

STRAWBERRY DIOXYGENASE: STRUCTURAL AND FUNCTIONAL STUDY BY HOMOLOGY MODELING

Ph.D. THESIS

by

ALOK KUMAR



**DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247 667 (INDIA)
DECEMBER, 2013**

STRAWBERRY DIOXYGENASE: STRUCTURAL AND FUNCTIONAL STUDY BY HOMOLOGY MODELING

A THESIS

*Submitted in partial fulfilment of the
Requirements for the award of the degree*

of

DOCTOR OF PHILOSOPHY

in

BIOTECHNOLOGY

by

ALOK KUMAR



**DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247 667 (INDIA)
DECEMBER, 2013**

**©INDIAN INSTITUTE OF TECHNOLOGY ROORKEE, ROORKEE-2013
ALL RIGHTS RESERVED**



INDIAN INSTITUTE OF TECHNOLOGY ROORKEE ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **“STRAWBERRY DIOXYGENASE: STRUCTURAL AND FUNCTIONAL STUDY BY HOMOLOGY MODELING”** in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the Department of Biotechnology of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from January, 2008 to December, 2013 under the supervision of Prof. Dr. Ramasare Prasad, Professor & Head, Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other Institute.

(ALOK KUMAR)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Dated: December, 2013

(Ramasare Prasad)

Supervisor

The Ph.D. Viva-Voce Examination of **Mr. Alok Kumar**, Research Scholar, has been held on..... .

Signature of Supervisor

Signature of Chairman, SRC

Signature of External Examiner

Signature of Head of the Dept./Chairman, ODC

ABSTRACT

Strawberry, in technical terms, is an aggregate accessory fruit and is being used in food and beverages, besides this, strawberry pigment extracts can be used as natural acid/base indicator due to different color of the conjugate acid and bases in the pigment. Strawberries contain Fisetin, an antioxidant which can be studied in relation to Alzheimers' disease and to kidney failure resulting from diabetes. Strawberries are of high nutritional value, and it is also a good source of flavonoids. Strawberry contains chemical compounds i.e. dimeric ellagitannin agrimoniin, which is an isomer of sanguin. Anaphylactoid reactions to the consumption of strawberries have also been reported, which indicate that strawberry contains some allergen also. Strawberry has potential health benefits. Strawberry leaves and fruit extracts are useful in cardiovascular diseases and other disorders. Strawberry consumption decreases high blood pressure and reduces the risk of heart disease. Polyphenols in strawberries help regulate blood sugar response with decreased risk of type 2 diabetes. Strawberry is a good source of antioxidants which can help boost immune system and fight free radicals. The flavonol and other secondary metabolites such as anthocyanidins, quercetin, kempferol, succinate, fumarate etc. have various functions including anti-inflammatory, decrease the effects of Alzheimer's disease, relieve rheumatoid arthritis and the acids might help remove stains from teeth. Strawberry extracts are very low in saturated fats, cholesterol and sodium. Strawberries help prevent macular degeneration, enhance memory functions and prevent types of cancers including esophageal, colon, breast, ovarian, cervical and lung cancer. These health benefits indicate that strawberries could be a potential dietary supplement.

In strawberries, dioxygenases are reported to be essential for oxidative transformation reactions including fruit ripening, flavor biogenesis and plant defense responses against pathogens. A dioxygenase is an enzyme which incorporates both the atoms of molecular oxygen into kinds of substrates, undergoing various types of mechanisms. One of the most important functions of dioxygenases would be cleavage of aromatic rings, which is an important step in the degradation of aromatic compounds. Based on the substrate, dioxygenases could be further classified broadly into carotenoid cleavage dioxygenases (CCDs) and specifically into 9-cis-epoxycarotenoid cleavage dioxygenases (NCEDs), which catalyzes the rate limiting step in ABA biosynthesis; supposed oxidative cleavage of

the carbon-carbon double bond of 9-cis-violaxanthin or 9-cis-neoxanthin. In maize (*Zea mays*), VP14 (Viviparous 14) is a kind of 9-cis-epoxycarotenoid dioxygenase and is one of the first characterized NCED through X-ray crystallography. NCEDs including VP14 are key regulators that determine ABA levels, which in turn control ABA-regulated processes. Homologs of VP14 have been identified in other plant species as well. Similar proteins have also been identified in variety of other plants, prokaryotes and animals. These results along with those studies into the physiological effects of ABA or apocarotenoid biosynthesis, laid the foundation for a more complete understanding of ABA biosynthesis and apocarotenoid biosynthesis or retinoid functions. Yet, although, various mechanisms for the rate limiting step in ABA biosynthesis and the 9-cis-carotenoid double bond cleavage have been proposed but the structure and the natural properties of specific determinants of specificity of dioxygenase pertaining to substrate specificity related amino acid motifs or encoding ring motifs are subject of intensive study.

To characterize the cleavage reaction mechanism, structural and other studies of strawberry dioxygenase would be of value. If the structure of the protein and the coordination of the iron in the active site are determined and its consistency with previously determined dioxygenase structures, also further analysis of the substrate activity relationships and identification of potential inhibitor compounds could provide clues to classify the strawberry dioxygenase and methods to understand how CCDs and NCEDs work. The dioxygenase – substrate structure model can also be used for the development of therapeutics as this will promote strawberry as a useful dietary supplement.

Keeping in view the above facts, the present study aims at understanding the strawberry dioxygenase structure by homology modeling approach, with the following objectives-

- Strawberry dioxygenase structure model analyses
- Strawberry dioxygenase and substrate activity interactions
- Identification of strawberry dioxygenase inhibitors

The Thesis is divided into 4 Chapters. Chapter 1 reviews the articles related to strawberry fruit ripening, flavor biogenesis; oxidative transformation and self hydroxylation reactions, herbicides and herbicide degradation and available dioxygenase structures, such as maize VP14, 4-HPPD, Tryptophan-2,3-dioxygenase, Indoleamine-2,3-dioxygenase and RPE65.

Chapter 2 describes Strawberry dioxygenase sequence and structural analysis. Dioxygenase amino acid sequence was retrieved from the NCBI database and after various analysis i.e. transit peptide, transmembrane fragments, hydropathy, conserved domain and motifs, post translational modification sites, e.g. glycosylation, amidation and phosphorylation etc., the template based structure model was generated by using different servers. The Structure model was further analyzed for quality assessment. The membrane orientation and 2D representation of dioxygenase shows that the N- and C-terminals lie in the intracellular side and dioxygenase structure is tilted by an angle ($46\pm 3^\circ$) from the membrane surface. The secondary structure and topology analysis indicates that the dioxygenase structure contains an α -helical domain and a β -propeller structure that is also consistent with dioxygenase structures of this class. The metal coordination or Fe^{2+} ion centre is consistent with other dioxygenase structures as the Fe^{2+} is harbored on the central axis of the tunnel that runs across the complete structure by four octahedrally bound histidine residues (His225, His273, His339, His529) located on long loops. This type of Fe^{2+} coordination is found only in a few enzymes such as superoxide dismutase, photosystem II, fumerate reductase. All attempts to refine the structure shows that molecular oxygen is the most likely ligand at this position from a crystallographic as well as from a mechanistic point of view. The structure and coordination of strawberry dioxygenase were compared with Human tryptophan-2,3-dioxygenase and Indoleamine-2,3-dioxygenase which indicated that its human counterparts are mainly membrane proteins with several transmembrane segments while dioxygenase is a monotopic membrane protein. Also, consistently the substrates are likely to be trapped within the transmembrane domain completely or minimally occluded from the intracellular side yet solvent exposed from the extracellular side through a pore of diameter 8-40 Å, which is presumably not too wide for substrate to escape. The structure also indicates that α -helical domain may be interacting with the membrane surface, acting as a gate for substrate entry, which is consistent with Maize VP14 and it is assumed that other dioxygenase structures would contain the same conformation and have similar substrate interaction mechanisms. The study also identifies specific amino acid residues based on sequence analysis, which may be involved in self hydroxylation reactions and enantiomeric and stereo selectivity of substrates.

Chapter 3 describes specific substrate binding domains in the dioxygenase structure. These substrate binding domains also provide reasons to assume that dioxygenase plays a very central role in the signaling pathway that connects catabolic processes to the plant defense

triggered by different allergens and also to the activated peripheral immune response during diseased conditions in humans. The study includes dioxygen interaction with dioxygenase which is the most fundamental interaction in these classes of enzymes. Also, some of the important interacting proteins and substrates include Allergen and lipids, Epoxycholesterol, Cell adhesion proteins, Ser/Thr kinases, Actin like ATPase, Dicer like protein, Peptide binding protein, Myosin ATPase and Immunoglobulin protein. These binding domains are imperative that it can be hypothesized that plant hormone interactions with dioxygenase in plants could pave the way to find potential new inhibitors or herbicides and the structure model could provide insights into designing targeted drugs pertaining to the treatment of various diseases and related conditions.

Chapter 4 describes dioxygenase interactions with plant hormone Abscisic acid receptor (PYL9) and other selective herbicide inhibitors including Indoleamine imidazole, 2,4-D, Dicamba etc. Compounds containing amines, imidazole are potential dioxygenase inhibitors such as 2,4-D and Dicamba, which are synthetic versions of plant hormone Auxin. It is quite known for some time that these compounds act on methyl group of substrates. The correct positioning of aromatic ring of Hydroxamic acid or Abamine derivative compounds in the active site serves the basis of enzyme inhibition as the positioning of aromatic ring in the active site competes with the substrate and could be a common mechanism for competitive inhibition of these group of enzymes. Some possible modifications of specific amino acid residues such as tyrosine hydroxylation and phosphorylation, glutamine and asparagine deamination are involved in the His¹-Glu/Asp-His² mechanism, which is widely applicable and it provides reasons to assume that self hydroxylation reactions are not very uncommon uncoupled reaction in this class of enzymes.

Chapter 5 describes overall methods used for analyzing transmembrane segments, transit peptide, folds and motifs in strawberry dioxygenase protein sequence retrieved from database. The structural model was generated and analyzed by using different servers and tools.

Finally, the main conclusions of the present study are summarized in Chapter 6. The structure of strawberry dioxygenase is consistent with other characterized structures of this family of dioxygenases. The structure of strawberry dioxygenase suggests some substrate and protein binding domains based on which potential protein partners and inhibitor

compounds are identified. The study also suggests the health benefits of strawberry dioxygenase by dioxygenase and possible drug substrate interactions. The structural similarity of strawberry dioxygenase and its human counterparts Tryptophan-2,3-dioxygenase and Indolamine-2,3-dioxygenase and Integrin receptors and substrate binding domains within the strawberry dioxygenase structure indicate that some dioxygenase functions may be common in plant and animal or human enzymes. This finding provide a reason to speculate the role played by dioxygenases in peripheral immune responses and it suggests that dioxygenase could be acting as an important node in the signaling pathway triggered by receptor tyrosine kinase or Integrin receptors, which sense mechanical constraints within the cell. It is hypothesized that some specific amino acid residues including arginine and tryptophan could act as dietary amino acids and may be involved in dioxygenase mediated regulation of T_{regs} and dendritic cells. This study on the basis of structure model also suggests a predictive role of dioxygenases under Hypoxia or similar conditions, Angiogenesis, Cell differentiation, Cell proliferation and migration; Tissue morphogenesis, Tumor growth i.e. Solid and circulating tumors and underlying pathways and Poly(ADP-ribose)polymerase mediated DNA damage associated cell cycle control. The study also predicts the role of dioxygenases in neuronal development and in the Axon guidance pathway. Moreover, based on the structural homology and dioxygenase structure several targeted drugs can be selected for analyses and in terms of their modes of action the underlying pathways can be studied in detail. Such targeted therapeutics would benefit patients with different type of diseases, including metabolic symptomatic disorders, neurogenerative diseases and certain forms of cancers.

ACKNOWLEDGEMENTS

This thesis is the result of six years of work whereby I have been accompanied and supported by many people. It is a pleasant aspect that I have now the opportunity to express my gratitude to all those people who contributed directly or indirectly to this endeavor. First and foremost, I would like to acknowledge my revered and esteemed supervisor Dr. Ramasare Prasad for his erudite guidance, outstanding advice, encouraging attitude, and abiding interest in this work. I thank him for his patience, kind nature, for many fruitful discussions and also for rectifying my mistakes. I would also thank Dr. Pravindra Kumar for his kind suggestions in deciding the PhD. topic and encouragement that have helped me all through the time of research. I would also like to thank him for providing invaluable expertise in field of Bioinformatics. I am privileged to have got opportunity to work with them. I feel overwhelmed in thanking to Prof. R Prasad (Present) and Prof. Ritu Barthwal (Former) Head, Department of Biotechnology for providing necessary facilities, support and cooperation in the department. I gratefully acknowledge the help rendered from time to time by the members of my Student Research Committee (SRC) Prof. G. S. Randhawa, Dr. Partha Roy (Present) and Dr. Ashwini Kumar Sharma (Internal member) and Dr. M.M. Mourya (External member, Department of chemistry) for their scholarly suggestions, prudent admonitions and immense interest that have made this task a success. I am courteous to the honorable faculty members, Professor H. S. Dhaliwal, Professor R.P. Singh, Dr. Ranjana Pathania, Dr. Naveen Navani, Dr. Vikas Pruthi, Dr. Partha Roy, Dr. Bijan Choudhury, Dr. Sanjoy Ghosh, Dr. M. S. Nair for their support and encouragement. I would like to express my sincere thanks towards Dr. Pravindra Kumar and Dr. Shailly Tomar, Department of Biotechnology for their unflinching help and support. I am much indebted to both of them for their valuable advice and precious times to read my work. I am obliged to my seniors and labmates Dr. Ashish, Dr. Neelesh Kumar Sharma, Dr. Seema Parveen, Dr. Sreela Dey, Dr. Sunity Singh and Ms. Rajbala for their sincere efforts, amiable attitude and cooperation while working in lab. I would like to thank Ms. Rajbala, Mr. Shailendra, Mr. Satish and Mr. Umesh for assistance, help and support in the successful completion of this thesis. Also to be acknowledged with love and appreciation in the memorable concern, affection and care from my dear friends Dr. P. Selva Kumar, Dr. Preeti, Mr. Shivendra, Dr. Dipak Mr. Paramesh, Ms. Manju, Ms. Sonali, Dr. Satya Tapas, Mr. Pramod, Mr. Aditya, Dr. Nagesh, Dr. Santosh, Dr. Rajnikant, Dr. Mrs. Supriya and many others who made my stay in IITR memorable and wonderful. I never felt home sick only because of their loving and caring attitude. I would like to mention all the technical and office staff of the Biotechnology Department in particular Mr. Ved Pal Singh Saini, Mr. Subhash Jain, Mrs. Shashi Prabha, Mr. Lokesh and Mr. Anil.

My parents deserve special mention for their unending love, inseparable support blessing, educating me and encouragement to pursue my interest. Without them this thesis would have been a distant dream. My deep appreciation goes to my admirable sisters who always encouraged me to move forward and meet the challenges held high in everything I try.

They were the core inspiration to reach great height and comfort when I occasionally falter. I would like to thank everybody who was important to the successful realization of thesis, as well as expressing my apology that I could not mention personally one by one.

I am also grateful beyond measure to University Grants Commission, Govt. of India for providing the financial support to conduct the research work.

(Alok Kumar)

LIST OF PUBLICATIONS

1. Jha, Alok., Prasad, Ramasare. Substrate based inhibitors of Strawberry Dioxygenase: Homology Models. International Journal of Scientific and Research Publications. 2013; 3,11:1-7.
2. Jha, Alok., Prasad, Ramasare. Strawberry Dioxygenase: Vorapaxar A model for antiplatelet therapy. IORS:JPBS. 2013; 8, 3: 24-32.

CONTENTS

	Page No.
CANDIDATE'S DECLARATION	i
ABSTRACT	ii-vi
ACKNOWLEDGEMENT	vii-viii
LIST OF PUBLICATIONS	ix
CONTENTS	x-xvi
CHAPTER 1	
INTRODUCTION	1-34
1.1. Introduction	2-5
1.1.1. Dioxygenase and oxidative transformation reactions	5-6
1.1.2. Oxygen activated by metal ions	6-7
1.1.3. Orbital overlap with a metal ion	7
1.1.4. Single electron transfer	7-8
1.1.5. Reaction with a substrate radical	8
1.1.6. Nucleophilic reactivity	8
1.1.7. Electrophilic reactivity	8-9
1.1.8. O-O bond cleavage via C-C bond migration	9
1.1.9. O-O cleavage via hemolytic cleavage	9
1.2. Strawberry metabolism	10
1.2.1. Recent advances in strawberry metabolomics	10
1.2.2. Strawberry fruit ripening and fruit texture	11-14
1.3. Abscisic acid and herbicide inhibitors	14-24
1.4. Strawberry flavor biogenesis	24-27
1.5. Strawberry health benefits	27-29
1.5.1. Strawberry as a Space food	29-30
1.6. Dioxygenase Structure and Function	30

1.6.1.	Dioxygenase structure model, substrate binding domains and inhibitor designing	30-32
1.7.	Loop modeling	33-34
1.8.	Research orientations	34

CHAPTER 2

SEQUENCE AND STRUCTURE ANALYSIS		35-132
2.1.	Introduction	35-37
2.2.	Strawberry dioxygenase amino acid sequence	37
2.2.1	Sequence analysis	37-38
2.2.2.	Alignments: %age similarity	38-41
2.2.3.	Protein BLAST result using PDB for the same protein (Dioxygenase [<i>Fragaria x ananassa</i>])	41-42
2.2.4.	Transit Peptides	42
2.2.5.	Membrane binding property	42
2.2.5.1.	Conserved domain analysis	42
2.2.5.2.	Description	43-44
2.2.6.	Hydropathy plot	44-47
2.2.7.	β -barrel outer membrane protein analysis	47
2.2.7.1.	Posterior probability plot (sequence)	47-48
2.2.8.	Folds and Motifs	49
2.2.8.1.	NAD-FAD Rossmann fold motif	49
2.2.8.2.	Motif analysis	49-57
2.2.9.	Glycosylation sites	58
2.2.9.1	Net C Glyc Prediction	58
2.2.9.2.	<i>Net N Glyc Prediction</i>	58
2.2.9.3.	Net O Glyc Prediction	59-61
2.2.9.4.	Phosphorylation sites	61-64

2.2.10.	Substrate Binding domains	64-65
2.2.11.	Inference	66
2.2.12.	Instability index	67
2.2.13.	Secondary structure	67-68
2.2.13.1	Model sequence analysis	68
2.2.13.2.	Model sequence	68
2.2.13.3.	Sequence search	69
2.2.13.4.	Hydropathy analysis	69-70
2.2.13.5.	β -barrel outer membrane protein analysis	70
2.2.13.5.1.	Posterior probability plot (model)	70-71
2.2.14.	Orientation of Proteins in Membranes	71-72
2.2.15.	Model Quality	72
2.2.15.1.	Coiled- coil probability	72
2.2.15.2.	Coiled coil probability	73
2.2.16.	Structure model validation and quality assessment	74
2.2.16.1.	Dioxygenase model Residue Error	74-77
2.2.16.2.	Ramachandran plot	78-79
2.2.17.	Fold and Motif analysis	79
2.2.17.1.	Fold analysis	79-80
2.2.17.2.	Motif analysis	80-87
2.2.17.3.	Phosphorylation sites	87-90
2.2.17.4.	Glycosylation sites	90
2.2.17.4.1.	NetCGlyc Prediction	90
2.2.17. 4.2	NetNGlyc Prediction	91
2.2.17.4.3.	NetOGlyc Prediction	91-93
2.2.18.	Reverse template comparison vs. structures in PDB	93
2.2.19.	Existing PDB structures	94

