example WW domain containing oxidoreductases (WWOX), a recently identified member of NADP⁺/NADPH-dependent oxidoreductases, plays an important role in various biological processes like cell migration, cell differentiation, apoptosis and steroid biogenesis². In addition to the SDR domain, WWOX contains two WW domains (Domain with two conserved Trp residues), which are shown to be involved in regulating diverse cellular signaling events. WW domains are found in various cytoplasmic signaling and structural proteins. Apart from WW domains, these proteins consist of a variety of distinct interaction modules including phosphorylated tyrosine binding domains (PTB) and FF domains^{3,4}, as well as protein localization domains like C2 domain and pleckstrin homology (PH) domains⁵. Additionally, these proteins also consist of a variety of catalytic domains including SDR (WWOX)⁶, E3 protein-ubiquitin ligase (NEDD4 family proteins)⁷, rotomerase/peptidyl prolyl isomerase (PIN1)⁸ and Rho GTPase-activating protein (CAMGAP1)⁹ and thus WW domain-containing proteins are shown to be involved in myriad cellular processes, including protein trafficking and stability, RNA transcription processing, receptor signaling, and cytoskeleton regulation *etc*¹⁰. In recent times, the WW domain containing proteins have grabbed attention because of their role in several human diseases, including muscular dystrophy, Alzheimer's, Huntington's diseases¹¹ and various cancers. Based on the ligand preferences, WW domains are categorized into five different classes^{2,12}. Class I WW domains are the largest group of WW domain-containing proteins including ITCH, WWOX and YAP^{7,13,14}. The WW domains of these proteins recognize an (L/P)PP(Y/Yp) motif (Yp indicates phosphorylated Tyr)^{15,16}. The phosphorylated Tyr known to

affect their binding affinities towards the binding partners of the protein. For example, Interaction of β -dystroglycan with dystrophin WW domain-containing protein gets disrupted by the Tyr phosphorylation in its PPxY motif¹⁷. The class II WW domains present in formin-binding proteins is shown to specifically recognize PPLPP motif¹⁸. On the contrary, class III WW domains are shown to usually bind PR rich motifs. The class IV WW domains, such as the PIN1 WW domain bind specifically to Pro residues preceded by phosphorylated serine or threonine^{19,20}. Finally, class V includes two repeated N-terminal WW domains of yeast PRP40 protein, binds to polyproline motif. Interestingly, WW domains of PRP40 can also binds to (L/P)PP(Y/Yp) and PPLPP motif motifs²¹. Some proteins possess tandem repeats of WW domains. However the exact function of these repeat domains is not clear.

WW domains with substitution of Tyr at the position of the second conserved tryptophan are classified as atypical WW domains. For example second WW domains (WW2) of the human MAGI1 (membrane associated guanylate kinase), eukaryotic Salvador (SAV1)²² and the human WWOX (WW domain-containing oxidoreductase). SAV1 form a β -clam-like homodimer with two monomers of WW2 domains. Unlike SAV1, the WW2 of MAGI1 exist in the monomeric state due to the presence of an Asp (corresponding to Ser in SAV1) substitution in MAGI1²². Interestingly, replacement of Asp with a Ser in MAGI1 resulted in the dimerization of the protein. Similar to MAGI1 WW2, human WWOX WW2 also suggested to be present in monomeric form (PDB code 1WMV) as this protein also contains an Asp at the position of Ser of SAV1 WW2. These atypical WW domains possesses similar fold to that of the typical first WW (WW1) domain.

5.1.1 WW domain structure:

The WW domain is one of the smallest naturally occurred domain composed of about 40 amino acids. Usually it is called by the name WW domain due to the presence of two conserved Trp separated by 20-23 amino acids. WW domain folds into three antiparallel beta-sheet motif²³. The conserved N terminal Trp interacts with Pro residue of conserved LP motif and stabilizes the N-terminal hook like structure. The C-terminal Trp on the third β -strand and a conserved Tyr or Phe of the second β -strand forms a hydrophobic surface and facilitates the binding of proline rich ligands²⁴. In case of PPxY motif, ligand binding is stabilized by the van der Waals interactions between C-terminal Trp and the two conserved prolines of PPxY motif.

On other hand conserved His of WW domain coordinated with Tyr of PPxY motif that fits into an extended hydrophobic pocket²⁵.

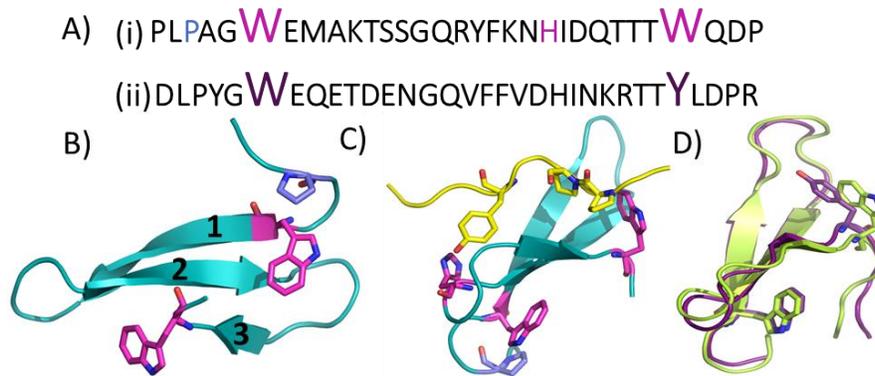


Fig 5.1. A) Consensus sequence of WW domains (i) sequence of YAP65 WW domain, highly conserved residues shown in magenta (ii) Sequence of second WW domain of WWOX (Atypical type), conserved residues shown in purple. B) Structure of WW domain of YAP65 with three anti parallel beta sheaths (PDB code : 1JM0), conserved Trp shown in magenta sticks, Pro shown in blue. C) Complex structure of WW domain (Cyan) and PPxY motif (yellow). Conserved residues in WW domain shown in magenta sticks and PP and Y of PPxY motif shown in yellow sticks. D) Superimposition of atypical WWOX WW2 domain (Purple, pdb code 1WMV) with the typical WW1 domain of Pin1 (Green, PDB code 1F8A) with an RMSD of 0.8Å. Conserved Trp and the atypical Tyr are shown in sticks.

Recent studies showed that alterations at genomic level or expression levels of WWOX implicated in several cancers²⁶. Cellular localization and presence of SDR domain in WWOX suggest the protein to be involved in steroid metabolism²⁷. However, the cognate substrate enzymatic mechanism and substrate specificity determinants of WWOX still remain elusive. The current chapter expounds on the putative substrates, substrate specificity and kinetics of this enzyme. Additionally, efforts to determine the structure of WWOX using X-ray crystallography are also discussed.

5.2 Methods:

5.2.1 Cloning, protein expression and purification:

To obtain high yields of homogenous recombinant WWOX, different constructs of the gene with varying polypeptide length (1-414, 51-414, 110-414) were cloned in several expression vectors (Table 6.1). Optimum expression was achieved in *E.coli* strain B834 (rare). The cells were grown at 37°C to an OD₆₀₀ of 0.5 and further induced with 0.5 mM IPTG and incubated for 14 hrs at 18°C. The culture was harvested by centrifugation. The cells were lysed in chilled lysis buffer containing buffer A (50 mM Tris pH-8.0, 500 mM NaCl, 2 mM EDTA,

2 mM β -mercaptoethanol, and 5% glycerol), one tablet of cOmplete EDTA free protease inhibitor and a pinch of lysozyme, employing sonication. The lysate was filtered and loaded on to the GST affinity column (GE Healthcare Biosciences, USA) for binding. The column was washed with buffer A and the protein was eluted with 20 mM glutathione reduced (Sigma) prepared in the same buffer in case of GST tagged protein and 200 mM imidazole for 6xHis-tagged protein. The elution fractions were dialyzed along with the PreScission protease (purified in our lab) overnight at 10°C. Glutathione, GST tag and PreScission protease were removed by the desalting followed by second GST affinity column. To improve purity of the protein, size exclusion chromatography was performed using Hiprep26/60 sephacryl S-300 (GE Healthcare Biosciences, USA) column in 10 mM Tris pH-8.0, 150 mM NaCl, 5% glycerol, 2 mM DTT. The peak fractions under specific molecular weight were collected and concentrated to perform crystallization experiments.

5.2.2 Thermal shift assay:

In order to check the protein stability, we perform protein thermal shift assay. The reaction mixture consist of 5 μ M of protein, 8x PTS dye, 5 μ l manufactures assay buffer. The final reaction volume of 20 μ l made up with buffer in which protein eluted (10 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol, 2 mM DTT). RT-PCR was used with the following program for melt curve analysis: STEP 1: 25 °C for 2 min with 100% ramp rate followed by 99°C with ramp rate of 1% in continuous mode. The obtained melt curve data was analyzed using Protein Thermal Shift Software (Thermo Fisher).

5.2.3 SEC-MALS:

In order to measure the accurate molecular weight and oligomeric state of the purified WWOX SEC-MALS analysis was performed. For SEC we used Superdex 200 10/300 GL column. (GE Healthcare) SEC column connected to an Agilent HPLC system coupled with the 18-angle light scattering detector (Wyatt Dawn HELIOS II) and a refractive index detector (Wyatt Optilab TrEX). Molecular weight was calculated using ASTRA software (WyattTechnologies).

5.2.4 Site-directed mutagenesis:

Selective mutants of $\Delta 110$ -WVOX were constructed by PCR using the QuikChange Lightning site-directed mutagenesis kit (Agilent) with the appropriate primer pairs. All primers were purchased in a reverse phase-HPLC-purified quality from Eurofins Bangalore. The pET22b expression plasmid containing $\Delta 110$ -WVOX was used as a template. PCR was performed in a program as follows: 30 sec at 95°C (denaturation), 18 cycles of 10 sec at 98°C (denaturation), 30 sec at 55°C (annealing), and 5 min at 68°C (extension). Mutated genes were sequenced, and the proteins were produced in *E.coli* (strain BL 21*) as described for the wild type protein.

5.2.4 Activity assay:

GC-MS based assay has been performed to analyze enzyme activity. The reaction mixture contained 100-200 μg of $\Delta 51$ -WVOX/ $\Delta 110$ -WVOX, 200 μM 10-estradiol/progesterone/pregnane-3, 20-dione and 200 μM NADPH/NADP⁺ in 500 μL of buffer containing 10 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol. Reaction mixtures were incubated at 25-30°C and was terminated after 2 hrs by adding 1 ml of TBME. Compounds were extracted in TBME, 1-5 μl of extracted compounds were injected into GC-MS for analysis using following method for progesterone: equilibration time for 2 min followed by 70°C for 0 min then 10°C/min to 230°C for 0 min then 2°C/min to 250°C for 0 min then 10°C/min to 280°C for 6 min and the overall run time is 39.5 min. The separation was performed on Rtx-5ms column (30 m x 0.25 mm x 0.25 μm) using helium as carrier gas at 1 ml min⁻¹ and with an injector temperature of 220°C.

5.2.5 Spectrophotometric assay:

The enzyme specific activity studies of wild type and mutant proteins were performed in presence of cofactor NADP⁺ and for substrates estradiol. For the determination of Michaelis-Menten parameters with estradiol, kinetics of NADP⁺ consumption was determined spectrophotometrically at 340 nm in cuvettes with 1 cm path length. Reactions contained 200 μM NADPH, 10 μM to 100 μM of estradiol, 2.5 μM $\Delta 51$ -WVOX/ $\Delta 110$ -WVOX enzyme in a total volume of 1 ml buffer containing 10 mM Tris-HCl pH 7.5, 150 mM NaCl, 5% glycerol. The reaction was equilibrated at room temperature and started by addition of enzyme. To obtain K_{cat} and K_m value of the reaction plots of initial velocities versus substrate concentration

were non-linearly fit to the Michaelis-Menten equation using GraphPad Prism5. Data were collected for 5 min at 1 sec intervals. All the assays were performed in triplicates.

5.2.6 Homology modelling:

The high throughput homology modelling was done using modeler 9.19²⁸. Gene sequences of WWOX obtained from Uniprot database. 3RD5, 3RIH, 4JRO, 1WMV structures were used as the model templates. A set of python script files executed for the sequence alignment and modelling. For each homolog five models were generated. The best model out of five was selected based on the discrete optimized protein energy (DOEP) score which is generated from the evaluation.

5.3 Results:

5.3.1 Cloning, expression and purification of WWOX:

Overexpression of eukaryotic gene in heterologous expression system like *E. coli* always pose challenges like levels of expression, protein stability and solubility, protein folding and its post-translational modifications *etc.* Although people have overcome these challenges and achieved good levels of soluble protein, in few cases it fails. By considering all these challenges, WWOX gene was synthesized from GeneArt, whose codon was optimized for expression in *E. coli*. Initially, gene was cloned in pGEX-6P-1 bacterial expression vector (Fig 5.2A). It contains N-terminal GST-tag followed by PreScission protease cleavage site. The protein was expressed in B834 rare strain. The protein was found to be soluble with GST-tag, as it appeared in supernatant. However, it was noticed that post PreScission treatment, the protein eluted in void fraction on gel filtration (Fig 5.2B), indicating protein to be aggregated or misfolded. This prompted us to perform protein disorder prediction. This analysis showed two major disordered regions between 45-54 amino acids (linker between WW1 and WW2) and 99-107 amino acids (linker between WW2 and SDR domains) (Fig 5.2C). Based on these observations we cloned two shorter constructs; Δ 51-WWOX (truncated by 50 amino acids from N-terminus and contains one WW domain followed by SDR domain) and Δ 110-WWOX (truncated by 110 amino acids from N-terminus and contains only SDR domain) in pGEX-6P-1. We expressed these truncated versions in bacterial system and purified. However, on gel filtration both the proteins found to be aggregated as they were eluted in void fractions.