2.2.20.	Model	94
2.2.20.1.	Model secondary structure	94-95
2.2.20.2.	Model secondary structure topology	96
2.2.21.	Dioxygenase protein structure	96-97
2.2.21.1.1.	Dioxygenase membrane orientation	97-98
2.2.21.1.2.	Transmembrane segments	98
2.2.21.1.3.	Orientation of dioxygenase in membranes	99-101
2.2.21.2.	Electrostatic patches	101
2.2.21.3.	Cysteine disulphide bonds	101-102
2.2.21.4.	Structure pore lining residues	102-103
2.2.21.4.1.	Pore shape	103
2.2.21.4.2.	Pore visualization	104
2.2.21.4.2.1.	Protein orientation	104-105
2.2.21.4.3.	Features of the cavity	105
2.2.21.4.3.1.	Protein orientation	105-106
2.2.21.5.	Dioxygenase protein structure	106-107
2.2.21.5.1.	Structure outline	107-109
2.2.21.5.2.	α -Helices	109-110
2.2.21.5.2.1.	Helix-helix interaction	110
2.2.21.5.2.2.	β -sheets	110
2.2.21.5.3.1.	β -strands	111-112
2.2.21.5.3.2.	β -hairpins	113
2.2.21.5.3.3.	β -turn	113-114
2.2.21.5.3.4.	β -bulges	115
2.2.21.5.3.5.	γ -(gamma) turn	115-116
2.2.21.5.4.	Loops and Kinks	116
2.2.21.5.4.1.	Large loop modeling	116-118
2.2.21.5.4.2.	Kinked α -helix	118-119

2.2.21.5.5.	Coordination	119-121
2.2.21.5.6.	Comparison	122-126
2.2.21.5.7.	Substrate binding pockets	127-129
2.2.21.5.8.	Substrate specificity and enantio-selectivity	129-130
2.2.21.5.9.	Self Hydroxylation	131-132

CHAPTER 3

DIOXYGENASE SUBSTRATE BINDING DOMAINS 133-186

3.1.	Introduction	133-134
3.2.	Dioxygenase sequence and substrate binding domains	134-137
3.3.	Dioxygen	137-138
3.3.1.	Dioxygenase and dioxygen binding	139-140
3.3.2.	Dioxygenase and dioxygen interactions	140-141
3.4.	Substrates binding domains	141
3.4.1.	Substrate binding domain 1	141
3.4.1.1.	Allergen and Lipids	141-143
3.4.1.2.	Allergen-lipid binding:Nematode Allergen-Lipid Binding	143-144
3.4.2.	Alpha-epoxycholesterol	145-146
3.4.2.1.	Alpha-epoxycholesterol binding	146
3.4.2..2.	Energy table	146-147
3.4.2.3.	Dioxygenase and Alpha-epoxycholesterol interactions	147-151
3.4.3.	Cell adhesion proteins	151-152
3.4.3.1.	Human Genetic Diseases	152
3.4.3.2.	Tumor adhesion	152
3.4.3.3.	Prokaryotes	152

3.4.3.4.	Viruses	152
3.4.3.5.	Focal adhesion kinase (FAK)	153-154
3.4.3.6..	Focal adhesion kinase (FAK) binding	155-156
3.4.3.6.1	Dioxygenase and Focal Adhesion Kinase (FAK) interactions	156-159
3.4.4.	Substrate binding domain 2	159
3.4.4.1.	Ser/Thr kinase binding	159
3.4.4.1.1.	Ser/Thr kinases	159-160
3.4.4.1.2.	Ser/Thr kinase binding	160-161
3.4.4.1.3.	Dioxygenase and Ser/Thr kinase HipA interactions	161-163
3.4.4.2.	Actin-like protein binding	163
3.4.4.2.1.	Actin like proteins	163-164
3.4.4.2.2.	Filamin-Immunoglobulin repeats N-terminal CFTR binding	164-165
3.4.4.3.3.	Dioxygenase and Filamin Immunoglobulin repeats N-terminal CFTR interactions	165-167
3.4.5.	Substrate binding domain 3	167
3.4.5.1.	Dicer-like protein binding	167
3.4.5.1.1	Dicer like proteins	167-168
3.4.5.1.2.	Dicer-like protein binding	168-169
3.4.5.1.3.	Dioxygenase and Dicer like protein (DCL4) interactions	169-171
3.4.5.2.	Peptide binding protein binding	171
3.4.5.2.1.	Peptide binding proteins	171-172
3.4.5.2.2.	ABC transporter peptide binding protein binding	172-173
3.4.5.2.3.	Dioxygenase and ABC transporter protein, binding protein interactions	173-175
3.4.5.3.	Myosin ATPase binding	175
3.4.5.3.1.	Myosin ATPase	175-176
3.4.5.3.2.	Myosin motor ADP-metvanadate-resveratrol binding	176-177

3.4.5.3.3.	Dioxygenase and Myosin ATPase interactions	177-179
3.4.5.4.	Immunoglobulin like protein binding	179
3.4.5.4.1.	Immunoglobulin like proteins	179-180
3.4.5.4.2.	Immunoglobulin kappa (κ) light chain dimer binding	180-181
3.4.5.4.3.	Dioxygenase and immunoglobulin kappa chain dimer interactions	181-185
3.5.	Inference	186
CHAPTER 4		
HERBICIDES AND DIOXYGENASE INHIBITORS		187-256
4.1.	Introduction	187-188
4.1.2.	Dioxygenase inhibitors	188-190
4.1.3.	Abscisic acid	190-197
4.1.4.	Abscisic acid signaling	197-212
4.1.5.	Signal Transduction	212-213
4.1.6.	Inhibition induced by ABA response or selective herbicides	214-231
4.2.	Important role of conserved residues and Inhibitor interactions	231
4.2.1.	Indoleamine imidazole binding	231-233
4.2.2.	Substrate binding and reaction mechanism of HPPD	233-237
4.2.3.	2, 4-D and Dicamba binding	237
4.2.3.1.	Dicamba binding	238
4.2.3.2.	2, 4-D binding	239-246
4.3.	Self Hydroxylation: Herbicide degradation and Flavor biogenesis in Strawberry	246-256
CHAPTER 5		257-260
METHODS		
CHAPTER 6		261-266
CONCLUSION		
REFERENCES		267-286

CHAPTER 1

INTRODUCTION

Introduction

A dioxygenase is an enzyme which incorporates both the atoms of molecular oxygen into kinds of substrate, undergoing various types of mechanisms. The cleavage of aromatic ring compounds could be an important function of dioxygenases, which result in the degradation of aromatic compounds. According to the modes of the cleavage of aromatic ring, the substrate for dioxygenases could be classified into two different groups or subfamilies based on sequence, folds, bound ligands and the orientation of amino acid residues in the active sites. One kind of dioxygenase is known as Intradiol dioxygenases, which cleave aromatic rings mainly between two hydroxyl groups (ortho- cleavage) using a non- haem (FeIII) for scission of the ring e.g. catechol 1,2-dioxygenase, chlorocatechol 1,2-dioxygenase etc. The second kind of dioxygenases is known as Extradiol dioxygenases, which cleave aromatic rings mainly between a hydroxylated carbon and other adjacent non-hydroxylated carbon (meta-cleavage) using non-haem (FeII) for scission of the ring e.g. catechol 2, 3-dioxygenases, biphenyl-2,3-diol 1,2-dioxygenase.

Based on the substrate, dioxygenases could be further classified broadly into carotenoid cleavage dioxygenases (CCDs) and specifically into 9-cis-epoxycarotenoid cleavage dioxygenases (NCEDs), which catalyze the rate limiting step in ABA biosynthesis, supposed oxidative cleavage of the carbon double bonds of 9-cis-neoxanthin or 9-cis-violaxanthin. The oxidation of aldehyde xanthoxin is followed by two subsequent reactions which result in the biologically active Abscisic Acid (ABA). In Maize (*Zea mays*), VP14 (Viviparous 14) is a kind of 9-cis-epoxycarotenoid dioxygenase. Thus, 9-cis-epoxycarotenoid dioxygenases (NCEDs) including VP14 are specific regulators which determine levels of ABA, which in turn control ABA-regulated processes. The VP14 was the first carotenoid cleavage enzyme to be cloned from any organism. VP14 also plays a

role in the strigolactones synthesis; the signaling molecules involved in shoot branching, symbiotic growth of arbuscular mycorrhizal fungi and parasitic weed germinations. VP14 homologues have been identified in other plant species as well. VP14 like other proteins have also been identified in different plants, prokaryotes and animals. The physiological effects of ABA or apocarotenoid biosynthesis and these results pave the way for a mechanistic understanding of ABA biosynthesis, apocarotenoid biosynthesis or retinoid functions. Yet, although, various mechanisms for the rate limiting step in ABA biosynthesis has been proposed for 9-cis-carotenoid double bond cleavage; the structure and the properties of specific determinants of dioxygenase specificity pertaining to substrate specificity related amino acid motifs or encoding ring motifs are subject of intensive study.

1.1. Introduction:

The pseudomonas lignostilbene dioxygenases catalyze a double bond cleavage reaction on a noncarotenoid substrate. The β , β - carotene 15-15'-dioxygenase involved in Vitamin A synthesis has been identified in drosophila, mouse and chicken. Genes encoding enzymes which have functions of catalyzing the cleavage of 9-10 carotenoid double bond has been identified in various plant species and mouse (von Lintig and Vogt 2000; von Lintig *et al.*, 2001; Kiefer *et al.*, 2001). Collectively, these enzymes are known as carotenoid cleavage dioxygenases (CCDs). In vertebrates, in the retinal epithelial tissues, highly expressed RPE65, has been identified which is another, similar kind of protein, involved in retinoid metabolism and in the visual cycle perhaps could be essential for the E/Z isomerization of trans retinol to 11-cis retinol (Kiser *et al.*, 2009).

Despite the significant biological functions of apo carotenoids in different organisms and the increase in the number of putative CCDs or NCEDs in sequence databases, the mechanisms followed by such enzymes to catalyze the reaction are only a little known.

Previous studies have shown that a 9-cis c=c bond configuration is essential for the violaxanthin and neoxanthin cleavage by the recombinant VP14 protein (Messing *et al.*, 2010). All trans isomers of these carotenoids are not cleaved by the VP14 and also on adding into the assays with cis-isomers, they do not show competitive inhibition. The enzyme could tolerate some flexibilities and variability in the ring region distal to the 9-cis double bond is indicated by the dissimilarity in the region (Sergeant *et al.*, 2009). The VP14 also cleaves 9-cis isomer of zeaxanthin containing two β rings and no epoxy groups. In plants different types of VP14 substrates cleaved by the enzyme is not significant because in most of the plant tissues; and the only carotenoids with a 9-cis configuration are violaxanthin and neoxanthin. However, binding and reaction mechanisms of VP14 and other CCDs this plasticity could be beneficial in understanding the mechanistic characteristics. In the building and optimization of CCDs or NCEDs, several structural features have been optimized; the structure of VP14 in maize indicates that VP14 has a seven blade β -propeller with four α -helical inserts forming an α -helical domain on top of the β -propeller (Messing *et al.*, 2010). A long tunnel runs across the centre of the structure from one end to the other, as in other β -propeller structures. For dioxygenase activity, the required Fe²⁺ is located inside the tunnel on the central axis of the β -propeller (Kloer *et al.*, 2005). Four His residues are located on the innermost strands of the blades which are octahedrally bound; a water molecule and an elongated density are interpreted as corresponding to a molecule of surrounding dioxygen. Only a few enzymes such as superoxide dismutase, photosystem II, fumarate reductase have such type of Fe²⁺ coordination (Messing *et al.*, 2010). All attempts to refine the structure indicate that molecular oxygen is presumably the most important ligand at this position from a crystallographic and mechanistic point of view. The observed coordination indicates that molecular oxygen supports a cleavage mechanism which involves a dioxocyclobutanyl

(dioxetane) intermediate. Also, the structure shows a shaft lined with charged residues beginning at the bottom of the base of β -propeller and ends at the metal coordination centre. The oxygen access to the active sites in apocarotenoid-dioxygenases (ACOs) and in VP14 is possibly through the tunnel that ends short of the iron blocked by histidine residues. This section of the carotenoids includes the 9-cis bond and the c=c double bond cleavage sites, which are held over the catalytic iron and oxygen by three phenylalanine residues. The residues Phe, Leu, Met, Val, Trp and Pro surround the 9-cis bond and its proximal methylenecyclohexane group. These residues are not only involved in making hydrophobic interactions which are non-specific about the substrate, but also provide reinforcement to the molecule such that c=c bonds align exactly over the catalytic iron and the oxygen. Analysis of the surface properties of α -helical domain of VP14 shows a hydrophobic patch formed by 25 hydrophobic residues penetrating the membrane beyond the head groups of the membrane phospholipids and interacts with the fatty acid residues in the membrane interior. In the thylakoid membrane the penetration of hydrophobic patch of VP14 places the substrate tunnel in close proximity to carotenoid. Importantly, the primary substrates of VP14, 9-cis-violaxanthin and 9-cis neoxanthin are 10 times more abundant in the thylakoid membrane than in other plant membranes. Throughout the CCD family, it may be significantly consistent and a significantly smaller hydrophobic patch was also observed in Prokaryotic Apocarotenoid dioxygenase structures (Messing *et al.*, 2010).

The structural features of β and ϵ -ring carotenoids include similarities and differences with 5, 6-epoxy carotenoid, the endogenous substrate for the biosynthesis of ABA (Finkelstein *et al.*, 2002). The configuration of ϵ -ring is almost identical to the 5, 6-epoxy ring at the polyene chain and the ring conjecture. However, the orientation of hydroxyl group, 3-hydroxyl in the β and 5, 6-epoxy rings makes it different from the ϵ -ring. In 9-cis lactucaxanthin, the presence of ϵ -rings at each end, due to steric constraints reduces the

activity of the enzyme which shows that the 3-hydroxyl group has a significant effect on the of substrates cleavage containing a ϵ -ring. Alternatively, the hydroxyl group could be involved in forming hydrogen bond with an active site residue.

The structural and other studies of VP14 protein can be significant to characterize the cleavage reaction mechanism, after determination of the structure of the protein and the iron coordination in the active site, further analysis of the activity of the substrate could provide a better understanding of how CCDs and NCEDs function (Messing *et al.*, 2010; Fritze *et al.*, 2004; Zhang *et al.*, 2007; Kiser *et al.*, 2009).

1.1.1. Dioxygenase and oxidative transformation reactions:

The oxidation of an unactivated alkane C-H bond, which are stereoselective a interconversion of functional groups, metal- dependent oxygenases or oxidases, most often carry out such reactions. The best characterized of these enzymes include the cytochrome P450 monooxygenases (Wong *et al.*, 1998), which belong to the oxygenases or oxidases which use iron as a cofactor and constitute redox enzymes 'superfamily', which also include both enzymes using diiron for example, ribonucleotide reductase and methane monooxygenase, and enzymes using monoiron that include those dependent on Fe (III), e.g. the intradiol cleaving catechol dioxygenases and lipoxygenase and those dependent on Fe(II), e.g., the extradiol cleaving catechol dioxygenases, benzene/naphthalene dioxygenases, Tyr/Phe hydroxylases and the 2-oxoglutarate (2OG)-dependent dioxygenases (Schofield and Zhang 1999).

These 2OG-dependent dioxygenase catalyzed oxidative reactions are mechanistic steps in the biosynthesis of types of metabolites in plants as well as mammals, which include flavor compounds and materials of agrochemical and medicinal importance, such as plant

hormones (abscisic acid, gibberellins) and antibiotics (the β -lactamase inhibitor clavulanic acid and cephalosporins).

The best characterized 2OG-dependent dioxygenase in mammals is prolyl-4-hydroxylase, which, in collagen, catalyzes the post-translational proline hydroxylation. The prolyl-4-hydroxylases regulate the hypoxia inducible factor (HIF) and its binding to the hypoxia response elements (HRE), involving ISWI and SWI/SNF histones (Flaus *et al.*, 2006; Kenneth *et al.*, 2009).

The 2OG-dependent dioxygenases require Fe (II) and these enzymes catalyze variety of two electron oxidations, including desaturation, hydroxylation and reactions involving oxidative ring closure. In all cases, the incorporation of one of the oxygens of the dioxygen molecule into succinate is the end result of the oxidation of prime substrate is coupled to the 2OG conversion into succinate and CO₂, whereas the other dioxygen-derived oxygen is presumably converted to water in the case of desaturation reactions. Hydroxylation reactions partially incorporate oxygen from dioxygen into the alcohol products with the observation of significant exchange of oxygen from water (Lloyd *et al.*, 1999; Baldwin *et al.*, 1993).

1.1.2 Oxygen activated by metal ions:

The dioxygen mediated oxidation of hydrocarbons is exothermic and despite of this exothermicity, in the absence of suitable catalyst dioxygen remains chemically unreactive. Probably, the reason could be that the dioxygen (³O₂) ground state in the highest occupied π^* orbitals contain two unpaired electrons and therefore, forbidden spin can react with spin-paired singlet species. In contrast, dioxygen (¹O₂), singlet excited state is highly reactive against alkenes and dienes because of the pair of valence electrons. However, for oxygenase enzymes it is not feasible to access this excited state, since, in energy ¹O₂ is

22kcal/mol higher than $^3\text{O}_2$. Thus, to activate dioxygen three strategies can be opted by transition metal ions having unpaired electrons.

1.1.3. Orbital overlap with a metal ion:

The complex formation of dioxygen with a transition metal ion having unpaired electrons in 3d orbital facilitate the overlap of unpaired electrons in the dioxygen π^* orbital with those on the metal ion (Jones *et al.*, 1979). In the transition metal-dioxygen complex such reactions with a singlet organic reagent are permissible, provided the overall number of unpaired electrons in the complex remains constant.

1.1.4. Single electron transfer:

In metallo-enzymes, the dioxygen activating transition metals have two available consecutive oxidation states e.g. Fe(II)/Fe(III) or Cu(I)/Cu(II), hence, the metal centre can carry out single electron transfer to bound oxygen. The $^3\text{O}_2$ oxygen ground state by receiving a single electron can form superoxide, thus being an allowed reaction, is one possible route for the activation of oxygen. Superoxide after generation can participate in different types of chemical reactions involving 1-or 2- electrons (Sawyer *et al.*, 1991). Since, the reaction is apparently energetically unfavorable, this route is not widely appreciated but if the enzyme could interact with bound superoxide but not with dioxygen (e.g. electrostatic interactions, protonation or hydrogen bonding), then the reaction becomes favored. In hemoglobin, such activation of dioxygen to superoxide has been observed (Winterhalter *et al.*, 1976) and EDTA-Fe(III) complexes (Brigelius *et al.*, 1974). Therefore, this route is feasible and can be used by metallo-enzymes such as extradiol catechol dioxygenases. The organic reagents catalyzed reactions that allow a stable radical intermediate are other examples of single electron transfer to dioxygen. Thus, flavoprotein hydroxylase enzyme with its reduced flavin cofactor via a single electron transfer activates

dioxygen resulting in the formation of a stable superoxide and semiquinone, which on recombination form a hydroperoxide intermediate (Massey *et al.*, 1994).

1.1.5. Reaction with a substrate radical:

Reactions involving substrate radicals are alternative possible mechanism because dioxygen reactions via a radical mechanism is a spin allowed process. It has been proposed that in the intradiol catechol dioxygenases, this kind of substrate activation mechanism usually occurs, where a hydroperoxide radical is formed by an attack on dioxygen by bound catechol semiquinone intermediate (Que *et al.*, 1996).

1.1.6. Nucleophilic reactivity:

Hydroperoxides, similar to other reagents containing adjacent heteroatoms (e.g. N_2H_4 , NH_2OH) possess a high order of nucleophilicity. Intermediates containing an adjacent carbonyl group after nucleophilic attack forms a dioxetane intermediate e.g. luciferase enzyme, where luminescence is caused by a reaction which involves formation of a dioxetane intermediate which decays and an excited state of the product is released, which causes luminescence (McCapra *et al.*, 1976). This type of reactivity is a common mechanism in the action of carboxylases dependent on vitamin K, where a conjugate addition on the adjacent alkene is caused by vitamin K hydroperoxide, and the subsequent reaction which results in exposed dioxetane to form an epoxide (Dowd *et al.*, 1991). In such cases, base catalysis is required to generate nucleophilic hydroperoxide anion.