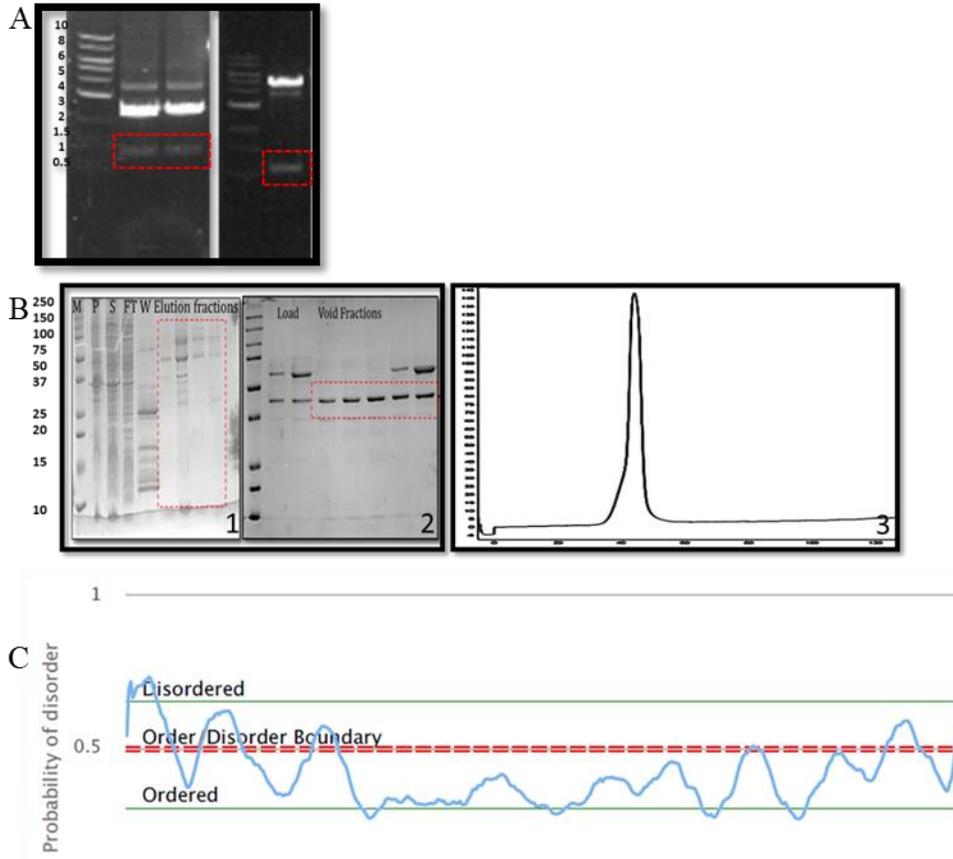


Fig 5.2 A) Agarose gel image of WWOX cloning, WWOX gene insert showed in red dotted box. B) SDS-PAGE image of WWOX-pGEX6P1 purification (1) affinity purification SDS-PAGE (2) Gel filtration SDS-PAGE (3) Gel filtration profile. C) RONN disorder prediction, disordered places displayed in black dotted box.

Elution of protein in void volume also suggests that the protein could be possessing higher molecular weight due to oligomerization, exceeding the capacity of the gel filtration column. Therefore, in order to confirm whether the protein is forming aggregates or higher oligomers we performed thermal shift assay, by measuring the temperature (T_m) at which a protein transits from a well-defined structure to a disorder conformation. In this analysis we used Elmo1 protein as positive control, and eluted fraction of full-length WWOX as test. However, no fluorescence was observed in case of WWOX, indicating that protein to be already aggregated (Fig 5.3).

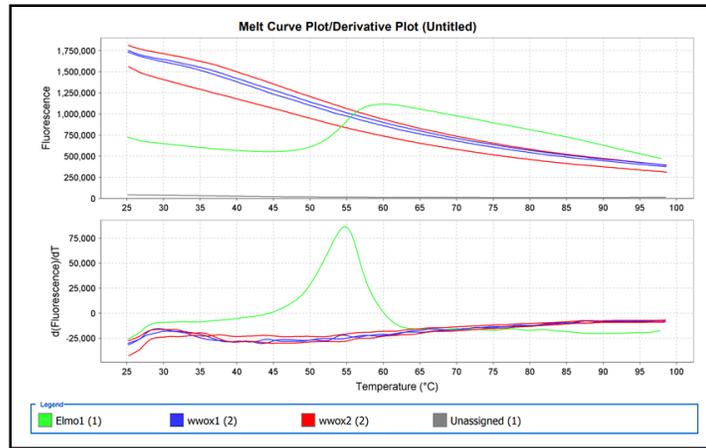


Fig 5.3. Thermal shift assay melt curve plot, red and blue lines represents WWOX and green line represents ELMO1(control).

Subsequently, to circumvent the aggregation problem we cloned full-length and truncated genes of WWOX into several other bacterial expression vectors as well as insect cell expression vectors. Protein obtained from other bacterial expression vectors like pET33b⁺ also exhibited aggregation. List of clones, their expression and/or their solubility are summarized in Table 5.1. Interestingly, protein cloned in pET22b vector with uncleavable C-terminus His-tag and expressed in B834 rare *E. coli* strain eluted at expected retention volume during gel filtration, indicating protein to be properly folded (Fig 5.4A). Similarly, other truncated versions of WWOX, cloned in the same expression vector followed similar trend. The protein of interest was further confirmed with western blotting (Fig 5.4A). Further, we performed MALS for the determination of protein molecular weight and oligomeric state. The molecular weight was ~100 kDa, which indicates that the protein exists as tetramer in solution (Fig 5.4B).

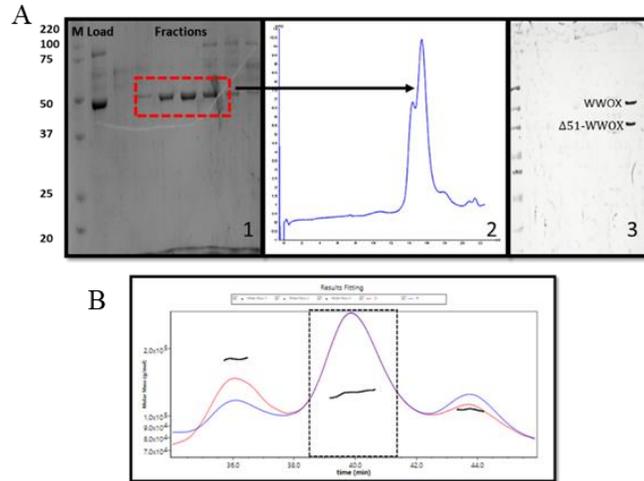


Fig 5.4. A) (1) SDS-PAGE of purification of WWOX-pET22b protein. (2) Gel filtration profile of WWOX-pET22b protein 3) Western-blotting of WWOX-pET22b B) MALS results fitting profile.

Table 5.1 summarized list of WWOX clones and their fate after purification.

GENE	Clone	Vector	Expression System	Remarks
<i>WWOX</i>	WWOX_FL	p3E	Bacteria	Aggregates
		pET33b ⁺	Bacteria	Aggregates
		pET22b	Bacteria	Working
		pET28A	Bacteria	Not tested
		pOPINSS	Bacteria	Not tested
		pOPINSSDS	Bacteria/Insect	Not tested
		pFBD	Insect	Not tested
	pHIS	Bacteria	No expression	
	Δ51-WWOX	p3E	Bacteria	Aggregates
		pET22b	Bacteria	Working
Δ110-WWOX	p3E	Bacteria	Aggregates	
	pET22b	Bacteria	Working	

5.3.2 Determining WWOX targets:

Till date, ample of literature has thrown light on WWOX interactome and depicted its significant role in several cancers especially, in breast and ovarian cancers. However, there are no direct evidences of its enzymatic role and cognate substrates. Base on sequence analysis, previous studies showed that WWOX could possess SDR domain similarity to that of retinol dehydrogenases. The presence of various conserved motifs in virtually identical positions

suggested WWOX could be a bona fide member of SDR. The consensus motifs, including GxxxGxG, involved in stabilizing the (NADP⁺)/NADPH cofactor, YxxxK the catalytic motif critical for the enzymatic activity of SDR domains further suggest presence of SDR domain in WWOX with reductase/dehydrogenase activity²⁹. The modeled structure of SDR domain of WWOX is shown in Fig 5.5A. Moreover, the earlier studies showed that the expression levels of WWOX were significantly high in testis, prostate, and ovary tissues³⁰. Based on this information, we hypothesized that it may be involved in steroid metabolism. We performed GCMS based assay to determine the enzyme activity with three different steroid molecules including pregnane-3, 20-dione, progesterone and estradiol. We incubated these steroid molecules with enzyme (Δ 51-WWOX) in presence of NADPH and NADP⁺ as individual reactions in assay buffer consists of 10 mM Tris-HCl pH 7.5, 150 mM NaCl and 5% glycerol. Surprisingly, enzyme showed activity towards all the three substrates (Fig 5.5). In order to verify if there is any role of WW domain in regulation of reductase function, we performed similar enzyme activity assay with only SDR domain (Δ 110-WWOX). The reductase activity of the SDR domain was similar to that of the untruncated protein. The enzyme kinetic studies also showed that there is no significant change in the enzymatic activity of Δ 51-WWOX compared to Δ 110-WWOX (Fig 5.5 B, Table 5.2). This indicates that WW domain might not have any effect on reductase activity, in isolation. It is interesting to note that amongst the three tested substrates, WWOX showed significant activity towards estradiol. Therefore, we surmise estradiol to be the major substrate of WWOX, which correlates well with the tissue specific expression of the protein. For subsequent activity studies we used estradiol as substrate.

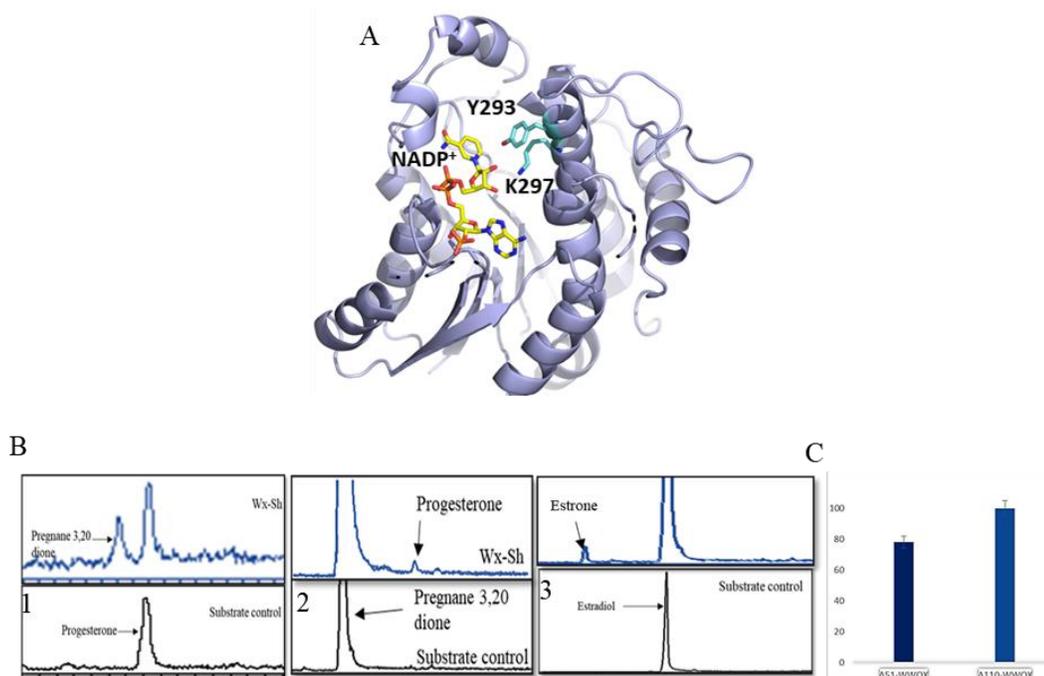


Fig 5.5. A) Modelled structure of WWOX SDR domain (Blue) NADP⁺ showed in yellow sticks, conserved Y293 and K297 shown in cyan sticks. B) Product profiles of GC-MS analysis with substrates (1) progesterone, (2) Pregnane-3, 20-dione (3) Estradiol C) Relative activities of Δ51-WWOX and Δ110-WWOX towards Estradiol.