1.1.7. Electrophilic reactivity:

The nucleophile attack resulting in the O-O bond cleavage of a hydroperoxide and hydroperoxide acting as an oxidizing agent for example, in the case of flavoprotein hydroxylases, the formation of flavin 4a hydroperoxide intermediate and p-

hydroxybenzoate hydroxylase, where, hydroxylation of the aromatic ring is the end result of the reaction of phenol nucleophile and flavin hydroperoxide (Massey *et al.*, 1994).

1.1.8. O-O bond cleavage via C-C bond migration:

The C-C bond migration onto electron deficient oxygen upon acid catalysis can occur. In the hydrogen peroxide and peracids mediated oxidation of ketones, such type of 1, 2-rearrangement can occur (Crow *et al.*, 1993). In these reactions, the migratory aptitude of migrating groups is influenced by the substituents effects; recent studies have shown that the selectivity of these rearrangements is determined by the stereoelectronic factors (Goodman *et al.*, 1998).

1.1.9. O-O cleavage via homolytic cleavage:

Organic peroxides fragmented by homolytic cleavage to generate an alkoxy radical in the presence of transition metal catalysts, such as iron salts, which can be further fragmented through homolytic cleavage of an adjacent carbon-carbon bond (Schreiber *et al.*, 1980). In these cases, the oxygen-oxygen bond cleavage is facilitated presumably by Fe(II) complex formation, while, iron(III) oxo complex is the other product in homolytic cleavage. In non-heme iron-dependent metallo-enzyme- and cytochrome P450 catalyzed reactions, catalytic intermediates are high oxidation state iron-oxo species (Sono *et al.*, 1996).

The hydroperoxides can undergo a number of possible types of reactions and hence dioxygenase enzyme catalyzed reactions have several different mechanistic possibilities, which can be observed in various dioxygenase-substrate complexes that involve different types of substrate groups ranging from substrates containing phenyl and dioxan rings to epoxy group exposed side chains (Bugg, Timothy. 2003).

1.2. Strawberry metabolism

1.2.1. Recent advances in strawberry metabolomics:

Particularly, strawberry has a secondary metabolite composition, similar to other Rosaceae family fruit including apples, plums, pear, peaches and raspberries. Hundreds of volatile and non-volatile compounds, the volatile ones are responsible for the typical fruit aroma. For decades, these metabolites have been the subject of intensive study. These focused studies include the most analyzed metabolites from strawberry are phenolic compounds such as phenolic acids, flavonols (quercetin and kaempferol derivatives), anthocyanin (pelargonidin and cyanidines derivatives), proanthocyanidins, galloylglucoses and ellagitannins. Additionally, some terpenoid compounds, some metabolites containing nitrogen and other volatile metabolites have been identified in strawberry. Metabolites such as micronutrients e.g. vitamin C and folate have been analyzed to determine its nutritional quality. The outcomes of these phytochemical analyses are included in the database for nutritional composition and are beneficial for health. These studies also have contributed to the knowledge about strawberry physiology.

Recent development in metabolomics technology has provided approaches to analyze the chemical composition of plants and by applying these approaches the strawberry metabolite repertoire has been extended which include number of variants of metabolites known so far, in addition to the identification of currently not known metabolite classes supposed to exist in strawberry. Also, in strawberry natural products generated by different organs and cell types in common to other plants, interact with the environment and forms a basis for chemical defense. In humans, these compounds upon consumption are responsible for the beneficial health effects (Hanhineva, Kati *et al.*, 2011).

1.2.2. Strawberry fruit ripening and fruit texture:

Climacteric and non-climacteric are two types of fleshy fruits, the climacteric fruits at the onset of ripening, show an increase in both ethylene production and respiration (Giovannoni, 2001). In climacteric fruit ripening, towards understanding the ethylene action many studies report their molecular mechanisms (Adams-Phillips *et al.*, 2004); however specific signal or signals activating the ripening in non-climacteric fruits are now identified as other studies indicate that the regulation of non-climacteric fruit ripening could be associated with the action of phytohormone abscisic acid (ABA) (Zhang *et al.*, 2009a). ABA has various plant growth and development related important functions and also has implications in plant responses to environmental stresses (Hirayama and Shinozaki 2007). In addition, in fleshy fruits, ABA promotes sugar accumulation and metabolism (Pan *et al.*, 2005).

Some studies reported that 9-cis-epoxycarotenoid dioxygenase (NCED) is a key enzyme, involved in ABA biosynthesis (Qin and Zeevaart, 1999), which regulates fruit ripening. Strawberry (*Fragaria ananassa*), a model to study non-climacteric fruits, involves plant hormone abscisic acid as a ripening inducer (Kano and Asahira, 1981; Manning, 1994; Perkins- Veazie, 1995; Jiang and Joyce, 2003).

In the reference plant *Arabidopsis* (*Arabidopsis thaliana*), molecular mechanisms of ABA signaling has been significantly improved, which can potentially explain the ABA mediated signaling and regulatory process in fruit ripening. Over the past decades, in *Arabidopsis*, various ABA- signaling components have been identified, such as phospholipases C/D and G-proteins; different protein kinases e.g. receptor like kinases, SNF1 related protein kinases (SnRKs), calcium-dependent protein kinases, mitogen-activated protein kinases and calcineurin B-type protein kinases; protein phosphatases e.g.

type- 2C /A protein phosphatases (PP2C/A); and various classes of transcription factors, e.g. members of the MYB/MYC, APETALA2, B3, bZIP and WRKY families (Nambara and Marion-Poll, 2005; Hirayama and Shinozaki, 2007; Wang and Zhang, 2008; Cutler *et al.*, 2010). A plasma membrane ABA receptor, GTG1/GTG2 (Pandey *et al.*, 2009), cytosolic ABA receptors PYR/ PYL/RCAR (Ma *et al.*, 2009; Park *et al.*, 2009) and a chloroplast/plastid ABA receptor, H-subunit of magnesium chelatase, ABAR/CHLH (Wu *et al.*, 2009; Shang *et al.*, 2010) have also been identified, supporting the existence of two different ABA signaling pathways in Arabidopsis. The PYR1-PP2C-SnRK2 core signaling network is the one signaling model, in which interaction between PP2C and PYR1 is facilitated by ABA, that results in the SnRK2 activation and PP2C inhibition, which in turn channelize the ABA signals through the phosphorylation of downstream factors i.e. AREB/ABF (Fujii *et al.*, 2009; Ma *et al.*, 2009; Park *et al.*, 2009). The ABAR-WRKY40-AB15 signaling network is the other model, include the ABAR-WRKY interaction, which is stimulated by ABA and AB15 gene is relieved from inhibition by repression of WRKY40 expression (Shang *et al.*, 2010). The ABA perception initiate the ABA action that activates downstream signaling cascades for the induction of physiological effects; in the strawberry fruit ripening the molecular mechanism of the ABA regulation indicates that the Arabidopsis ABA receptor gene ABAR and the H-subunit of magnesium chelatase, namely FaCHLH/ ABAR homolog and closest ortholog from peach and castor bean, is present as a single copy in the strawberry genome. The FaCHLH/ABAR functions throughout the plant, is apparent from the expression pattern of this gene as, in the strawberry plant FaCHLH/ABAR gene is ubiquitously expressed. In the developing receptacles, the transcript of FaCHLH/ABAR and protein contents indicated that during ripening the green fruits had greater expression of FaCHLH/ABAR than in nongreen fruits and the expression increased further with the concomitant onset of red coloration, which

supports that FaCHLH/ABAR is associated with strawberry fruit ripening also that in ripening strawberry fruit a key component of the ABA signaling pathway is FaCHLH/ABAR. The evidences support the hypothesis that presumably strawberry fruit ripening is positively regulated by ABA and sugars, especially Suc, promote partly the fruit ripening by regulating ABA contents. Further, in strawberry fruit ripening, ABA-mediated signaling include down-regulation of positive regulators of ABA signaling (SnRK2 and ABI3, ABI4, ABI5), up-regulation of a negative regulator of ABA signaling (ABI1) and pigment biosynthesis and sugar metabolism related genes (SigE, CHS and AMY) agree with the positive role for FaCHLH/ABAR (Jia *et al.*, 2011).

The beginning of strawberry fruit ripening is followed by accumulation of sugar, decreased acid content and increased cell-responsive cellular changes to promote ABA accumulation such as sugar, osmosis stress, turgor and pH act as initial and early signals. Two ABA receptors: FaPYR1 and FaCHLH/ABAR or G-protein coupled receptors strawberry homolog (GTGs/GCR2 (the upstream of GPA1, termed as FaGTG/GCR), receive the ABA signal (at least). The signals from these receptors are channelized respectively as follows: Signals from FaABAR/CHLH through transcription factors SigE or WRKY, FaPYR1 signals through PP2C (ABII) and signals from FaGTG/GCR presumably through GPA1. Thus, second messengers, protein kinases, protein phosphatases 2Cs and different transcription factors participate in the ABA signaling pathway. The fruit sugar and pigment metabolism are regulated by signals from FaABAR/CHLH and FaPYR1 transmitted through transcription factors such as ABF and SigE; the fruit softening (fruit texture) is regulated by FaGTG/GCR2 signaling through Ca²⁺ ion channels (Chunli li *et al.*, 2011).

In different cell compartments certain isoforms of lipoxygenases (LOX) have been identified and their activity has been linked to membranes of different origins, such as lipid bodies (Hause *et al.*, 2000), chloroplast (Rangel *et al.*, 2002), isolated plasma membranes

(Vianello *et al.*, 1995) and tonoplasts (Trabanger *et al.*, 1991); the enzymatic and chemical properties of plant lipoxygenases have been implicated in the plant metabolism. Since, a number of lipoxygenase isoforms are present in the different subcellular compartments; this makes it difficult that each LOX isoform would be assigned a specific function. Some evidences support that processes, such as regulation of growth, maturation, senescence and wound- and pathogen- induced defense are correlated with LOXs (Antonella leone *et al.*, 2006).

1.3. Absciscic acid and herbicide inhibitors:

Herbicides are weed killers. Those which kill specific targets are known as selective herbicides, while the desired crops are left relatively unharmed. A few of such herbicides are synthetic ‘versions’ of plant hormones that act by interfering with the growth of the weed and there are some natural herbicides produced by plants e.g. genus *Juglans* (walnuts). The interactions of natural herbicides and other related chemicals in the environment are called allelopathy. Herbicides are used in agriculture as an agent for landscape turf management. Over decades, all efforts to investigate the growth regulator or auxin herbicides and the physiological and biochemical basis of their phytotoxicity and their molecular interactions had been elucidated in recent studies (Grossmann *et al.*, 2003). The identification of receptors for auxin perception (Dharmasiri *et al.*, 2005; Kepinski *et al.*, 2005; Tan *et al.*, 2007) and hormone interactions in signaling between ethylene, auxin and the up-regulation of NCED in ABA biosynthesis (Kraft *et al.*, 2007), leading to ABA accumulation in dicot plants are some important steps towards understanding of auxin herbicide action. In crop production auxin herbicides are implicated in a new mechanism for weed control as in the plant their systemic mobility and in cereal crops, their specific action preferably against dicot weeds. Auxin herbicides belong to different chemical classes, which include benzoic acids phenoxycarboxylic acids, aromatic carboxymethyl

derivatives, pyridine-carboxylic acids and quinoline carboxylic acids. Moreover, the carboxyl group of the dissociated molecule containing a strong negative charge separated from planar aromatic ring containing the weaker positive charge with a distinct distance is the structural requirement for their activity (Grossmann *et al.*, 2003).

Auxin herbicides in higher plants mimic the main auxin indole-3-acetic acid (IAA) mode of action. However, their higher stability in the plant causes that they are long lasting and therefore, more effective than IAA. At the cellular sites of action, when present at low concentration, auxin herbicides stimulate various processes including growth and development. However, as the auxin concentration and tissue activity increase, the growth is disturbed which causes lethal damage to plants. The molecular viewpoint about auxin action supports the extensive perception of auxin receptors and presumably, the primary mechanism of herbicide action is subsequent signal transduction process because of a result of more than optimal endogenous auxin concentration. The identification of transport inhibitor response 1 (TIR1)-type auxin receptors, initiated the importance of a perception mechanism which could account for a large part of the response mediated by auxin. The F-box protein TIR1 is the recognition component of Skp-cullin-F-box-ubiquitin-protein-ligase (SCF^{TIR1}) complex, which participates in the ubiquitin-proteasome pathway for protein degradation (Dharmasiri *et al.*, 2005; Kepinski *et al.*, 2005). The TIR1 substrates i.e. in an auxin-dependent manner, transcriptional repressor proteins of Auxin/IAA are recruited to TIR1 and after binding to TIR1 these repressor proteins are degraded (Dharmasiri *et al.*, 2005; Kepinski *et al.*, 2005; Hagens *et al.*, 2002). IAA binds to the base of same TIR1 pocket that put Aux/IAA substrate on top of IAA filling the rest of the TIR1 pocket as reported in some crystallographic studies (Tan *et al.*, 2007). To enhance TIR1-Aux/IAA interactions IAA seems to function as a 'molecular glue' (Tan *et al.*, 2007), resulting in subsequent ubiquitination of Aux/IAA proteins, tagging them as substrates to

be degraded by proteasomes (Dharmasiri *et al.*, 2005; Kepinski *et al.*, 2005; Tan *et al.*, 2007). The Aux/IAA repressor proteins degradation leads to pre-existing DNA-binding transcriptional activator proteins derepression, the auxin response factors, ARFs (Hagen *et al.*, 2002); Transcription of auxin- responsive genes are continuously activated by ARFs until auxin concentrations remain high. TIR1 binds and functionally responds to the auxin herbicide, 2,4-dichlorophenoxyacetic acid (2,4-D) and 1-naphthalene acetic acid (1-NAA), which is consistent with the hypothesized action of auxin herbicides as synthetic versions of IAA (Dharmasiri *et al.*, 2005; Kepinski *et al.*, 2005). In addition, it is clear from the crystal structures of TIR1 complexed with auxin ligands, IAA, 2, 4-D and 1-NAA that IAA binds to a promiscuous site at the base of the TIR1 pocket which also accommodate auxin analogues (Tan *et al.*, 2007). Furthermore, the binding interaction shows that IAA binds to the TIR1 with greater affinity, which involves the ring system as well as side chain carboxyl group (Tan *et al.*, 2007). A more important factor which is responsible for growth inhibition and in the actual phytotoxic response to auxin is the ABA overproduction (Grossmann *et al.*, 2003). Previous studies on a variety of dicot species demonstrated that induction of ABA accumulation was the result of auxin herbicides from different chemical classes and IAA (Grossmann *et al.*, 2003). For the first time, Hansen *et al.*, 2000 reported that auxins trigger *de novo* ABA biosynthesis in the highly sensitive dicot cleavers (*Galium aparine*). Moreover, further analysis of intermediates of this pathway indicated that ABA biosynthesis is exclusively induced in the shoot tissue by increasing xanthophyll cleavage, resulting in increased production of xanthoxin which is a precursor of ABA (Hansen *et al.*, 2000) and the plastid enzyme 9-cis-epoxycarotenoid dioxygenase (NCED) catalyzes this key regulated step in the pathway; this enzyme is encoded by the family of NCED genes (Taylor *et al.*, 2005). The study had additionally excluded that the NCED was presumably regulated by auxin treatment as ABA accumulation could be due to the specific events such

as enhanced NCED precursor supply, steps in the pathway after NCED or reduced ABA degradation and conjugation (Hansen *et al.*, 2000). Varieties of annual and perennial broadleaf weeds are treated with synthetic auxin herbicides which are shown to be effective. The most commonly used synthetic auxin herbicides are 2,4-dichlorophenoxy acetic acid (2,4-D) and 3,6-dichloro-2-pyridinecarboxylic acid (Dicamba) (Hagood *et al.*, 2001; Heap *et al.*, 2002). These compounds have similar mode of action to the endogenous auxin (IAA), thereby disrupting cell wall functions and nucleic acid metabolism, leading to abnormal cell division and plant growth, with a concurrent increase in ethylene production. Plant twisting and growth inhibition followed by necrosis and plant death are some of apparent symptoms (Ahrens *et al.*, 1994), thus, the application in field corn should be made early to avoid injury symptoms associated with later applications such as stalk brittleness, brace root formation. Volatilization of these herbicides can injure sensitive non target plants (Ahrens *et al.*, 1994; Hagood *et al.*, 2001). Sulfonyl urea herbicides was the first group of registered herbicides for the inhibition of acetolactate synthase (ALS) enzymes, which is responsible for the bio synthesis of branched chain amino acids, leucine, isoleucine and valine (Beyer *et al.*, 1988; Ahrens *et al.*, 1994). Various ALS inhibitors such as nicosulfuron, primisulfuron, rimsulfuron, thifensulfuron, halosulfuron and flumetsulum are registered for control annual and perennial weeds and are applied on all major field crops and some minor crops. For corn weed management, herbicides with different mode of action are available and new chemistries are needed to generate herbicide tolerance by genetically modifying corn hybrids and to control a broad-spectrum of weeds that involves the utilization of this additional knowledge. Further, to manage weed resistance and weed-species shifts, there are certain advantages of mixing herbicides with different modes of actions. Herbicide families that are effective in plant systems without detriment effects to the environment or humans are the major concern of researchers.