Table 5.2 Kinetic parameters of NADP⁺ consumption with estradiol

Protein	Substrate (Estradiol) K_m	Substrate K_{cat}	K_{cat}/K_m
Δ51-WWOX	64.55	0.082	0.0013
Δ110-WWOX	59.09	0.11	0.0018

5.3.3 Site-directed mutagenesis studies:

Based on the sequence analysis and enzymatic activity of the WWOX, like in other SDR family, we presumed that conserved catalytic motif YxxxK could play important role in catalysis. Usually in SDR family proteins, Tyr of the catalytic motif mediate the electron transfer between the cofactor and substrate whereas Lys residue reprotoxate the Tyr. Thus to delineate the role of Y293 and K297 of YxxxK motif of WWOX we cloned two WWOX SDR mutants, carrying Y293A and K297A mutations independently. In line with the prediction, none of these mutants showed activity towards estradiol (Fig 5.6). This indicated that Y293 and K297 could play conserved role in the catalytic activity of WWOX.

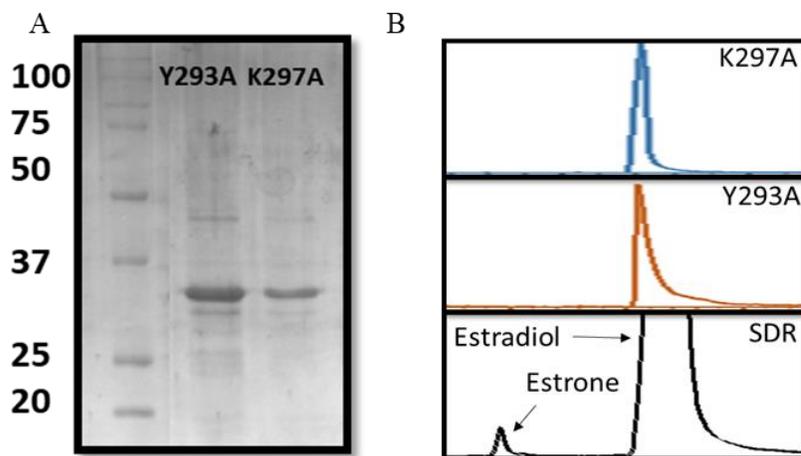


Fig. 5.6 A) SDS PAGE of mutants Y293A and K297A. B) Product profiles of GC-MS analysis with SDR and mutants.

5.3.4 Crystallization of WWOX:

For further understanding of the mechanism of enzymatic activity, we subjected the full length and $\Delta 51$ -WWOX protein, along with 5 molar excess of NADPH, to crystallization trails using commercially available screens. In one of the buffer condition 0.2 M NaCl, 0.1 M Na/K phosphate pH 6.5, 35% PEG 1000 we obtained tiny needle-like crystals. These crystals were very small in size and they were not optimum for X-ray diffraction studies. Therefore, we have optimized the crystallization buffer conditions by changing buffer strength, buffer types, salt types and different molecular weight of PEG. However there was not much improvement in the quality of the crystals. Further we tried buffer optimization using Dragonfly (TTP labtech). Interestingly, we observed elongated needles in one of the conditions containing 0.1 M Tris, 0.4 M NaCl, 33% PEG 1000. However, these crystals were also not up to the mark to collect diffraction data. Therefore, as further efforts to improve the quality of the crystals we performed macroseeding using needle-like crystals as seeds. This resulted in improved quality crystals, in terms of their shape and size. These crystals were used for X-ray diffraction data collection (Fig 5.7).

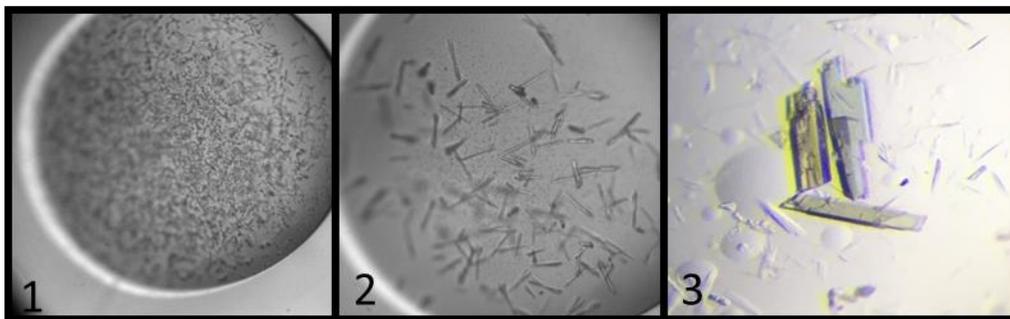


Fig 5.7. Crystals of WWOX (1) microcrystals (2) Needle-like crystals (3) crystals obtained after macroseeding.

5.3.5 Data collection and structure determination:

With improved crystals X-ray diffraction data was collected at BM14 beamline of Diamond synchrotron. The crystals diffracted up to 2.8 Å. The data was indexed using XDS and then scaled using SCALA (Table 5.3). The protein crystallized in P1 space group with unit cell dimensions (75.14, 90.9, 117.20, 69.84°, 73.68°, 89.63°) From Mathew's coefficient suggest presence of six copies of $\Delta 51$ -WWOX in the asymmetric unit (asu) with solvent fraction of 56.8%. This observation is in line with our MALS data, hinting presence of three dimers in the asu. For few data sets self-rotation function suggested presence of pseudo translation (Fig 5.8). However the data set used for structure solution was with slightly different unit cell parameters and did not show presence of pseudo translation. For structure determination using molecular replacement homologous structure with 30% sequence homology was needed. Since for the present case there is no available structure with sequence homology better than 30%, we used structure of putative uncharacterized SDR from *Mycobacterium paratuberculosis* (PDB Code: 3RD5) as search model. The search model (3RD5) exhibits sequence homology of ~25% with $\Delta 51$ -WWOX. However, several attempts with modified search parameters and altered models did not yield any success (Table 5.4). Hence, to obtain phase information with experimental phasing techniques, we soaked the crystals in heavy atom (Au, Hg and Pb) containing cryo buffer. Amongst the three heavy atoms, in terms of their diffractibility, crystals soaked with $K[Au(CN)_2]$ and $HgCl_2$ were found to be promising. For these crystals we collected data at the wavelength corresponding to the absorption edge of the respective heavy atom. Data processing showed the anomalous signal in the diffraction data to be poor, which failed to provide phase information. To overcome this,

we labelled protein with selenomethionine by replacing methionine residues in proteins with selenomethionine through protein overexpression in bacteria grown in selenomethionine containing medium³¹. The labeled protein was purified and crystallized by following the protocols similar to that of the unlabeled protein (Fig 5.9A). Thus, we have successfully crystallized selenomethionine labelled WWOX for structure determination (Fig 5.9B). Due to lower intrinsic (crystal) symmetry and higher radiation sensitivity of the crystals obtaining “phasable” data still remains challenging, which would be taken up by the group.

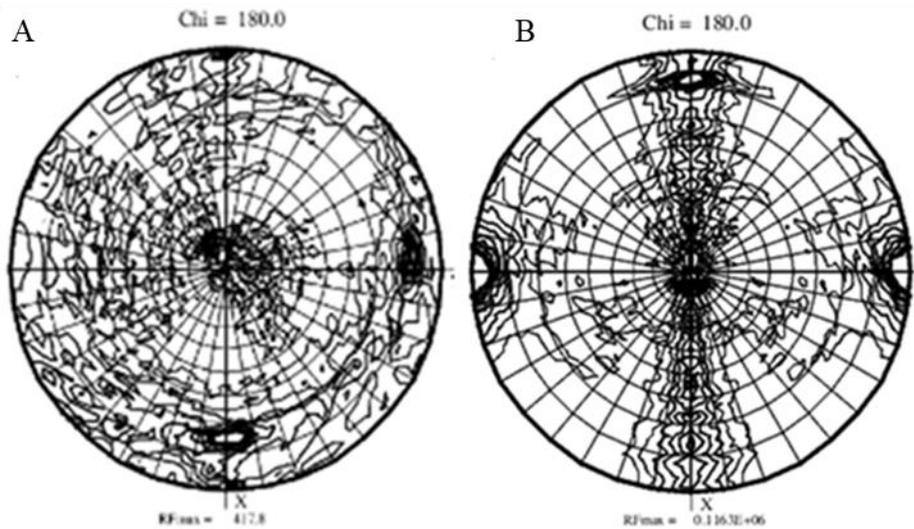


Fig 5.8. Self rotation function of WWOX data sets. A) Peaks in the sphere corresponds to the presence of three dimers each at chi 180° in asymmetric unit. Peak close the origin indicates pseudo translation. B) Peaks corresponds to the presence of three dimers each at chi 180° without pseudo translation.

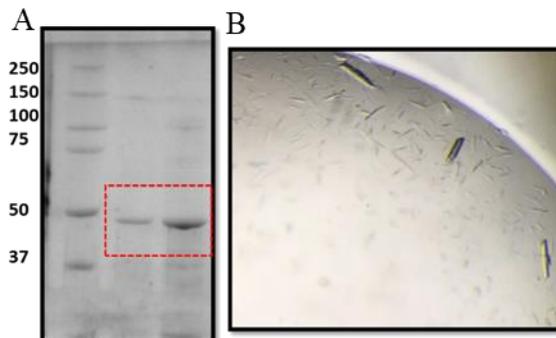


Fig 5.9. A) SDS PAGE of selenomethionine containing WWOX. B) Crystals of selenomethionine containing crystals.

Table 5.3 Data collection statistics of WWOX diffraction data

Data collection statistics	
Space group	P1
a (Å)	75.14
b (Å)	90.9
c (Å)	117.20
α(°)	69.84
β(°)	73.68
γ(°)	89.63
Resolution limits (Å)	29.61-2.8
R_{merge}	0.17 (0.81)
I/s (I)	11.2 (2.5)
Unique reflections	71574 (4569)
Completeness (%)	99 (98)
Multiplicity	11.1 (11.3)

Table 5.4 Summarized molecular replacement runs with WWOX X-ray diffraction data

Search model	Parameters			
	No of copies in search	Rotation function search target	Translation function search target	Remarks
3RD5,	1,2,3,4,5,6	Fast, Brute	Fast	No significant solution
Truncated 3RD5	1,2,3,4,5,6	Fast, Brute	Fast	No significant solution
3RIH	1,2,3,4,5,6	Fast	Fast	No significant solution
4JRO	1,2,3,4,5,6	Fast	Fast	No significant solution

Modelled WWOX	1,2,3,4,5,6	Fast, Brute	Fast	No significant solution
1WMV	1,2,3,4,5,6	Fast	Fast	No significant solution
Multiple ensembles (3RD5, 3RIH, 1WMV, 4JRO)	6	Fast	Fast	No significant solution

5.4 Discussion:

WWOX is the only member of SDR family containing two WW domains, known till date. The SDR domain is responsible for its dehydrogenase/reductase activity whereas WW domain is shown to be involved in protein-protein interactions, exerting a plethora of cellular signaling events. Since the past decades efforts are being made to understand the patho-physiological roles of WWOX. However, there is lacunae in our undersnaing of specific and functions of this protein in cellular metabolism and tumour suppression. The lack of structural information on this protein is the major limitation in delineating its biochemical role in patho-physicilogicall conditions. Thus elucidating the structure of WWOX and its cognate substrates would provide insights on the functioning of this protein and also facilitate designing inhibitors that could potentially act as anti-metastatic drugs.

In this persuit, we have produced several constructuts of WWOX, compatible for hetrologous over-expression in bacterial and insect cell system. Initially we were unable to obtain soluble protein. Howvere, from extensive screening we obtained soluble, stable protein from the construct cloned in pET22b. We examined the biological assembly of WWOX through MALS and shown it to be dimer. Based on the reported cell biology data we hypothesize it to be involved in steroid metabolism. Thus based on this observation we tested the enzymatic activity of WWOX towards progesterone, pregnane-3, 20-dione, and estradiol, in presence of NADP/NADPH. Although, it showed activity towards all the three substrates but the activity towards estradiol was highest. Further, from site-directed mutagenesis studies

we showed that residues Y293 and K297 could play important role in catalysis. For further understanding of the mechanism of enzymatic activity, we successfully crystallized WWOX using macroseeding method. The crystals with spacegroup P1, diffracted to 2.8Å resolution. However, the diffraction data could not be phased with molecular replacement. Hence we went ahead to obtain phase information using experimental phasing techniques. To achieve this we have labelled protein with selenium heavy atom by replace methionine residues in proteins with selenomethionine. We have successfully crystallized selenomethionine labelled WWOX which would pave ways for its structure elucidation.