Photosynthesis, chlorophyll and carotenoid production in plant systems offer areas for inhibition of plant development which has limited risk to animals. Almost, all photosynthesizing organisms contain pigments such as carotenoids, which absorb light at the 400-500 nm range, effectively broadening chlorophyll light absorption capacity. Absorbed light is then transferred through resonance transfer to chlorophyll from carotenoids. However, protecting plants from photo-degradation is another important function of carotenoids (Cunningham and Gantt, 1998). Chlorophyll usually receives more light during photosynthesis, than it can dissipate; the excess energy is transferred to chlorophyll through resonance transfer, which in turn excite the molecules to a triplet state (^3Chl) and oxygen can interact with these excited molecules, resulting in the formation of highly reactive singlet oxygen species. Singlet oxygen species are essential for the peroxidation of lipids and also in the degradation of membranes. This peroxidation involves extraction of a hydrogen molecule from unsaturated lipids, leading to the formation of lipid radicals. These reactions ultimately result in cell destruction and plant death. Carotenoids have an important role in preventing this chain of events as they are involved in quenching ^3Chl before it can excite different oxygen species. Carotenoids by collecting toxic oxygen species can also protect chloroplasts by quenching the excess energy of oxygen species before they can damage membrane (Mortenson and Skibsted, 1997). Other studies indicate that in response to various environmental stresses plants produce different carotenoids (Young *et al.*, 1991; Haldimann *et al.*, 1996). Isopentenyl pyrophosphate (IPP) is the central molecule, which form carotenoids and other isoprenoid compounds. This 5-carbon molecule integrates in to larger skeletal structures in multiples of 5-carbons, e.g. the 20-carbon molecule geranylgeranyl pyrophosphate (GGPP), which is a precursor molecule in the isoprenoid pathway for the formation of other molecules. The basic structure of plant carotenoids are formed by two molecules of GGPP (Cunningham

and Gantt, 1998). In terms of carotenoid production, many enzymes and cofactors are involved, which could be potential target for herbicide inhibition of these processes (Britton *et al.*, 1989). Certain chemicals through inhibition of plant physiological process that produce chloroplast pigments, cause chlorotic or white foliage, are known as bleaching herbicides (Britton *et al.*, 1989); these herbicides are categorized into three groups on the basis of their modes of action, some of which including, fluorochloridone, norflurazone and fluridone effectively inhibit the conversion of phytoene to phytofluene which is regulated by the enzyme phytoene desaturase. For control of triazine resistant weeds in field corn, the herbicide Fluorochloridone was investigated, which shows that this herbicide has marginal effects on crop safety (Buhler *et al.*, 1988). Moreover, enzymes in the isoprenoid pathway are being inhibited by Clomazone, which is implicated in the production of farnesyl pyrophosphate (Ahrens *et al.*, 1994). Amitrole is one herbicide which causes carotenoid precursor accumulation in susceptible plants therefore; enzymes in carotenoid, chlorophyll or histidine synthesis could be inhibited. The function of other class of carotenoid inhibitors is through 4-hydroxyphenylpyruvate dioxygenase (HPPD) inhibition. HPPD is the enzyme which facilitates the conversion of 4-hydroxyphenylpyruvic acid to homogentisic acid (Norris *et al.*, 1998; Pallet *et al.*, 1998; Viviani *et al.*, 1998). Since, 4-HPPD is a part of the pathway that converts amino acid tyrosine to plastoquinone, which is a cofactor for the enzyme phytoene desaturase, an important enzyme in the production of carotenoids. In the absence of carotenoids, photo oxidation can damage plants (Norris *et al.*, 1998). Between the Q_B site of Photosystem II and Cyt *b₆/f*, plastoquinone serves as an electron carrier and is involved in transferring electrons in other photosynthetic processes (Ort *et al.*, 1986; Wise and Cook, 1998). The reactive oxygen species generated within the chloroplast are also scavenged by plastoquinone (Wise and Cook, 1998). The production of α -tocopherol (Vitamin E) can be blocked by the inhibition

of HPPD, α -tocopherol which is also an antioxidant compound which is involved in strengthening membrane structures and functions in scavenging free radicals within chloroplast membranes (Pallet *et al.*, 1998). The HPPD inhibitors include three known herbicide families: (i) benzoyl isoxazoles, (ii) pyrazoles and (iii) triketones. The benzoyl isoxazoles and triketones are herbicides mainly for weed management in field corn. The detailed structural information on the molecular target sites of agrochemicals now allows different target site-based approaches to improve both selectivity and potency toward the pathogens, specifically away from non-target organisms. 4-hydroxyphenylpyruvate dioxygenase (HPPD) is the enzyme which is compatible to such an approach, as the target site for recently commercialized herbicides i.e. isoxaflutole and mesotrione and therefore, this could be of great interest for designing novel herbicides (Pallett *et al.*, 1998; Mitchell *et al.*, 2001). The enzyme is present in microbes, mammals and plants and has range of functions in various organisms. In plant and mammalian HPPDs, this enzyme has a relatively low overall sequence homology (21% and 29% amino acid identity, respectively). Fritze *et al.*, 2004 reported the crystal structures of HPPDs from the plants *Arabidopsis thaliana* and *Zea mays*. However, substrate analogues or inhibitors which could be used as herbicides or for drug designing, no HPPD structures complexed with such compounds have yet been described. HPPDs are example of non-heme Fe^{2+} -dependent dioxygenases which are mechanistically related to α -ketoglutarate dependent dioxygenases i.e. prolyl hydroxylase and clavaminic synthase (Ryle *et al.*, 2002). In HPPDs, the acidic side chain of hydroxyphenylpyruvate (HPPA) is acting as the α -ketoacid moiety. The catalyzed reaction is mechanistically complex as it involves chain of reactions such as an oxidative decarboxylation, alkyl group ring migration (an NIH shift) and ring hydroxylation (Crouch *et al.*, 1997). Various HPPD inhibitor classes have been described and 1, 3-diketone moiety in different forms is almost present in all types of inhibitors

(Pallett *et al.*, 2001; Lee *et al.*, 1998; Meazza *et al.*, 2002). On kinetic analyses of potent HPPD inhibitors, it was indicated that they exhibit the characteristics of slow and high affinity binding inhibitors (Garcia *et al.*, 2000; Ellis *et al.*, 1996; Kavana *et al.*, 2003; Neidig *et al.*, 2004). Interestingly, majority of HPPD inhibitors can be used as both plant and mammalian HPPD inhibitors; indeed the physiological effects of HPPD inhibition in both plants and mammals were observed in early studies on HPPD inhibitors (Lock *et al.*, 1998; Lee *et al.*, 1997). The structural basis that how herbicidal benzoylpyrazoles could be potent inhibitor of the plant enzyme and the possible modes of selectivity for the plant enzyme versus the mammalian enzyme, was studied by Cheng *et al.*, 2004; crystal structures of both plant and mammalian enzymes in complex with novel selective and nonselective ligands, was investigated, which serves the basis for herbicide designing. In the study, the inhibitor DAS869 was selected for co-crystallization with *Arabidopsis thaliana* HPPD as it exhibits good herbicidal activity and is a potent inhibitor of both plant and rat HPPDs (Siddall *et al.*, 2003). Because of the bulky substituents, which DAS869 contains relative to other HPPD inhibitors, the binding pocket topography could be potentially defined to a greater extent. The plant-enzyme-inhibitor complex (AtHPPD-DAS869) structure was solved by molecular replacement and was refined to 1.8 Å resolution. The superimposed structure of AtHPPD-869 and that of AtHPPD was studied using program O (Jones *et al.*, 1991). Analysis of the superimposed structure, however, did not indicate any gross difference in the main chain atoms between the unbound and ligand-bound forms. Within the active site pocket, the overall orientation of the inhibitor DAS869 represents the metal ion coordinating three amino acids remain the same but the 1, 3-diketone moiety displaced two water molecules after the incorporation of DAS869 inhibitor. The distances from metal ion to the oxygen atoms were refined to approximately 1.9-2.4 Å, which, maintained the octahedral geometry and provided a strong ligand

orientation and binding force. All reported structural classes of potent HPPD inhibitors, contain an acidic 1, 3-diketone moiety (Pallett *et al.*, 2001; Meazza *et al.*, 2002) and for tight binding, the essential element is the coordination of both inhibitor oxygens to the active site metal ion (Cheng *et al.*, 2004). In addition to metal coordination, the side chains of several residues are involved at inhibitor binding site, such as the phenyl groups of Phe, which form a π -stacking interaction with the benzoyl moiety of the inhibitor compound DAS869. For an intensive study, the detailed comparison of inhibitor interactions between a plant and mammalian enzymes, the rat HPPD (RnHPPD) was cocrystallized with RnHPPD-869, however, only 29% amino acid sequence identity was found between AtHPPD and RnHPPD, yet overall they have very similar structures. Moreover, super imposed structures indicate that there are some regions showing some structural differences. The RnHPPD has very comparable ligand binding features to AtHPPD-869. These results were similar to another study, which reported a selectivity of inhibitors for plants. The large majority of inhibitors were equally effective on both plant and mammalian enzymes, but some of the identified compounds showed significant selectivity towards the inhibition of the plant enzyme. All the selected compounds could be substituted for standard residues that included pyrazole ring undergoes a substitution for a phenyl substitution and a ter-butyl substitution or both. Both of these features were present in DAS645 and shows remarkable selectivity for the plant enzyme. To probe the structural basis of plant selectivity because of its herbicidal activity DAS645 is a useful compound (Cheng *et al.*, 2004). The structure of DAS645 complexed with AtHPPD (AtHPPD-645) was solved to a refinement of 1.9 Å resolution, the ligand binding site which has some notable difference from AtHPPD-869, indicated differences in the positioning of the π -stacking network around the benzoyl moiety in AtHPPD-645. The phenylalanine residue was rotated away from the inhibitor which is because of the steric presence of the 3-(2,4-

dichlorophenyl) substitution on the pyrazole; this significant movement was the result of 3-phenyl pyrazole substitution as the structures of different enzyme-inhibitor complexes without such substitution did not show such movement. Another, significant difference was the positions of the C-terminal helix (residues 400 and above) of AtHPPD, AtHPPD-645 and AtHPPD-869 which exposes the active site more than the C-terminal helix does in RnHPPD-869. Thus, Phe is displaced distally from the active site centre in the plant enzyme as compared to the mammalian enzyme, which can accommodate larger inhibitor molecules in order to bind in the plant active site (Cheng *et al.*, 2004). The rat enzyme RnHPPD, similar to the *Pseudomonas fluorescens* HPPD (PfHPPD), contains a short five residue turn (His354-Gly358) preceding the C-terminal helix, whereas, in Arabidopsis, the enzyme structure contains a unique long loop (Val378-Gly397), the AtHPPD structure also contains Gly-Gly pairs before and after Cys residues, just before the start of the C-terminal helix, which is absent in the rat sequence and could act as hinge regions, thus, the C-terminal helix of AtHPPD appears to be slightly hinged, whereas, that of the RnHPPD is not. The C-terminal helix packs against another helix in both the plant and mammalian HPPD structures. This helix rearrangement was possibly difficult to achieve in the rat enzyme because a short unhinged turn is present preceding this helix and the hydrogen-bond chain. This could explain that DAS-645 has very high affinity for the plant enzyme however inhibition of the rat enzyme was yet undetectable. Combining the observed three key structural differences, which has been rationalized that plant inhibitor selectivity of DAS645 and other plant selective inhibitors of HPPD, elucidated in AtHPPD-869, the Phe residue is distally displaced from the metal ion binding site than corresponding phenylalanine in RnHPPD-869. Moreover, AtHPPD has a more flexible C-terminal helix and more dynamic proline residue than the rat enzyme (Cheng *et al.*, 2004). DAS645 is an inhibitor that contains both an N1-ter-butyl group and a substituted 3-phenyl pyrazole ring,

and in terms of binding to the metal ion it avoids steric clashes by relocating phenylalanine and displacing proline distally, thereby accommodating the larger and bulky inhibitors in the plant enzyme. In contrast, the inhibitor DAS645 collides with the RnHPPD Phe residue because the inhibitor is displaced toward phenylalanine and away from proline, while due to the C-terminal helix inflexibility, this Phe cannot rearrange itself. For generating selectivity in inhibitors; the size, geometry and substitution patterns of the inhibitor are key factors and these steric barriers can reduce the binding affinity dramatically (Cheng *et al.*, 2004). The study elucidated detailed information of inhibitor binding to both plant and mammalian HPPDs and provides useful approach for designing novel herbicides and drugs, which target HPPD. The identification of highly selective inhibitors of the plant enzyme and insights into the structural basis of plant selectivity provided unique and powerful tools for agrochemicals design with improved toxicological activity. Other similar studies that use other target sites and combining available structural information for selective inhibitors provide a wider variety of targets shared by both pathogens and vertebrates which could be exploited for safer and more effective crop protection from chemicals.

1.4. Strawberry flavor biogenesis:

Strawberry flavor is extremely popular worldwide as part of the fruit or as a flavoring added in many manufactured foodstuffs. Therefore, studying the origin of strawberry flavor is a worthwhile and challenging area of food research. Strawberry flavor, being a very complex mixture of nonvolatile and volatile compounds, is one of the most complicated fruit flavors with about 350 components (Latrasse *et al.*, 1991). Flavor research has progressed from qualitative and quantitative analysis of volatile compounds towards the sensory evaluation of the most important ones. Biosynthetic studies have been carried out on these most sensory active compounds, recently genetic studies on these

important flavor components have also been carried out. The enzyme of importance are α -amino transferase, α -ketoacid decarboxylase, α -keto dehydrogenase, alcohol dehydrogenase (ADH) and alcohol acyl transferase (AAT), however, the ADH and AAT enzymes are the most widely studied due to their high presence and activity in ripe fruit. Evidences that amino acids are the precursor of ester can be provided by chiral analysis, for example if the source of 2-methylbutyl moiety of the ester is Isoleucine and stereo specifically derived then it will exhibit the S-configuration. Since, amino acids are playing important role in the development of esters as their precursors for the effects of flavor generation through ripening, it was predicted that the distribution of the different types of esters in the strawberry fruit aroma could be due to the free amino acid composition. To understand their role in flavor formation, the amino acids were quantified separately (Perez *et al.*, 1992). Aliphatic and branched chain alcohols, acids, carbonyl and esters are the outcome of amino acid metabolism. The asparagine, glutamine and alanine identified as important amino acids to be quantitatively determined and the results show that these amino acids are the most abundant compounds but alanine shows most significant changes in concentration during fruit ripening. Leucine isoleucine, lysine and phenylalanine were also found, but at levels of less than 0.25mg/100g of fruit. It was observed that from 30 to 36 days, the alanine concentration sharply increased, stayed invariant until day 41 and then declined at the fully ripe stage. It was concluded that the ester biosynthesis is directly linked and proportional to the alanine concentration and thus when alanine reaches its highest level ester biosynthesis was enhanced. The study demonstrates the importance of amino acids in the development of the different varieties of ester compounds in strawberries aroma. Another study explaining the enzymatic aspects of flavor biogenesis in strawberry fruits and specifically the important role played by AAT enzyme in the biosynthesis of volatile esters was carried out by Perez *et al.*, 1993, shows that all the types

of alcohols such as primary and secondary, saturated and unsaturated, straight chain and branched were present as acetate esters and among all of the acetate esters which are formed by enzymatic reactions, strawberry aroma contains hexyl acetate (4.5%) as the most abundant acetate ester, followed by butyl acetate (1.9%) and amyl and isoamyl acetate (1%) respectively. There is also a clear correlation between composition of ester and the available alcohols. For flavor development even carbon number alcohols such as butanol and hexanol were the most important (Perez *et al.*, 1993). Therefore, in aroma production AAT is a key enzyme which shows some specificity and reacts with a large group of alcohols. An important study reported that, the AAT specific activity increases through maturation in all strawberry varieties. Thus, it was deduced that high AAT activity leads to higher ester production, which subsequently results in fruits with enhanced aroma (Perez *et al.*, 1996a). The AAT substrate specificity studied by Olias *et al.*, 1995 observed that for determining the volatile ester composition in the fruit, the availability of the substrates (acyl CoAs and alcohols) and the inherited properties of the AAT enzyme are the two main factors. The *in vivo* pathway of strawberry flavor volatiles includes the NAD-dependent and NADP-dependent ADH enzymes as important components, as characterized by Mitchell and Jelenkovic, 1995. In strawberries the NAD- and NADP-dependent ADH activities were found to have broad substrate specificities in reactions which include alcohols and aldehydes responsible for strawberry aroma and flavor, the reaction mechanism involves their ester products or directly, thus as a result of the activity of these enzymes a wide range of esters is produced. It was demonstrated in the study that commonly short chain alcohols are the target of NAD-dependent activities and aromatic and terpene alcohols usually undergo NADP-dependent activities (Mitchell and Jelenkovic, 1995). These studies serve as evidences that provide information about major volatiles which are emitted during fruit ripening specifically in the later stages are products of the

genetically modified SAAT enzyme. It is explained by the presence and activity of SAAT enzymes that how variety of compounds are produced in strawberry which is a fruit with a relatively low metabolic energy. These processes seem to be energetically efficient that only a few enzymes are responsible for producing a large number of different volatiles through a single pathway, rather than a number of different pathways all leading to different compounds. However, a large variety of esters originated from a common pathway and due to wide substrate specificity of this enzyme; carbohydrates, different types of esters and furanones are biosynthesized (K.G Bood *et al.*, 2002).

1.5. Strawberry health benefits:

Scientists have been studying the beneficial effects of strawberries on the cardiovascular health, that how they prevent the development of diabetes and heart diseases. The strawberry extracts positively activate a protein called 'Nrf2' which is involved in increasing antioxidant concentration and other protective activities. This protein has functional role in decreasing blood lipids and cholesterol, which can lead to cardiovascular problems. Strawberry consumption has previously been found to counter post-meal blood glucose and low density lipoprotein or cholesterol and therefore decrease risk of diabetes and heart disease, but it is interesting that strawberry extracts have been actively stimulating proteins that offer protection against disease (Thomalley *et al.*, 2012).

Researchers have demonstrated that strawberries have protecting effect in a mammalian stomach that has been damaged by alcohol. Suarez *et al.*, 2011, reported that the strawberry extract can prevent damage to mucous membrane of stomach. The study explains that the positive effects of strawberries are linked to their antioxidant capacity and high content of phenolic compounds e.g. anthocyanins and also to the fact that they can activate antioxidant defenses and enzymes of the human body. The conclusions of the study state that gastric illnesses that are related to the generation of free radicals or other reactive

oxygen species, can be prevented by a diet rich in strawberries. This fruit could slow down the formation of stomach ulcers in humans. Gastritis or inflammation of the stomach mucous membrane is related to alcohol consumption but can also be caused by viral infections or by nonsteroidal anti-inflammatory medication (such as aspirin) or medication used to treat against the *Helicobacter pylori*. Suarez *et al.*, 2011, found that strawberry extract can reduce ulcerations. This study was not meant to mitigate the effects of getting drunk but rather as an approach to discover protecting molecules in the stomach membrane that work against the damaging effects of different agents." New protective medicines with antioxidant properties are currently the need of treatments for ulcers and other gastric pathologies. The compounds present in strawberries could be useful for this purpose. Scientists have assessed the outcome of absorbing specific antioxidant compounds present in strawberries on human consumption. High antioxidant containing foods are the excellent source of compounds which are beneficial for health (Carkeet *et al.*, 2008), and it is speculative that how these compounds are absorbed and utilized within the human body. In the United States strawberries are the fifth most consumed fresh fruit and consumption has more than doubled in the past decade. Strawberry's antioxidants belong to both vitamins and newly defined plant chemicals. Berries contain compounds particularly anthocyanins, which is the source of the berries' blue, purple and red pigments. The study demonstrates that more anthocyanins pigments can be assimilated as intakes increase. The results can explain the health benefits of individual anthocyanins and provide plant breeders, approaches for developing varieties with optimal anthocyanins content.

The nutrition-loaded edibles which promote health but discourage diseases are termed as 'super food' in common parlance. Foods high in antioxidants and phytochemicals which block the development of cancer cells are beneficial against the potentially devastating disease. Strawberries are among such familiar super foods. These tasty red berries are a

significant source of vitamin C, a natural antioxidant capable of attracting and neutralizing free radicals, the invasive and highly reactive molecules which damage the body's endogenous cancer fighting cells. In the human body, the antioxidants can prevent cellular and tissue damage. Wang *et al.*, 2007, led a recent study in multiple species of wild strawberries that investigated the antioxidant capacity and anticancer activity. The study demonstrated that "antioxidants are endogenous plant chemicals which have important roles in promoting human health, while it has been known that wild strawberries can be a good source for obtaining desirable traits by genetic approaches; very little information was available on certain compounds present in specific species of wild strawberries such as their antioxidant activities and their inhibitory effects on the growth of cancer cells." The study demonstrated that such compounds show differing antioxidant capacity and anticancer activity among types of wild strawberries. The study explains about seven different types of wild strawberries containing higher antioxidant levels, with more potential to reduce cancer risk. These seven types can be used for developing cultivars with greater anticancer activity. Among studied genotypes, these types show remarkably greater anti-proliferation effects than other genotypes. Moreover, the outcome of the research will be valuable for different communities including scientists, produce growers and fruit breeders interested in producing berries that are high in antioxidant contents (Wang *et al.*, 2007).

1.5.1. Strawberry as a Space food:

Researchers have found a healthy candidate to help satisfy a sweet tooth of astronauts on long space missions, the strawberry, which requires little maintenance and energy. Scientists, tested several cultivars of strawberries and identified one variety, called Seascape, which seems to meet the requirements to be used as a space crop. Massa *et al.*, 2010; demonstrated that Seascape strawberries are day-neutral, which means they aren't sensitive to the length of available daylight to flower. Seascape was tested with varying

light intensity for different duration as little as 10 hours and as much as 20 hours of daylight. While less light produces fewer strawberries, each berry was larger and the volume of the yields was statistically the same. Moreover, the findings indicated that, since, Seascope strawberry cultivar meets several guidelines set by NASA; hence, it could be a good candidate for a space crop. Since, Strawberry plants are relatively small; it meets mass and volume restrictions. Since, Seascope cultivars provide fewer, but larger, berries under short days, crew members who pollinate and harvest the plants by hand have to tend less labor. It would need less light cuts down energy requirements not only for light lamps, but also for systems which remove heat created by those lights. Strawberries are easy to maintain and there's little waste. Seascope cultivars also had less cycling, because it steadily supplied fruit throughout the investigation period. The plants after starting to flower kept producing fruit for about six months; the earliest space crops will find its use as a part of a "salad machine," particularly a small growth unit that will provide fresh produce which can supplement traditional space meals. Lettuces, radishes and tomatoes are certain crops being considered. Strawberries are considered as a sweet fruit. The idea is to supplement the human diet with fresh berries. The study next plans to grow Seascope strawberries using LED lighting, hydroponics and at different temperature ranges (Massa *et al.*, 2007).