5.5 Bibliography:

1. Kavanagh, K., Jörnvall, H., Persson, B. & Oppermann, U. The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *FASEB J.* (2008).
2. Del Mare, S., Salah, Z. & Aqeilan, R. I. WWOX: Its genomics, partners, and functions. *Journal of Cellular Biochemistry* (2009). doi:10.1002/jcb.22298
3. Yan, K. S., Kuti, M. & Zhou, M. M. PTB or not PTB - That is the question. *FEBS Letters* (2002). doi:10.1016/S0014-5793(01)03305-1
4. Smith, M. J., Kulkarni, S. & Pawson, T. FF Domains of CA150 Bind Transcription and Splicing Factors through Multiple Weak Interactions. *Mol. Cell. Biol.* (2004). doi:10.1128/mcb.24.21.9274-9285.2004
5. Clark, C. E. J., Nourse, C. C. & Cooper, H. M. The tangled web of non-canonical Wnt signalling in neural migration. *Neurosignals*. **20**, 202–20 (2012).
6. Jörnvall, H. *et al.* Short-Chain Dehydrogenases/Reductases (SDR). *Biochemistry* (1995). doi:10.1021/bi00018a001
7. Ingham, R. J., Gish, G. & Pawson, T. The Nedd4 family of E3 ubiquitin ligases: Functional diversity within a common modular architecture. *Oncogene* (2004). doi:10.1038/sj.onc.1207436
8. Göthel, S. F. & Marahiel, M. A. Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts. *Cellular and Molecular Life Sciences* (1999). doi:10.1007/s000180050299
9. Moon, S. Y. & Zheng, Y. Rho GTPase-activating proteins in cell regulation. *Trends Cell Biol.* (2003). doi:10.1016/S0962-8924(02)00004-1
10. Ingham, R. J. *et al.* WW domains provide a platform for the assembly of multiprotein networks. *Mol. Cell. Biol.* (2005). doi:10.1128/MCB.25.16.7092-7106.2005
11. Sudol, M., Sliwa, K. & Russo, T. Functions of WW domains in the nucleus. *FEBS*

- Letters* (2001). doi:10.1016/S0014-5793(01)02122-6
12. Salah, Z., Aqeilan, R. & Huebner, K. WWOX gene and gene product: tumor suppression through specific protein interactions. *Future oncology (London, England)* (2010). doi:10.2217/fon.09.152
 13. Sudol, M. Structure and function of the WW domain. *Progress in Biophysics and Molecular Biology* (1996). doi:10.1016/S0079-6107(96)00008-9
 14. Chen, H. I. *et al.* Characterization of the WW domain of human Yes-associated protein and its polyproline-containing ligands. *J. Biol. Chem.* (1997). doi:10.1074/jbc.272.27.17070
 15. Linn, H. *et al.* Using molecular repertoires to identify high-affinity peptide ligands of the WW domain of human and mouse YAP. *Biol. Chem.* (1997). doi:10.1515/bchm.1997.378.6.531
 16. Aasland, R. *et al.* Normalization of nomenclature for peptide motifs as ligands of modular protein domains. *FEBS Letters* (2002). doi:10.1016/S0014-5793(01)03295-1
 17. Ilsley, J. L., Sudol, M. & Winder, S. J. The interaction of dystrophin with β -dystroglycan is regulated by tyrosine phosphorylation. *Cell. Signal.* (2001). doi:10.1016/S0898-6568(01)00188-7
 18. Bedford, M. T., Reed, R. & Leder, P. WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: The proline glycine and methionine-rich motif. *Proc. Natl. Acad. Sci.* (2002). doi:10.1073/pnas.95.18.10602
 19. Otte, L. *et al.* WW domain sequence activity relationships identified using ligand recognition propensities of 42 WW domains. *Protein Sci.* (2003). doi:10.1110/ps.0233203
 20. Lu, P. J., Zhou, X. Z., Shen, M. & Lu, K. P. Function of WW domains as phosphoserine- or phosphothreonine-binding modules. *Science* (80-.). (1999). doi:10.1126/science.283.5406.1325
 21. Wiesner, S., Stier, G., Sattler, M. & Macias, M. J. Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40. *J. Mol. Biol.* (2002). doi:10.1016/S0022-2836(02)01145-2
 22. Ohnishi, S. *et al.* Structural basis for controlling the dimerization and stability of the WW domains of an atypical subfamily. *Protein Sci.* (2008). doi:10.1110/ps.035329.108
 23. Ibragimova, G. T. & Wade, R. C. Stability of the β -sheet of the WW domain: A molecular dynamics simulation study. *Biophys. J.* (1999). doi:10.1016/S0006-3495(99)77059-1
 24. Jäger, M., Nguyen, H., Dendle, M., Gruebele, M. & Kelly, J. W. Influence of hPin1

- WW N-terminal domain boundaries on function, protein stability, and folding. *Protein Sci.* (2007). doi:10.1110/ps.072775507
25. Blank, P. N. *et al.* Substitution of Aromatic Residues with Polar Residues in the Active Site Pocket of epi-Isozizaene Synthase Leads to the Generation of New Cyclic Sesquiterpenes. *Biochemistry* **56**, 5798–5811 (2017).
 26. Paige, A. J. W. *et al.* WWOX: A candidate tumor suppressor gene involved in multiple tumor types. *Proc. Natl. Acad. Sci.* (2002). doi:10.1073/pnas.191175898
 27. Ludes-Meyers, J. H. *et al.* WWOX binds the specific proline-rich ligand PPXY: Identification of candidate interacting proteins. *Oncogene* (2004). doi:10.1038/sj.onc.1207680
 28. Andrej Šali. MODELLER A Program for Protein Structure Modeling. *Comp. protein Model. by Satisf. Spat. restraints.* (1993). doi:10.1006/jmbi.1993.1626
 29. Farooq, A. Structural insights into the functional versatility of WW domain-containing oxidoreductase tumor suppressor. *Exp. Biol. Med.* (2015). doi:10.1177/1535370214561586
 30. Bednarek, A. K. *et al.* WWOX, a novel WW domain-containing protein mapping to human chromosome 16q23.3-24.1, a region frequently affected in breast cancer. *Cancer Res.* (2000).
 31. Walden, H. Selenium incorporation using recombinant techniques. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010). doi:10.1107/s0907444909038207

Chapter6

Summary and future scope

Oxidoreductases constitute a large class of enzymes, which catalyze the transfer of an electron from an electron donor to an electron acceptor. In these reactions, they use a variety of electron acceptors and electron-donating substrates, yielding many products of pharmacological and biotechnological interest. Also, oxidoreductases are involved in various energy production processes like photosynthesis, cellular respiration. They also play a role in an array of primary and secondary metabolic pathways and thus these are essential for life. Among all oxidoreductases, ~50% are NADPH-dependent. Industrially these enzymes pose great interest as they catalyze crucial steps in the production of various chemical compounds which are difficult to synthesize under viable conditions. They have been used largely in synthetic biology for the production of various bioactive compounds such as artemisinin, taxol, ajmalicine, vinblastine and vincristine *etc* in large scale where vinblastine and vincristine are widely used anticancer drugs and are majorly synthesized in plants through precursor molecules like iridoids. Furthermore, many of these enzymes are drug targets, as they are implicated in several diseases like cancer, cardiovascular dysfunction etc. This thesis comprises of three atypical NADPH-dependent oxidoreductase systems which possess uncommon mechanisms in addition to their redox mechanism, unlike typical oxidoreductases. Two of them are involved in the regulation of iridoid biosynthesis, a precursor for vincristine and vinblastine (iridoid synthase & 10-hydroxygeraniol dehydrogenase) and third WW domain containing oxidoreductases is a newly identified drug target for breast cancer. We explored structural and functional aspects of these atypical oxidoreductases employing X-ray crystallography and Cryo-EM and other biochemical and biophysical techniques.