1.6. Dioxygenase Structure and Function:

1.6.1. Dioxygenase structure model, substrate binding domains and inhibitor designing.

For protein structure prediction, comparative modeling is the most reliable and routinely used approach (Eswar *et al.*, 2008; Ginalski *et al.*, 2006). The alternative term homology modeling is more commonly used. Based on the homology between the target protein and the templates (although not always), a structural template (or templates) is identified

typically which is usually reflected by a certain percentage of sequence similarity. When for identifying a template, some advanced Fold recognition (FR) techniques are being used, sometimes templates without any obvious homology relations can be identified on the basis of their structural similarity to the target. This could be a genuine example of convergent evolution or (more frequently) the case when remote homology just cannot be found. Template independent, *de novo* structure prediction is quite difficult and less reliable, although in this area of computational biology a steady progress has been observed (Kolinski *et al.*, 2005; Moult *et al.*, 2007). The contemporary approaches for *de novo* structure predictions involve certain aspect of evolutionary correlation between protein sequences and structures. For the derivation of statistical potentials for *de novo* modeling the evolutionary approaches are essential and/or are implicated in various strategies for extracting building blocks of structure from known protein structures (Pokarowski *et al.*, 2005; Rohl *et al.*, 2004). Classical homology modeling involves three steps. First, a modeling template is identified and sequence alignment between the template and target sequences is generated. Usually, standard tools are being used for template identification, such as PSI-BLAST and the resulting alignments are subsequently corrected by different tools and eventually manually by expert corrections. To identify remote templates FR procedures can also be used (Friedberg *et al.*, 2004). The decreasing sequence similarity implies increasing evolutionary distance and implicated in increasing variation in structure between the template and the target; makes alignments more ambiguous. The template-target alignment heavily determines the accuracy of classical comparative modeling. The second stage involves the generation of corresponding fragments of the target structure, for this aligned fragments of templates are used. When a single template is available, according to the alignment, template coordinates are copied only but when multiple templates are available, a consensus scaffold could be generated, for instance via the

rearrangement of the spatial restraints known from the templates, as it is employed in the MODELLER method (Sali *et al.*, 1993). The important thing of this stage of modeling is construction of missing loop regions in the template scaffold. In the advanced approaches to comparative modeling, using templates as sources of restraints various types of complete structure of the target is built (Kolinski *et al.*, 2005; Baker *et al.*, 2001). Being able to build a model of the target structure which has much similarity to the true structure of the target than to any of the templates used, is the main aim and challenges of such approaches for example distant homology based modeling. The third and final stage of modeling is structure refinement which involves side chains repacking and energy minimization of the whole structure (Kmieciak *et al.*, 2007; Jamroz and Kolinski, 2010). Natural proteins different from random heteropolymers, fold into unique structure patterns. However, only by energetically unfavorable non-ideal features, it is complicated to understand how these are encoded in amino-acid sequences, for example, bulged β -strands, kinked α -helices, buried polar groups and strained loops and that arise in proteins from neutral drift or evolutionary selection for biological function. In the study an approach for ideal protein structures designing has been stabilized which involves completely consistent local and non-local interactions; this is based on relating secondary structure patterns to protein tertiary motifs, making it possible for designing of funnel-shaped protein folding energy landscapes which lead into the target folded state. The study also includes designed sequences which are specifically predicted to fold into ideal protein structures that consists same α -helices, β -strands and minimal loops. Five different topologies, which were designed in the study, were found to be monomeric and very stable to adopt structures in solution nearly identical to the computational models. The study demonstrates how independent from natural evolution the folding funnels of natural proteins arise and provide the foundation for designing functional proteins (Koga *et al.*, 2012).

1.7. Loop modeling:

For predicting protein structure the template-based methods provide models for a significant portion of the protein but often contain conformations such as insertions or chain ends which are difficult to determine. The problem of modeling of the InsEnds onto the rest of the protein is taken in to account by the local structure prediction. In crystal structures loops of >12 residues are considered as standard. However, InsEnds may contain as many as ~50 amino acids making the template-based model of the protein imperfect. To address these issues, Adhikari *et al.* 2011 presented a modeling method for predicting the local structure of loops and large InsEnds in both template based models and crystal structures. The approach uses single amino acid torsional angle flexible moves of the protein backbone labeling at *C β* level representation. Nevertheless, the accuracy for loops is as comparable to other existing methods. They have applied a more constrained method for the structure prediction and refinement categories of the CASP9 structure competitions for improving the quality of homology based models by modeling longer InsEnds, as sizes are generally inaccessible to other existing loop prediction methods. The approach ranks as one of the best in the CASP9 refinement category that involves improving template-based models for crystallographic structure determination as they can function as molecular replacement models to solve the phase problem (Adhikari *et al.*, 2012). Different from the connecting secondary structure elements in protein structures, loop fragments in protein chains are often highly flexible even in usually stable proteins. In terms of a protein's stability, the structural variability of loops is often central to understand the folding and even biological function. The important biological processes are mediated by loops, such as protein-protein interactions, protein-ligand binding and signaling. Under physiological conditions the modeling conformations of a loop remains a problem in computational biology. Other studies including the article by Shehu *et al.*, 2012, reviews computational

methods in loop modeling, highlighting progress and challenges. These studies put insights on potential directions for future research in the area of structural and computational biology.

1.8. Research orientations:

- The basis for designing the pathway for dioxygenase functions.
- These studies provide route to quality homology modeling for crystal structures.
- These studies show great opportunities for inhibitor designing.
- These studies could be basis for exploring different substrate binding domains within the dioxygenase structure and investigating dioxygenase interaction with other proteins that may participate in a common or different signaling pathway.
- The studies related to dietary supplement is a boost for the strawberry implications as dietary supplement for various diseased patients suffering from cardiovascular and metabolic syndrome.
- Luxary diet strawberry can be a potential diet supplement for people suffering from malnutrition.
- These studies provide basis for newly emerging research area that includes **structure aided drug designing**.

CHAPTER 2

SEQUENCE AND STRUCTURE ANALYSIS

SEQUENCE AND STRUCTURE ANALYSIS

2.1. Introduction:

In nature there are a large number of enzymes that are able to catalyze the dioxygen activation and use it to influence a variety of reactions. Most of such reactions are performed by a group of enzymes which are able to incorporate molecular oxygen into kinds of substrates. The stereo selective oxidations of unreactive alkane C-H bond are most often carried out by metal dependent oxygenases or oxidases including dioxygenases. The cytochrome p450 monooxygenase are the best characterized of these enzymes (Wong *et al.*, 1998) that uses iron as a cofactor and constitute a redox enzyme superfamily. The group extends to diiron using enzymes such as ribonucleotide reductase etc. and monoiron using enzymes including Fe (III) dependent Lipxygenases and intradiol cleaving catechol dioxygenase and Fe (II) dependent extradiol cleaving catechol dioxygenase and 2-OG dependent dioxygenase (Schofield and Zhang *et al.*, 1999). The 2-OG dependent dioxygenase studied so far, have an absolute requirement for Fe (II), and use to catalyze two electron oxidation reactions including hydroxylation, desaturation and oxidative ring closure reactions (Lloyd *et al.*, 1999; Baldwin *et al.*, 1993). The dioxygen activating transition metal present in metallo-enzymes have two different oxidation states e.g. Fe (II)/Fe (III), thus facilitating the electron transfer from metal centre to bound oxygen. The route to 1, 2 electron chemical reactions involve the formation of superoxide but since, this reaction is energetically unfavorable this route is often not followed, but if the enzyme in lieu of dioxygen forms stabilizing interaction with bound superoxide then the reaction becomes favorable. Another feasible reaction of dioxygenases with dioxygen includes the spin allowed radical mechanism where reaction with a substrate radical is alternate possibility. It has been proposed that intradiol catechol dioxygenases share such substrate activation and reactions, where a bound catechol semiquinone intermediate attack

dioxygen to form a hydroperoxide radical (Massey *et al.*, 1994; Que *et al.*, 1996). Hydroperoxides in common with other reagents which contain adjacent heteroatoms (NH₂, NH₂OH) shows highly nucleophilic nature (McCapra *et al.*, 1976). A nucleophilic attack is the possibility for intermediates containing an adjacent carbonyl group in terms of forming a dioxetane intermediate. In addition, on the adjacent alkene hydroperoxide forms a conjugate addition which leads to dioxetane to be exposed to form an epoxide (Dowd *et al.*, 1991). In such cases the base catalysis is required to generate hydroperoxide anion such that water could act as a base. The presence of transition metal catalysts, specifically iron salts favor reactions that organic peroxides can be fragment via hemolytic cleavage to produce an alkoxy radical that can be further fragmented via hemolytic cleavage of an adjacent carbon-carbon bond (Schreiber *et al.*, 1980). The oxygen-oxygen bond is cleaved presumably by integration of Fe (II) in such cases, whereas Fe (III)-oxo intermediate is another result of hemolytic cleavage (Sono *et al.*, 1996). Thus, hydroperoxides may undergo a number of possible reactions and hence there are different possibilities for dioxygenase catalyzed reactions that can be observed in dioxygenase-substrate complexes, which involve substrates containing different chemical groups (Bugg, Timothy. 2003).

Strawberry, in technical terms is an aggregate accessory fruit which is being used in food and beverage. Scientists have been studying its beneficial effects on the cardiovascular health and its other protective effects that the strawberry extracts can prevent damage to stomach mucous membrane (Hanhineva, Kati *et al.*, 2011). The antioxidant capacity is not the only beneficial effects of strawberries but they can also activate the antioxidant defenses and enzymes of the human body. Strawberry extracts can also reduce alcerations. Strawberries contain chemical compounds including phenolic compounds and allergens. In the ABA biosynthesis, a 9-cis-epoxycarotenoid dioxygenase (NCED), VP14 catalyzes the rate determining step i.e. the oxidative cleavage of the 11,12-C=C bond of either 9-cis-

neoxanthin or 9-cis-violaxanthin (Messing *et al.*, 2010). The final steps in the Abscisic acid pathway include the oxidation of the C-15 aldehyde xanthoxin, which is converted to the biologically active Abscisic acid through two subsequent reactions (Qin and Zeevaart, 1999). This enzyme and several other dioxygenases have been crystallized and characterized that includes Plant and Bacterial 4-Hydroxyphenylpyruvate dioxygenases (Fritze *et al.*, 2004; Johnson-Winters *et al.*, 2003), bovine RPE65 (Kiser *et al.*, 2009), Human tryptophan-2, 3-dioxygenase, Indoleamine-2, 3-dioxygenase (Zhang *et al.*, 2007).

Strawberry Dioxygenase

2.2. Strawberry dioxygenase amino acid sequence:

```
VSNDGIVVPNPKPSKGLTSKLVDLVEKLVKLMYDSSQPHHYLAGNFTPVIEETPPT
KDLNVIGHLPDCLNGEFVRVGPNPKEFAPVAGYHWFDGDGMIHGMRIKDGKATYV
SRYVKTSRLKQEEYFGGAKFMKIGDLKGLFGLLMVNMQQLRAKLVVVDLSYGH
GTSNTALIYHHGKLLALSEGDKPYVLKVLEDGDLQTVGLLDYDKRLKVEKPHVT
DDHDDVKQKLHSFTAHPKVDPFTGEMFTFGYSHDPPYVMYRVVSKDGMHDPVP
ITVPAPIMMHDFAITENYAIFMDLPLYFKPKEMVKEGKLIFSDETCKARFGVLPKY
AKDDLIRWFELPNCIFIHANAWEEDEVLITCRLENPDLDMVNGPIKKKLDTF
KNELYEMRFLKTGLATQKKLSESAVDFPRVNESYTGRKQRFVYGTTLDSIAKVT
GIVKFDLHAAPEVGKTKIEVGGNIQGLYDLGPGRFGSEAIFVPRVPGITSEEDDGYL
IFFVHDENTGKSAIHVLDAKTMSNDPVAVVELPHRVVPGYGFHAFVTEEQQLQEANL
```

2.2.1 Sequence analysis:

BLAST Results for the sequence provided against non redundant protein sequences:

Id-gi/167989404/gb/ACA13522.1/dioxygenase (Fragaria x ananassa)

Id-gi/163881523/gb/ABY47994.1/carotenoid cleavage dioxygenase 1 (Rosa x damascene)

Id-gi/82548252/gb/ABB82946.1/carotenoid cleavage dioxygenase (cucumis melo)

Id-gi/134285450/gb/ABO69703.1/9-cis-epoxycarotenoid dioxygenase (castanea mollissima)

Id-gi/195972576/emb/CAR57918.1/ carotenoid cleavage dioxygenase 1 (Medicago truncatula)

Id-gi/76560796/gb/ABA43900.1/carotenoid cleavage dioxygenase 1 (Coffea canephora)

Id-gi/307592509/gb/ADN65332.1/carotenoid cleavage dioxygenase 1 (Manihot esculenta)

Id-gi/76560804/gb/ABA43904.1/carotenoid cleavage dioxygenase 1 (Coffea arabica)

Id-gi/61654494/gb/AAX48772.1/9,10[9',10']carotenoid cleavage dioxygenase (Vitis vinifera)

Id-gi/119486151/ref/ZP_01620211.1/Retinal pigment epithelial membrane protein (Lyngbya sp. PCC8106)

Id-gi/119511691/ref/ZP_01630796.1/Retinal pigment epithelial membrane protein (Nodularia spumigena CCY9414)

Id-gi/307154549/ref/YP_003889933.1/9-cis-epoxycarotenoid dioxygenase (Cyanotheca sp. PCC 7822)

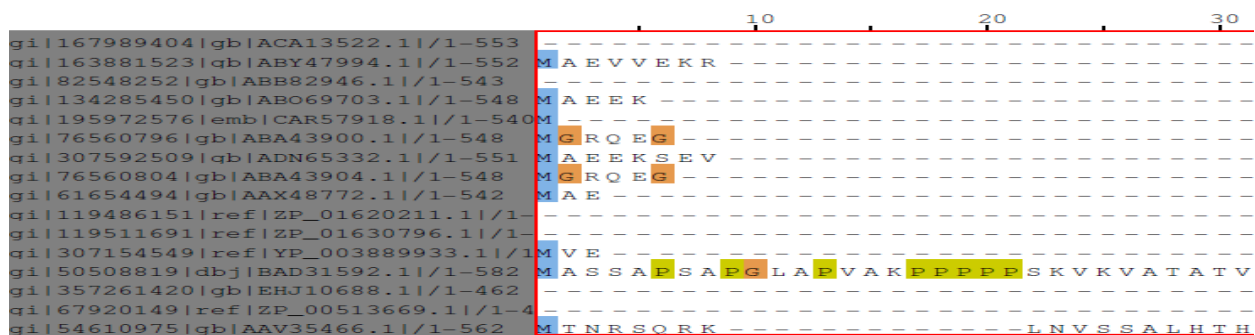
Id-gi/50508819/dbj/BAD31592.1/putative viviparous -14 protein (Oryza sativa Japonica group)

Id-gi/357261420/gb/EHJ10688.1/Retinal pigment epithelial membrane protein (Crocospaera watsonii WH0003)

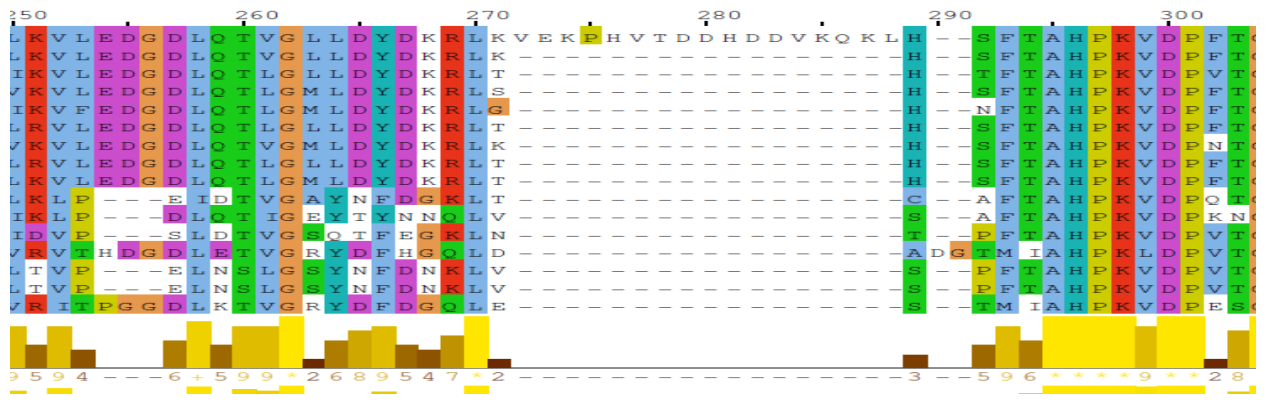
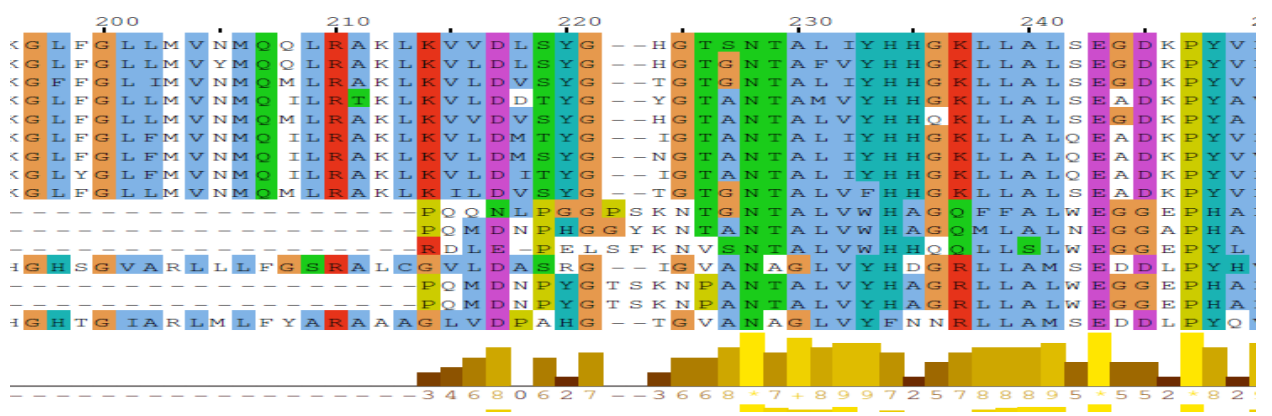
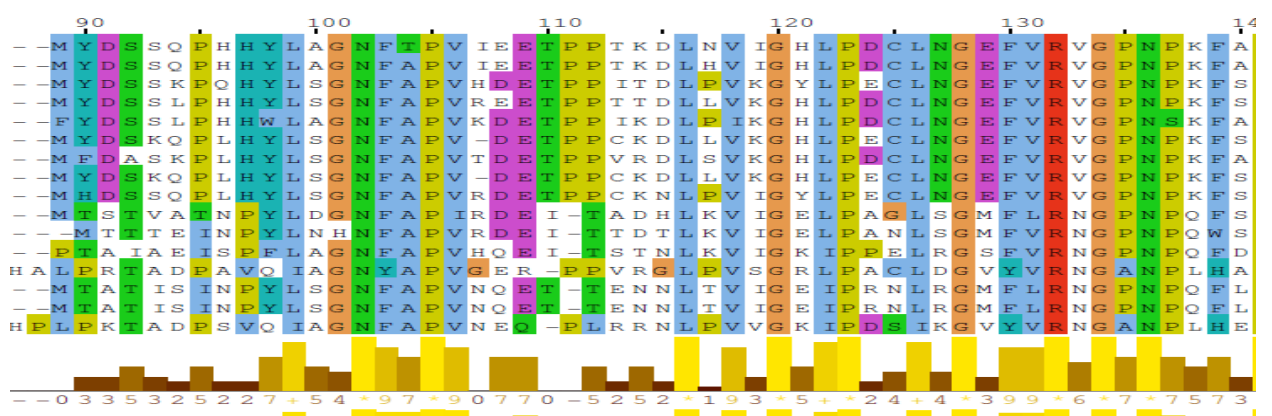
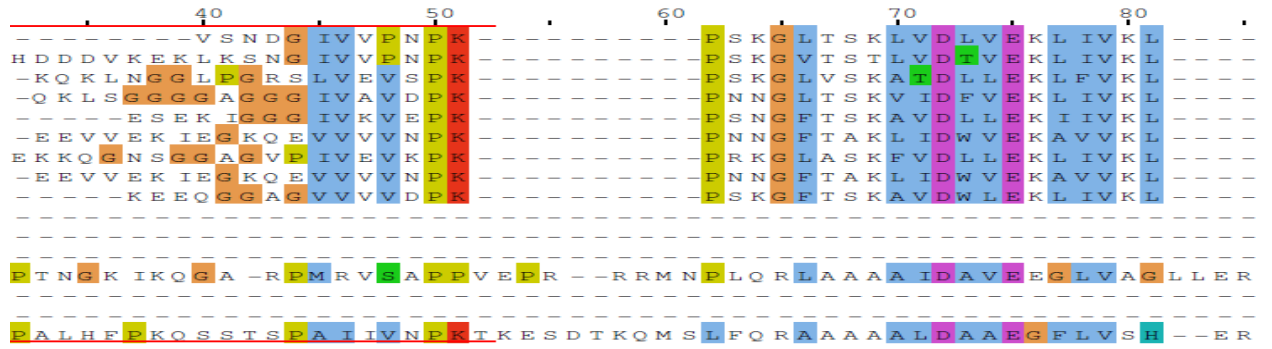
Id-gi/67920149/ref/ZP_00513669.1/Retinal pigment epithelial membrane protein (Crocospaera watsonii WH 8501)

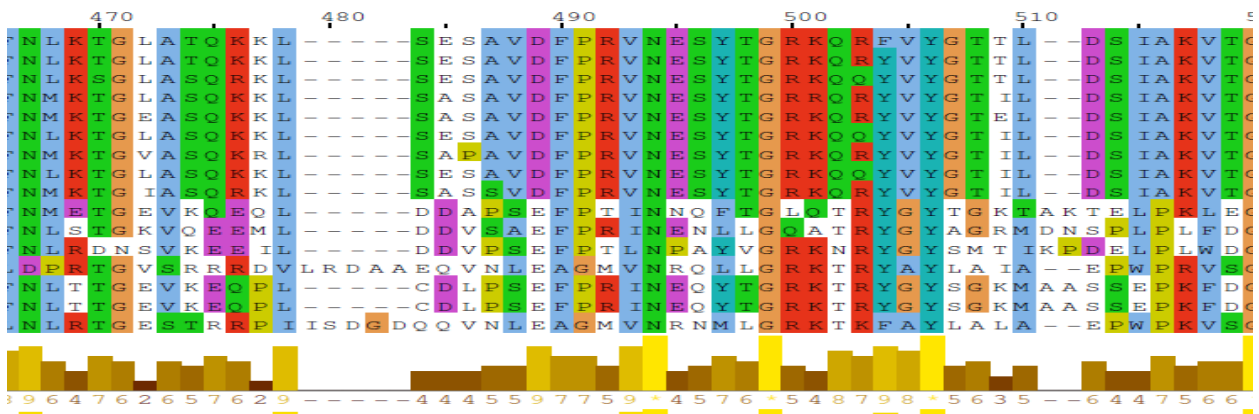
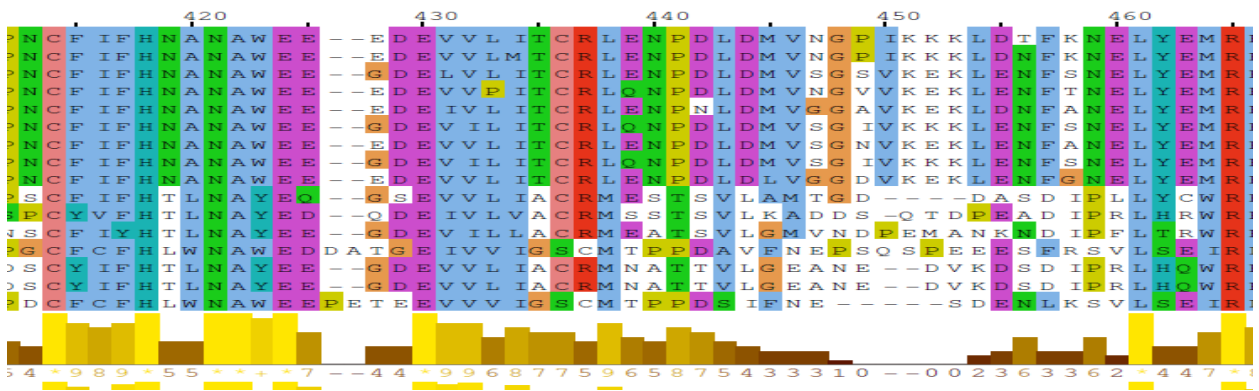
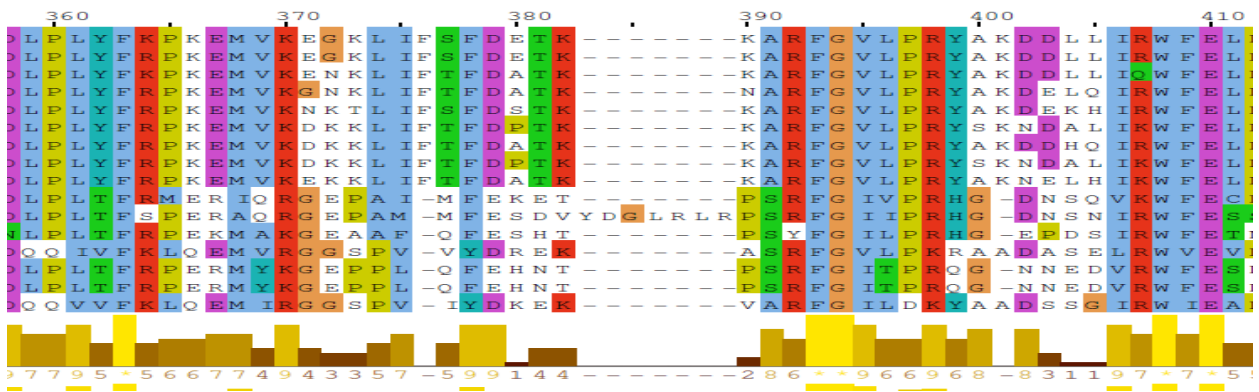
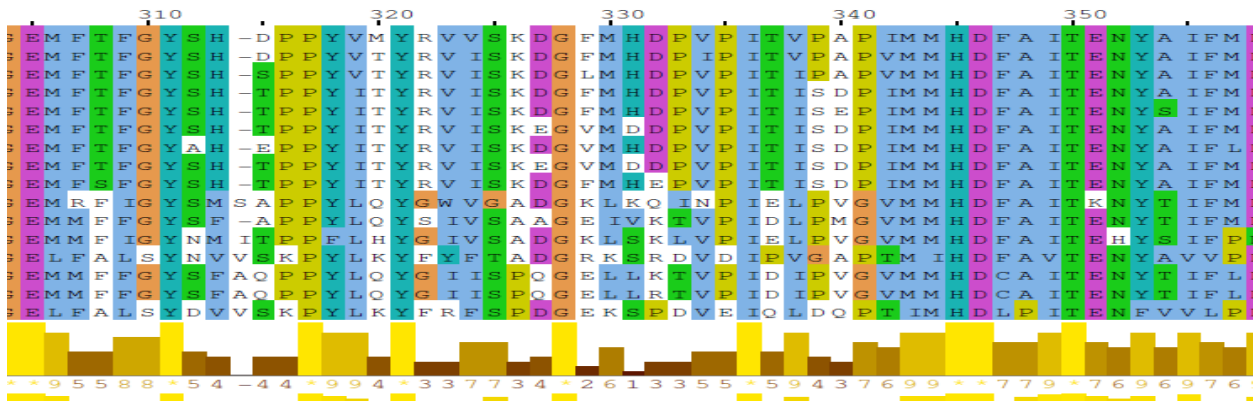
Id-gi/54610975/gb/AAV35466.1/9-cis-epoxycarotenoid dioxygenase (Brassica rapa subsp. Pekinensis)

2.2.2. Alignments: %age similarity



Conservatio:





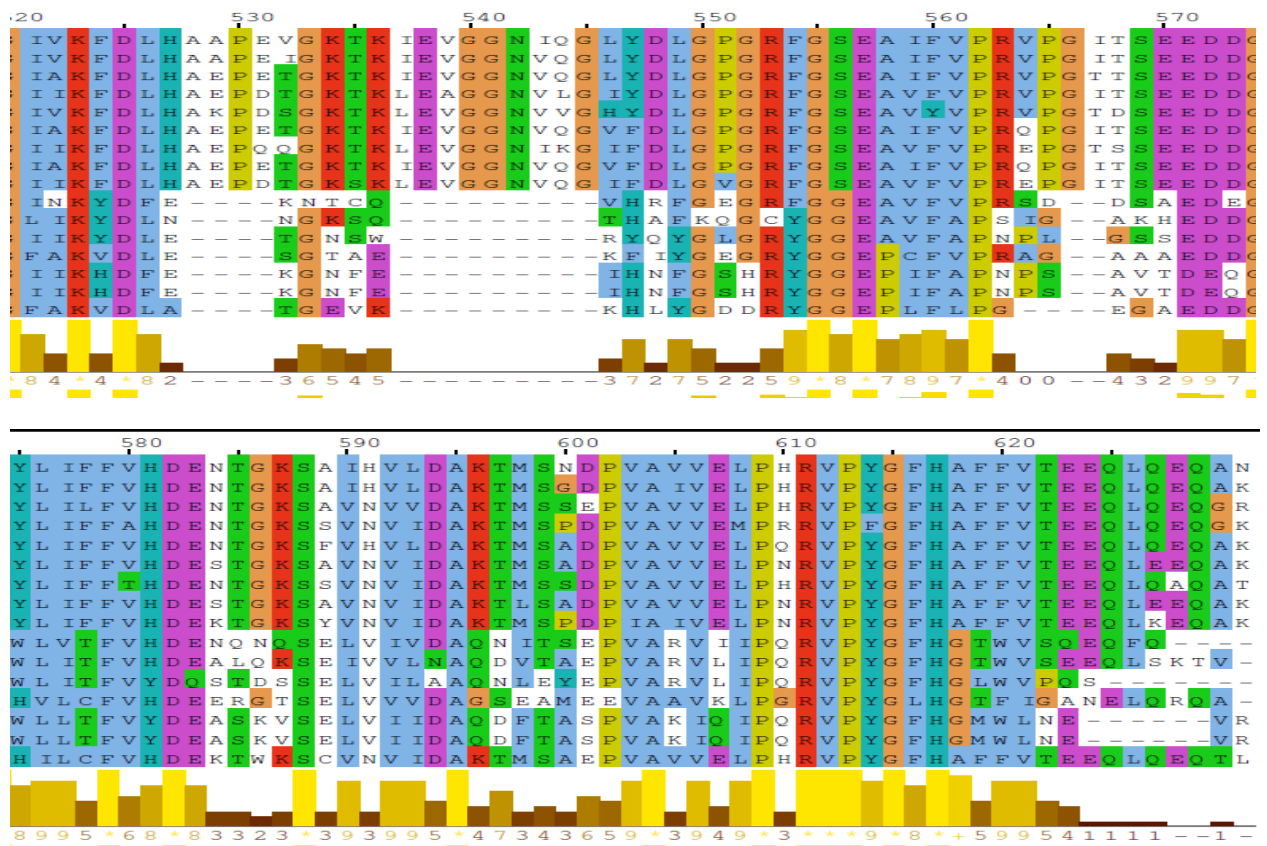


Fig. 2.1 Alignment of dioxygenases from different plant and bacteria

2.2.3. Protein BLAST result using PDB for the same protein (Dioxygenase [*Fragaria x ananassa*])

3NPE_A Chain A structure of VP14 in complex with oxygen

2BIW_A Crystal structure of apocarotenoid cleavage oxygenase

3FSN_A Crystal structure of Rpe65 at 2.14 angstrom resolution

3EOY_A The crystal structure a conserved domain from a protein

3P4G_A Chain A, X-ray crystal structure of a hyperactive, Ca²⁺ dependent kinase

2PFJ_A Crystal structure of T7 endo I resolvase

2Q6U_A Semet-substitute form of Nikd

3D40_A Crystal structure of fosfomycin resistance kinase

2OLN_A Chain A Nikd, An unusual amino acid oxidase essential for Nikkomycin biosynthesis

2EWE_A Chain A crystal structure of pectate lyase CR218k mutant

1PLU_A Chain A pectate lyase C from *Erwinia Chrysanthem*i with 1 Lu³⁺ ion

3HZL_A Tyr258phe mutant of Nikd, an unusual amino acid oxidase

2.2.4. Transit Peptides:

Further, it is assumed that the sequence must not contain any transit peptide amino acid sequence (chloroplast or mitochondria) and the protein must be cytoplasmic, and this was confirmed by analysis of the sequence. The sequence does not have any transit peptide amino acid sequence, and there are examples of proteins like squalene-hopene cyclase and lanosterol synthase, which lack an N-terminal signal peptide, thus, for being as an integral membrane protein the lack of potential transmembrane segments and/or an N-terminal signal peptide in the primary amino acid sequence of a protein cannot be the determining factor. Therefore, structural and functional information must be taken into account when membrane interactions of a protein are classified. After, removing first 41 residues, still there was no any transit peptide amino acid sequence. Amino acid sequences were also developed and adding those amino acids in place of removed amino acids, it acted like chloroplast and mitochondria targeting peptides.

2.2.5. Membrane binding property:

Complete sequence was analyzed for protein conserved domains:

2.2.5.1. Conserved domain analysis:

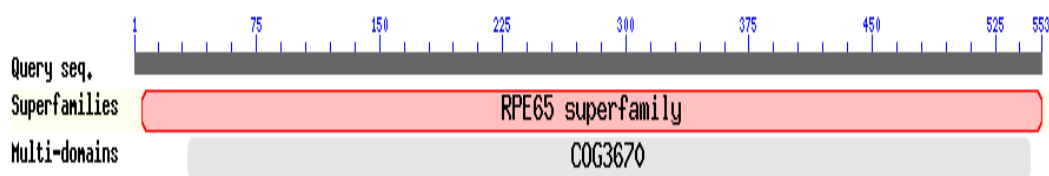


Fig. 2.2 Conserved Domain

In the whole sequence, two domains were found:

RPE65 super family (c110080) Retinal pigment epithelial membrane protein. This belongs to a family of retinal pigment epithelial membrane receptors.

COG3670 (COG3670) Lignostilbene-alpha, beta dioxygenase and related enzymes and secondary metabolites. It is a multidomain structure.

2.2.5.2. Description:

RPE65 superfamily [c110080]: Retinal pigment epithelial membrane protein; this family includes an epithelial membrane receptor for retinal pigment which is abundantly found in retinal pigment epithelium and complexes with plasma retinal binding protein. The family also includes the plant sequence related neoxanthin cleavage enzyme and bacterial lignostilbene-alpha, beta-dioxygenase.

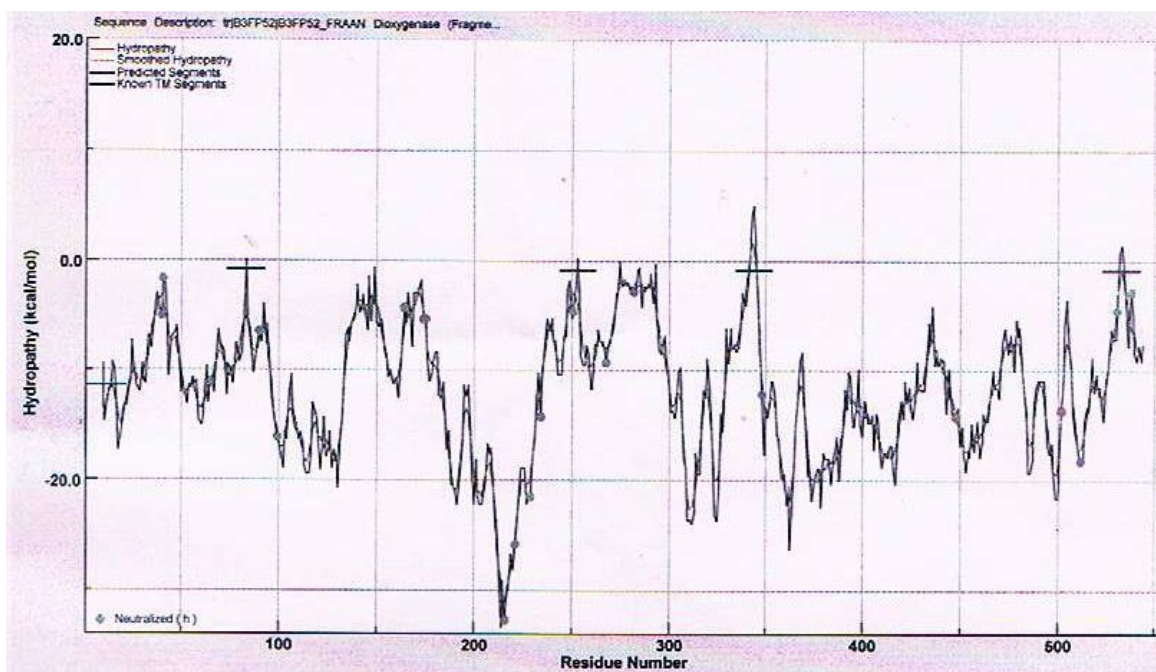
COG3670 [COG3670]: Lignostilbene-alpha, beta-dioxygenase and related enzymes [Biosynthesis of secondary metabolites, their transport and catabolism]

We know that, RPE 65 superfamily and related 9-cis-epoxycarotenoid dioxygenases are involved in Retinoid biosynthesis in retinal tissues (Kiser *et al.*, 2009) and carotenoid biosynthesis and metabolism (Messing *et al.*, 2010). Interestingly, some RPE65 relatives like 9-cisepoxycarotenoid dioxygenases, appears to be exclusively membrane bound, while β -carotene dioxygenases shows water solubility. Hydrophobic carotenoids which are not water soluble are the substrates for 9-cisepoxy dioxygenase and since the active sites are deeply located within the carotenoid oxygenases rather than superficially anchored near their surfaces the substrate must be physically displaced from the membrane for enzymatic activity. These enzymes require delivery of substrates in micelles or bilayer for their enzymatic processing, thus it is the possibility that the enzyme must interact with micelles or bilayers for substrate extraction regardless of the degree of their water solubility. To further investigate, this hypothesis that this might be a membrane bound protein, the hydrophobicity of the sequence and the probability of finding pores in the structure model were analyzed (Eyre, Tina. A. *et al.*, 2004). Hydrophobic pore residues on the surface surround the orifice of a large tunnel leading to the catalytic site of the enzyme (Messing *et al.*, 2010). In apocarotenoid dioxygenases and RPE65 the hydrophobic patches are variable allowing the enzyme to penetrate the membranes or micelles to extract carotenoid

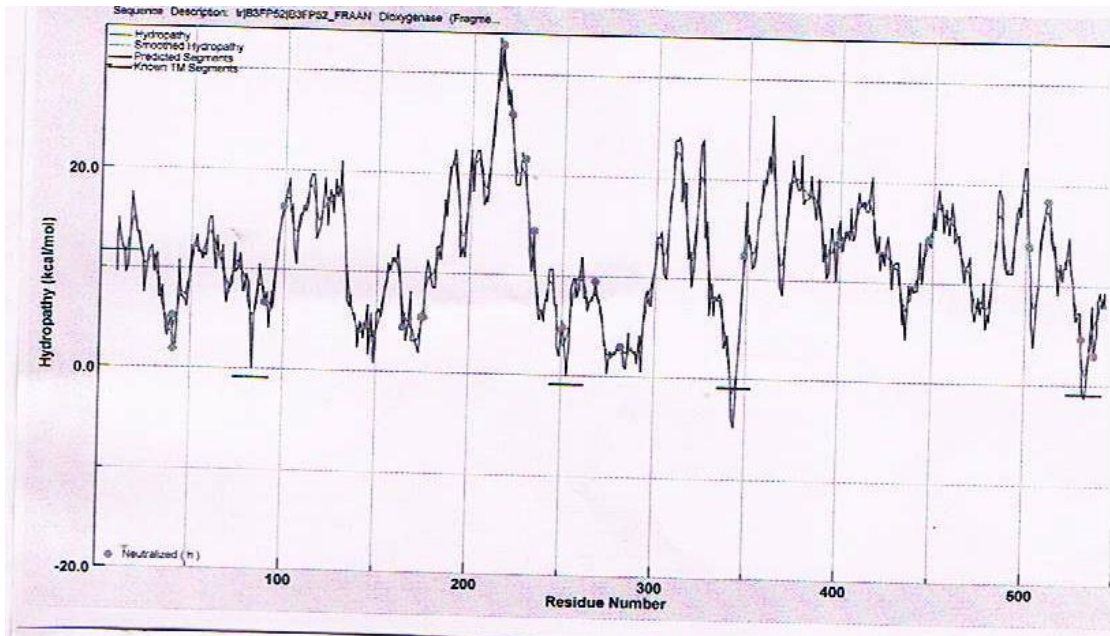
substrates; the comparison of structures indicates that RPE65 has more extensive hydrophobic patches than that of apocarotenoid dioxygenases.