Recent studies on oxidoreductases systems in *C. roseus* has provided a novel reducto cyclase IS and also dehydrogenase 10-HGO involved in the biosynthesis of iridoids. IS not only acts as a reductase but also claimed to be involved in cyclization of its substrate 10-oxogeraniol which is the product of its precursor enzyme 10-HGO. These novel features of IS makes a non-canonical reducto cyclase. Despite of having biochemical information on the chemical reaction involve in the IS mechanistic details and substrate specificity was unclear.

By elucidating the crystal structure of IS catalytic mechanism and substrate specificity were delineated. Although having significant structural similarity with P5 β R, IS has subtle differences in exerting reduction catalyzing reduction reaction of 10-oxogeranial. Tyr185 orients the substrate and stabilize the intermediates. Perhaps the wide catalytic pocket of the enzyme could be involved in hetero stereo selectivity of the enzyme. Another surprising fact about IS is substrate specificity compared to its structural homologue P5 β R. The former reduces 10-oxogeranial whereas later specific to progesterone. Comparative structural studies, site-directed mutagenesis and MD simulation presented here showed that two loops L1 and L2 at opening substrate entry channel are involved in generating specificity. The two residues EG (SY in P5 β R) present in L1 modulates the substrate entry channel and orientation of the substrate and hence govern the specificity of enzyme. Another interesting fact about 10-HGO and IS is their concerted action. When these 10-hydroxgeranial is treated with 10-HGO and IS is in a single pot reaction, nepetalactol is a major product on the contrary when these enzyme act in tandem they produce heterogeneous products where nepetalactol is not a major product. We delineated the structural basis of this concerted action using chimera of 10-HGO-IS as a surrogate system. Using protein complementation assay and single particle Cryo-EM we showed that these two enzymes physically interact in a manner to form a substrate or product passing channel. With this interaction, the product of 10-HGO is directly transferred to IS as substrate making it less amiable for stereo heterogeneity thus the ‘hand-shake’ transfer of product and substrate hint at the co-evolution of these two enzymes.

WVOX is another reductase from humans which was investigated in this study. In addition to classical SDR domain, this enzyme possess two tandem WW domain. By virtue of these two signaling domain and its reductase function, WVOX is shown to be involved in myriad signaling elements. Aberration in the function of WVOX has been implicated in numerous cancer. Aligning with our theme of this thesis we probed the function of this SDR domain containing reductase. From biochemical studies, we established that progesterone, pregnane-3, 20-dione, estradiol could be the substrates of this enzyme. In the pursuits of elucidating WVOX molecular mechanism at atomic level, we successfully crystallized WVOX and collected diffraction data however due to lack of phase information structure remains unsolved. To obtain phase information using experimental phasing techniques we

have labelled protein with selenomethionine thus the current work on WWOX will pave ways for elucidating atomic structure and also designing novel drugs targeting this enzyme.

Thus the structural and biochemical investigations conducted here on SDR provided valuable insights on their function and substrate specificity. Given the biotechnological and pharmaceutical importance of these enzymes, the study presented here will advance the technology in developing large scale production of MIAS and /or novel drugs for cancer.

Supplementary information

List of Primers:

Primer Name	Primer Sequence
IS_FWD	GGTGGTCATATGGAAAATGGAGTTTGTAAAAGCTACAAGAG TGT
IS_REV	GGTGGTGCGGCCGCCTAAGGAATAAACCTATAATCCCTCAT CTTAT
P5 β R_S155E&Y156G_FWD	ATGGGTCCGTTTGAAGAGGGTGGCAAATCGAAAGC
P5 β R_S155E&Y156G_REV	GCTTTCGATTTTGCCACCCTCTTCAAACGGACCCAT
IDS_E161S&G162Y_FWD	TATTTTGGTATTTTTGAATCATATTCTAAAGTTGTACCACAT
IDS_E161S&G162Y_REV	ATGTGGTACAACCTTTAGAATATGATTCAAAAATACCAAAT A
IDS_N358A_FWD	TTTTGTTCTGAGGCTTTGATTAGTAGT
IDS_N358A_REV	ACTACTAATCAAAGCCTCAGAACAAAA
IDS_DHFR2FP	GGTGGTCCATGGATGGAAAATGGA GTTTGTAAAAGCTACAAG
IDS_DHFR2RP	GGTGGTGCGGCCGCAGGAATAAACCTATAATCCCTCATCTT A
10HGO_DHFR3FP	GGTGGTCCATGGATGCATCATCATCATCATGCGAAATC ACCGGAAGTCGAG
10HGO_DHFR3RP	GGTGGTGCGGCCGCCAGATTCAGTGTGTTGGCTACATCA
10HGO_IS_FP1	CGGGGAUATGGCGAAATCACCGGAAGTCGAG
10HGO_IS_RP1	ACCACCACCU GCAGATTTTCAGTGTGTT
10HGO_IS_FP2	AGGTGGTGGUGGTGGCGACCA
10HGO_IS_RP2	CCCCGAUCTAAGGAATAAACCTATAATCCCTCATCTTATCAA
10HGO_FP	GGTGGTACCGGTGAAAACCTGTACTTCCAGAGC GGG ATG GCGAAATCACCGGAAGTCGAG
10HGO_RP	GGTGAATTCTTATGCAGATTTTCAGTGTGTTGGCTACATCAA
WX_FP	GGTGGTCATATGGCAGCACTGCGTTATGCAGG
WVOX_FP57	GGTCATATGGGTGATCTGCCGTATGGCTGG
WVOX_FP110	GGTCATATGGGTAGCACCACCGCAATGGAAATT
WVOX_RP	GGTGCGGCCGCTTAACGTTCTTGAATCAGACGTTGCTCA
WX_352 RP	GGTGGTGCGGCCGCTTAAAGCTTCATGCTTTTGGTAAACGG ACGTGC
WX_Y295AFP	CAATGCTGGCAGCCAATCGTAGCAA
WX_Y295ARP	TTTGCTACGATTGGCTGCCAGCATTG
WX_K299AFP	AATCGTAGCGCACTGTGCAATATTCTG
WX_K299ARP	CAGAATATTGCACAGTGCCTACGATT