2.2.6. Hydropathy plot:

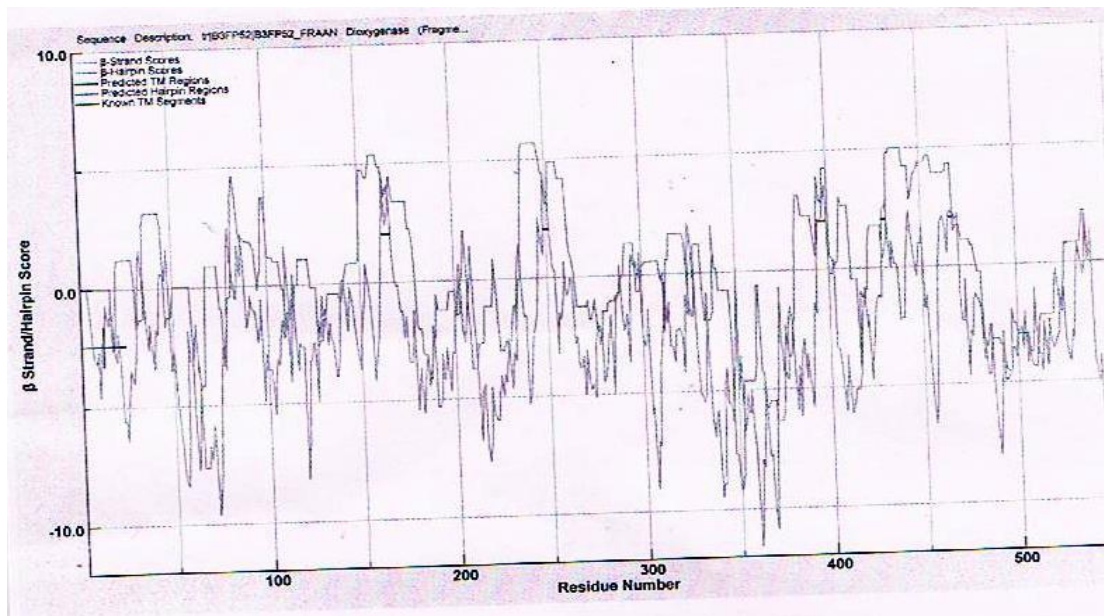
To visualize the hydrophobicity of the strawberry dioxygenase over complete peptide sequence length; hydropathy plots have been analyzed for that a hydropathy scale of all 20 amino acid residues based on their hydrophobic and hydrophilic properties are used. Such plots have significant role in determining membrane spanning segments of membrane bound proteins and the hydrophobic interior portions of globular proteins.



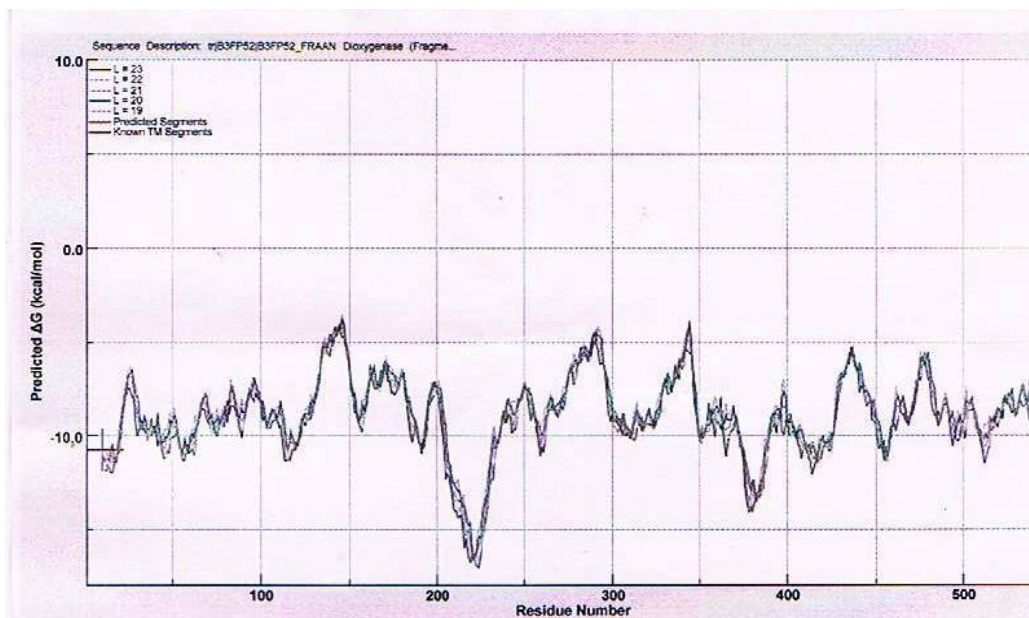
(a) Hydropathy Plot (Bilayer to Water Partitioning)



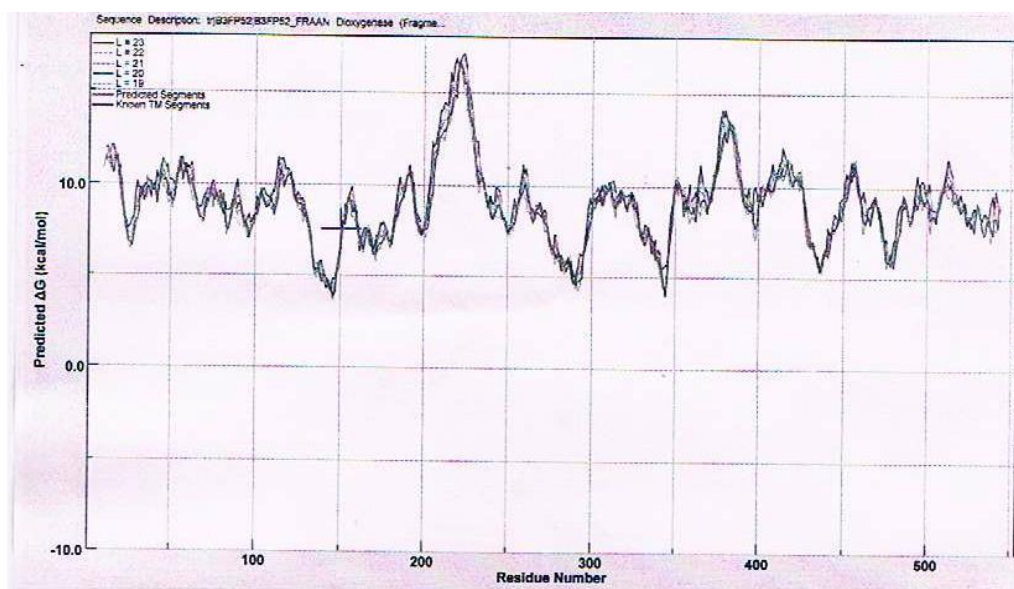
(b) Hydropathy Plot (Water to Bilayer Partitioning)



(c) β -strand/Hairpin (Bilayer to Water partitioning)



(d) Translocon segments (Bilayer to Water Partitioning)



(e) Translocon segments (Water to Bilayer Partitioning)

Fig. 2.3. Hydropathy Plots

In the hydropathy plot, most of the residues are in the negative side which indicates polar and aromatic amino acid residues. However, in both the partitioning system bilayer to water and water to bilayer, there are some hydropathy predicted segments, which are also transmembrane segments. Translocon TM analysis shows, the following segments:

(1) FVRVGNPKFAPVAGYhWF

(2) MFTFGYShDPPYVMYRVVS

(3) LIRWFELPNCFIFhNANAW

(4) VAVVELPhRVPYGFhAFFV

Moreover, there are 14-15 membranes spanning segments or region, which in this case, are mostly β -barrels. Therefore, it is assumed that the sequence belongs to a transmembrane protein, but to confirm whether, it is monotopic or integral, the sequence was further investigated.

2.2.7. β -barrel outer membrane protein analysis:

2.2.7.1. Posterior probability plot (sequence):

Table 2.1 Posterior probability

<i>Posterior decoding method</i>												
	1		2		3		4		5		6	
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	
0000	V	S	N	D	G	I	V	V	P	N	P	K
0060	N	V	I	G	H	L	P	D	C	L	N	G
0120	K	Q	E	E	F	G	G	A	F	M	K	I
0180	L	S	E	G	D	K	P	Y	V	L	K	V
0240	T	G	E	M	F	T	F	G	Y	S	H	D
0300	K	P	K	E	M	V	K	E	G	L	I	F
0360	L	I	T	C	R	L	E	N	P	D	L	M
0420	T	G	R	K	Q	R	F	V	G	T	T	L
0480	F	V	P	R	V	P	G	I	T	S	E	E
0540	F	V	T	E	E	Q	L	Q	E	Q	A	N

Table 2.2 Membrane Inside and Outside Residue Position

in	1	40	tm	204	212	out	354	375			
tm	41	47	out	213	230	tm	376	384			
out	48	89	tm	231	237	in	385	389			
tm	90	102	in	238	244	tm	390	400			
in	103	108	tm	245	251	out	401	423			
tm	109	117	out	252	289	tm	424	433			
									tm	496	502

out	118	136	tm	290	300	in	434	436	in	503	553
tm	137	146	in	301	306	tm	437	447			
in	147	148	tm	307	315	out	448	457			
tm	149	157	out	316	332	tm	458	468			
out	158	187	tm	333	341	in	469	474			
tm	188	200	in	342	344	tm	475	481			
in	201	203	tm	345	353	out	482	495			

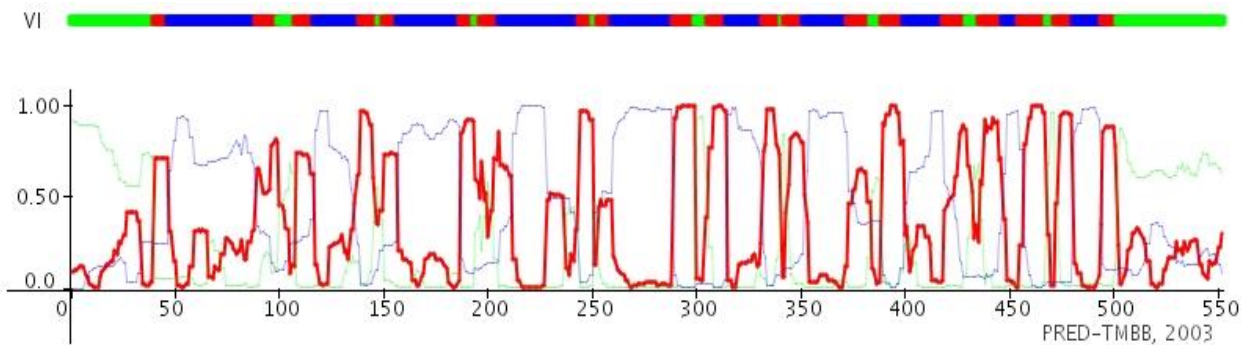


Fig. 2.4. Posterior Probability Plot

2D representation (sequence):

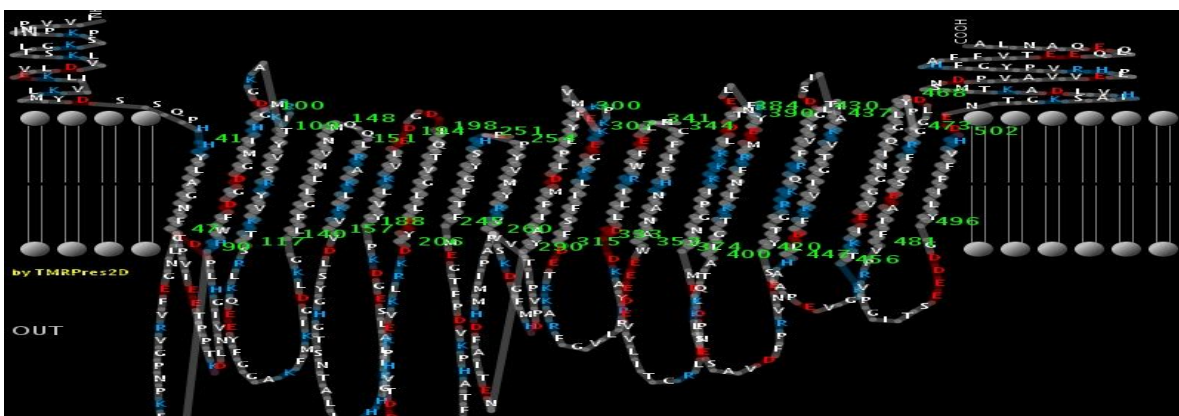


Fig. 2.5. Strawberry Dioxygenase sequence 2D representation in the membrane

The posterior probability plot and electrostatic 2D representation of the dioxygenase sequence show its membrane bound property and it also indicates transmembrane residues (shown in red), Inside residues (shown in green) and outside residues (shown in blue) in tables 2.1 & 2.2.

2.2.8. Folds and Motifs:

2.2.8.1. NAD-FAD Rossmann fold motif:

On NAD-FAD Rossmann fold motif analysis, no such motifs in the given amino acid sequence were found. Consequently, it is unlikely that the protein adopts both Rossmann fold and binds NAD and FAD. However, the protein may be a Rossmann fold that does not bind NADP and FAD. Furthermore, the protein may adopt a fold other than Rossmann fold and bind FAD and NADP. Here, it is pretty certain, that the protein binds NADP and FAD.

2.2.8.2. Motif analysis:


Table 2.3: Strawberry Dioxygenase sequence motifs and possible post translational modifications

Residue position		Modification type
Start	End	
421	424	AMIDATION
417	420	ASN GLYCOSYLATION
449	456	ATP GTP A
404	407	CAMP PHOSPHO SITE
182	185	CK2 PHOSPHO SITE
314	317	CK2 PHOSPHO SITE
409	412	CK2 PHOSPHO SITE
431	434	CK2 PHOSPHO SITE
456	459	CK2 PHOSPHO SITE
489	492	CK2 PHOSPHO SITE
96	101	MYRISTYL
165	170	MYRISTYL

462	467	MYRISTYL
18	20	PKC PHOSPHO SITE
117	119	PKC PHOSPHO SITE
318	320	PKC PHOSPHO SITE
363	365	PKC PHOSPHO SITE
384	386	PKC PHOSPHO SITE
402	404	PKC PHOSPHO SITE
421	423	PKC PHOSPHO SITE
506	508	PKC PHOSPHO SITE
156	162	TYR PHOSPHO SITE
148	179	PIPLC X DOMAIN
30	544	RPE65
30	544	RPE65

The sequence was analyzed for motifs; the search indicates certain amidation, glycosylation and phosphorylation sites along with the RPE65 superfamily motif which also indicates the probable modifications of specific residues at specific sites by certain specific kinases; the structural and functional roles and the importance of the residues involved needs further investigation by incorporating different substrates in the structural model and by the study of their interaction pattern.

Table 2.4: Post translational modification sites

	pos.: 421-424	AMIDATION <i>Amidation site.</i> Legends: 1 amidation.
---	----------------------	---

<pre> 1 Y W V T S R R Q N M L L K K I H H G G F E D D C T C Y W V T S R R Q N M L L K K I H H G G F E D D C T C NASA I::: NESY </pre>	<p>pos.: 417-420</p>	<p>ASN_GLYCOSYLATION <i>N-glycosylation site.</i> Legends: 1 carbohydrate.</p>
---	---------------------------------	---

<pre> G T A:::XGKS : : I I: AAPEVGKT </pre>	<p>pos.: 449-456</p>	<p>ATP GTP A <i>ATP/GTP-binding site motif A (P-loop).</i></p>
---	-----------------------------	--

<pre> 1 KK T RRXS ::: KKLS </pre>	<p>pos.: 404-407</p>	<p>CAMP PHOSPHO SITE <i>cAMP- and cGMP-dependent protein kinase phosphorylation site.</i> Legends: 1 phosphorylation.</p>
-----------------------------------	-----------------------------	--

<pre> 1 T E SXXD : : SEGD </pre>	<p>pos: 182-185</p>	<p>CK2 PHOSPHO SITE <i>Casein kinase II phosphorylation site.</i> Legends: 1 phosphorylation.</p>
<pre> 1 T E SXXD : : SFDE </pre>	<p>pos.: 314-317</p>	
<pre> 1 T E SXXD : : SAVD </pre>	<p>pos.: 409-412</p>	
<pre> 1 T E SXXD : : TTLD </pre>	<p>pos.: 431-434</p>	

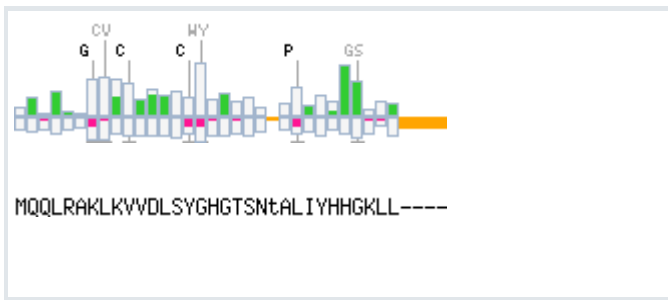
<pre> 1 T E \$XXD : TKIE </pre>	<p>pos.: 456-459</p>
<pre> 1 T E \$XXD : TSEE </pre>	<p>pos.: 489-492</p>

<pre> 1 Y W V T S R R Q N M L K I H V T S Q N M L I G C GAXXSA I: :: GMIHGM </pre>	<p>pos.: 96-101</p>	
<pre> 1 Y W V T S R R Q N M L K I H V T S Q N M L I G C GAXXSA I: :: GTSNTA </pre>	<p>pos.: 165-170</p>	<p>MYRISTYL <i>N-myristoylation site.</i> Legends: 1, myristyl.</p>
<pre> 1 Y W V T S R R Q N M L K I H V T S Q N M L I G C GAXXSA I: :: GNIQGL </pre>	<p>pos.: 462-467</p>	

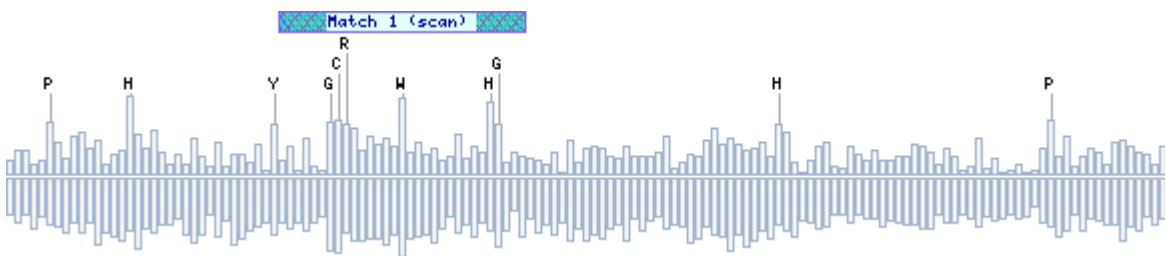
<pre> 1 T K SXR : : : : TSK </pre>	pos.: 18-20	<p>PKC PHOSPHO SITE <i>Protein kinase C phosphorylation site.</i> Legends: 1, phosphorylation.</p>
<pre> 1 T K SXR : : : : TSR </pre>	pos.: 117-119	
<pre> 1 T K SXR : : : : TKK </pre>	pos.: 318-320	
<pre> 1 T K SXR : : : : TCR </pre>	pos.: 363-365	
<pre> 1 T K SXR : : : : TFK </pre>	pos.: 384-386	
<pre> 1 T K SXR : : : : TQK </pre>	pos.: 402-404	
<pre> 1 T K SXR : : : : TGR </pre>	pos.: 421-423	
<pre> 1 T K SXR : : : : TGK </pre>	pos.: 506-508	
<pre> 1 K E RXXDXXY : : : : KVVQLSY </pre>	pos.: 156-162	<p>TYR PHOSPHO SITE <i>Tyrosine kinase phosphorylation site.</i> Legends: 1, phosphorylation.</p>

The abovementioned amino acid residues are amidated, glycosylated and phosphorylated by specific kinases and other proteins, some of them are modified already in the natural conditions or being modified after substrate binding as for the activation of specific kinases some of the residues in the kinase proteins must be modified by other environmental effects.

Table 2.5: Motifs

 <p>MQQLRAKLVVWDSYGHGTSNTALIIYHHGKLL----</p>	<p>pos.: 148-179 raw-score = 357 N-score = 6.811 E-value = 3.3</p>	<p>PIPLC X DOMAIN <i>Phosphatidylinositol-specific phospholipase X-box domain profile.</i></p>
---	---	---

ID PIPLC X DOMAIN; MATRIX.
DE Phosphatidylinositol-specific phospholipase X-box domain profile.
CC The scoring system depicted below is approximate.



ID RAW TEXT TEMPORARY; PRT; 553 AA.

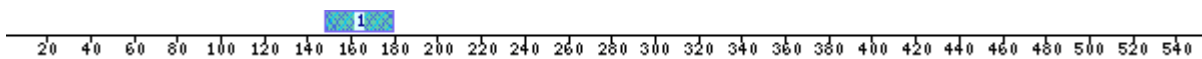
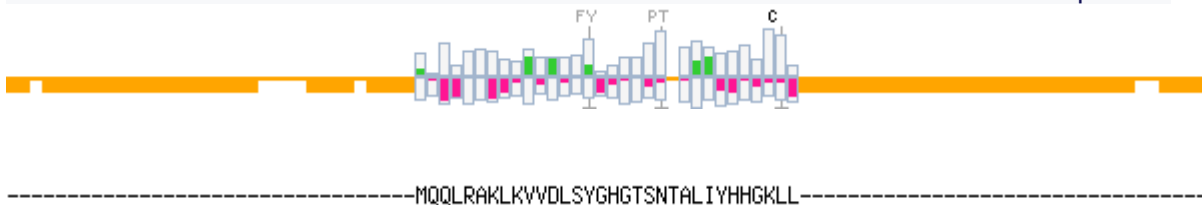


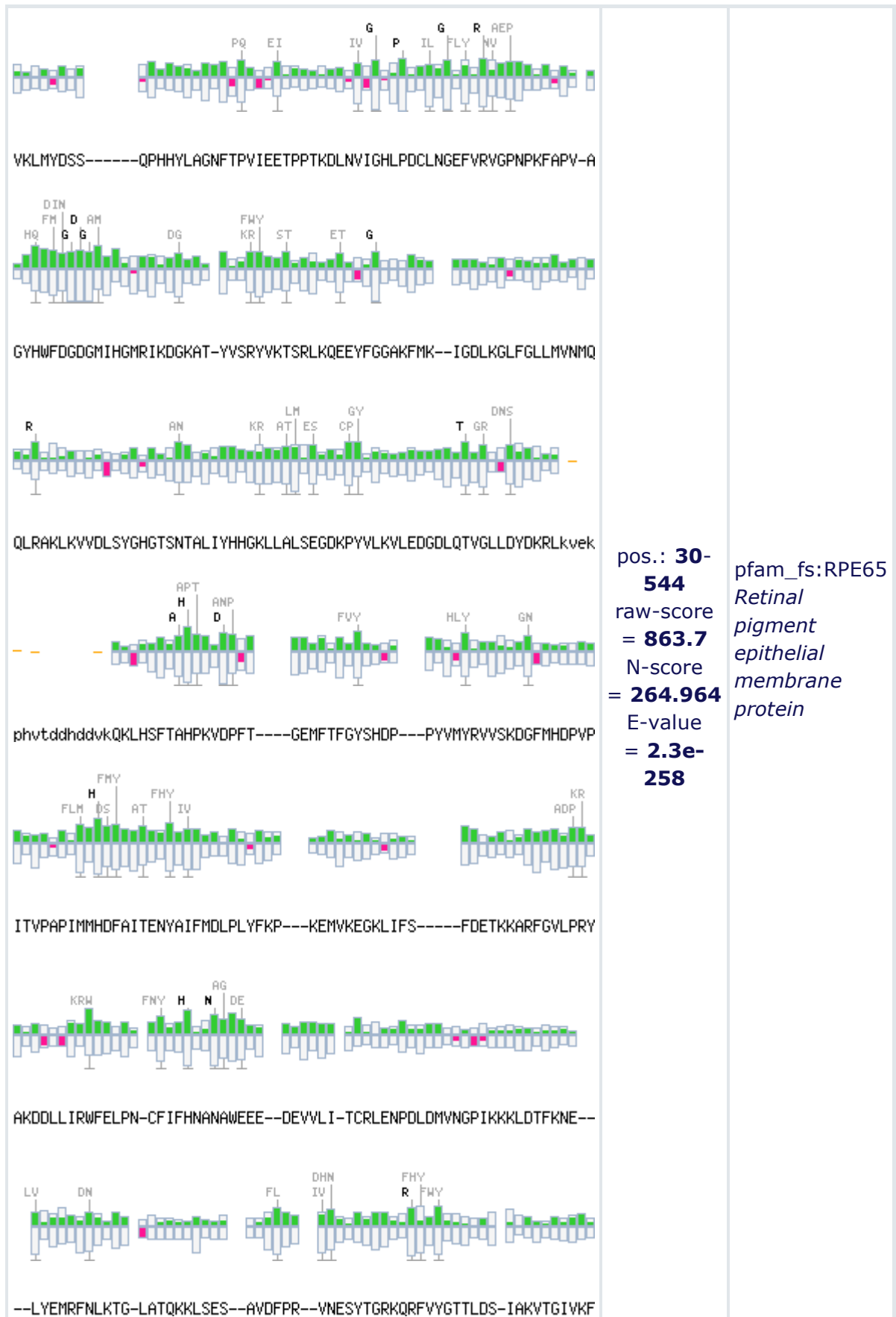
Fig. 2.6. Dioxygenase Motif scan

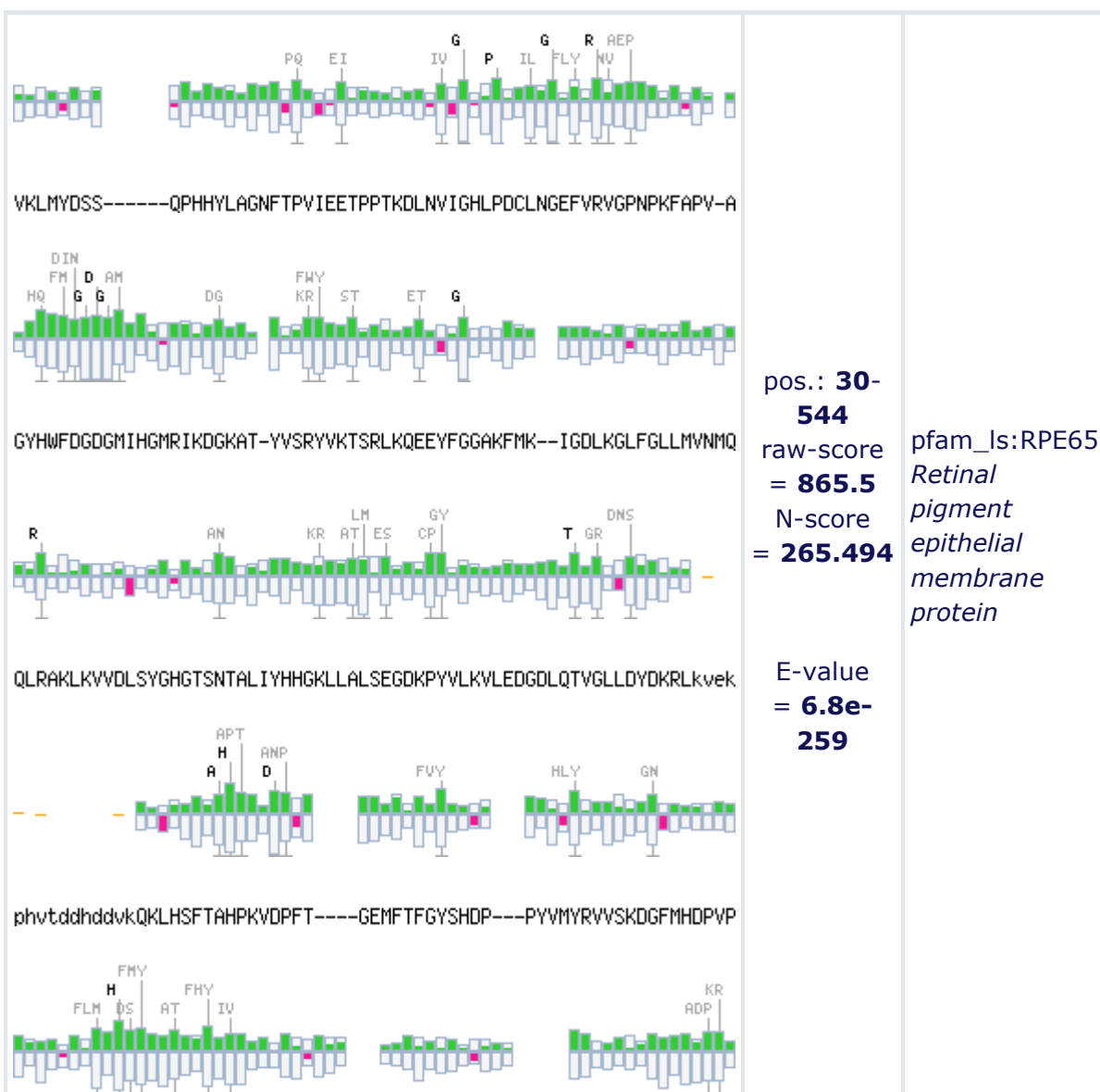
```
>prf:PIPLC_X_DOMAIN 6.811 357 pos. 148 - 179 [35, -81]
PS50007|prf:PIPLC_X_DOMAIN Phosphatidylinositol-specific phospholipase X-box domain profile.
```

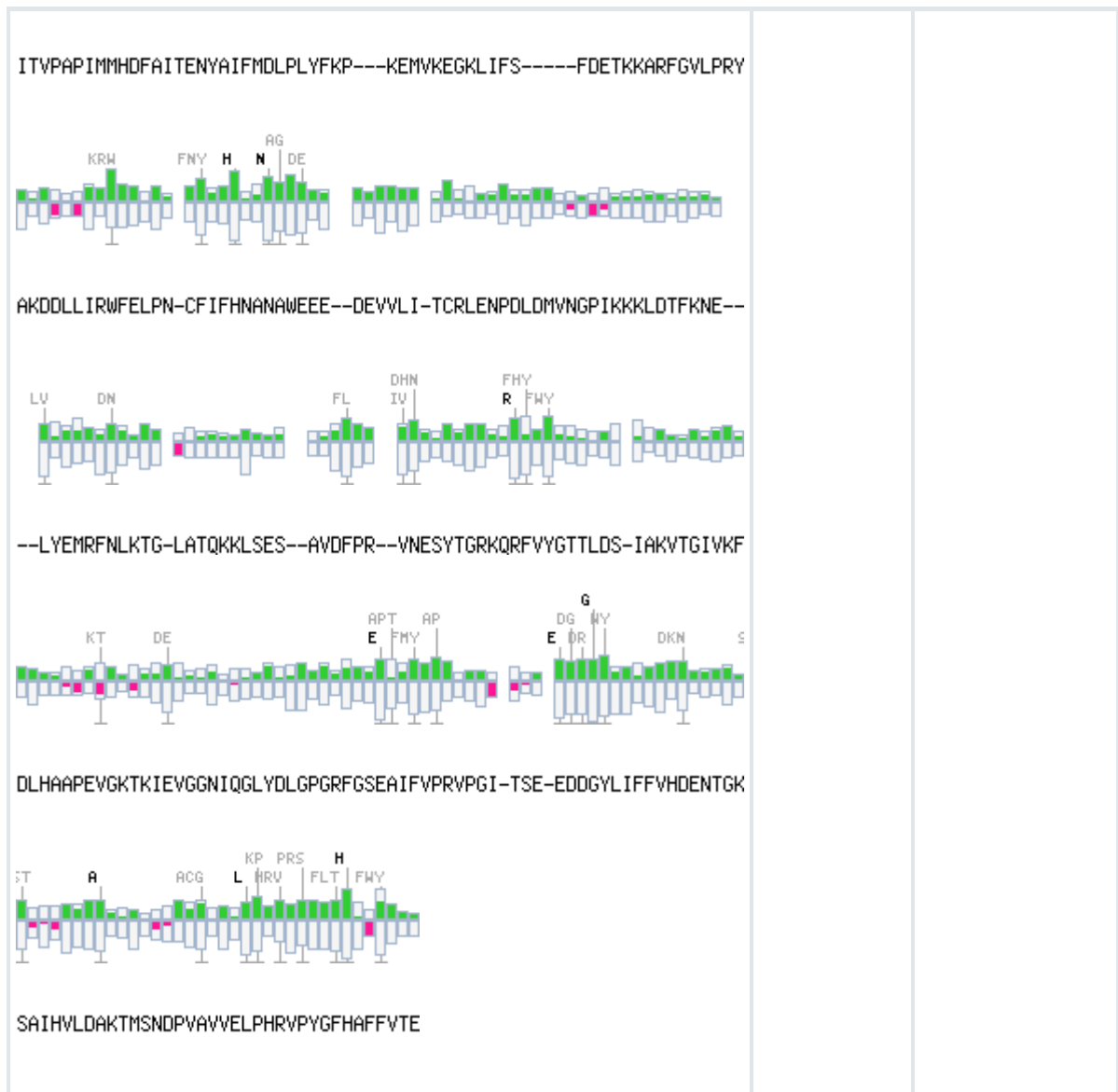


Legends: 1, Match 1 (scan).

Table 2.6: Motif







The known motif RPE65 superfamily belongs to the group of enzymes involved in retinol biosynthesis in the retinal endothelium, also the BLAST analysis and multiple sequence analysis indicated such motifs to be present in the sequence, which is conserved among certain species and the sequence and structural similarity is well beyond 70-90% or almost identical.

These motifs will enhance the implication of the substrate binding as depending on these results it is possible to explore the role of dioxygenases in conditions which are very supportive of occurring of such modifications.

2.2.9. Glycosylation sites:

2.2.9.1 NetCGlyc Prediction

Table 2.7: Net C Glycosylation sites:

Feature	Start	Stop
C-manno	91	91
C-manno	338	338
C-manno	353	353

C-mannose glycosylation sites have been predicted that are shown in table 2.7, and the corresponding residues where this modification is most possible are at positions 91, 338 and 353.

2.2.9.2. NetNGlyc Prediction

Table: 2.8 Net N-Glycosylation sites

Position	Potential
46	NFTP
417	NESY

N-glycosylation sites have been predicted that are shown in table 2.8; the corresponding residues at positions 46 and 417 are also mentioned.

2.2.9.3. NetOGlyc Prediction:

Table 2.9: Net O Glycosylation sites

S/T	Position	G-score	I-score	Y/N	Comment
S	2	0.450	0.068	.	-
S	14	0.206	0.106	.	-
T	18	0.222	0.076	.	-
S	19	0.153	0.069	.	-
S	36	0.110	0.046	.	-
S	37	0.122	0.027	.	-
T	48	0.368	0.077	.	-
T	54	0.328	0.114	.	-
T	57	0.278	0.494	.	-
T	109	0.172	0.041	.	-
S	112	0.122	0.075	.	-
T	117	0.174	0.054	.	-
S	118	0.111	0.034	.	-
S	161	0.103	0.061	.	-
T	166	0.134	0.032	.	-
S	167	0.093	0.025	.	-
T	169	0.139	0.031	.	-
S	182	0.093	0.064	.	-
T	200	0.131	0.090	.	-
T	218	0.188	0.247	.	-
S	230	0.229	0.052	.	-
T	232	0.298	0.049	.	-
T	241	0.330	0.078	.	-
T	246	0.336	0.074	.	-
S	250	0.185	0.052	.	-

S	262	0.252	0.055	.	-
T	274	0.325	0.681	T	-
T	287	0.236	0.064	.	-
S	314	0.109	0.051	.	-
T	318	0.147	0.035	.	-
T	363	0.199	0.079	.	-
T	384	0.133	0.051	.	-
T	398	0.157	0.034	.	-
T	402	0.173	0.054	.	-
S	407	0.168	0.063	.	-
S	409	0.161	0.019	.	-
S	419	0.157	0.020	.	-
T	421	0.267	0.059	.	-
T	431	0.209	0.055	.	-
T	432	0.192	0.025	.	-
S	435	0.126	0.030	.	-
T	440	0.219	0.076	.	-
T	456	0.183	0.094	.	-
S	477	0.200	0.072	.	-
T	489	0.247	0.096	.	-
S	490	0.157	0.080	.	-
T	506	0.162	0.028	.	-
S	509	0.127	0.018	.	-
T	518	0.291	0.043	.	-
S	520	0.221	0.077	.	-
T	543	0.245	0.054	.	-

The O-glycosylation sites also have been predicted and the corresponding residue Threonine (T) at the position T274 (shown in brown), the table 2.9 also indicates that other residues can be modified as well.

2.2.9.4. Phosphorylation sites:

Table: 2.10 Net Phosphorylation sites and Specific Kinases

Position	Residue	Score	Kinase
2	S	0.45	CaM-II
14	S	0.5	PKA
18	T	0.58	PKC
19	S	0.44	CaM-II
36	S	0.47	GSK3
37	S	0.65	DNAPK
48	T	0.48	GSK3
54	T	0.52	CKI
57	T	0.46	GSK3
109	T	0.47	PKC
112	S	0.55	PKC
117	T	0.89	PKC
118	S	0.59	PKC
161	S	0.55	PKA
166	T	0.44	GSK3
167	S	0.49	PKA
169	T	0.52	cdc2
182	S	0.49	CKI
200	T	0.47	CKII
218	T	0.62	CKII
230	S	0.85	PKC

232	T	0.48	PKG
241	T	0.45	GSK3
246	T	0.57	PKC
250	S	0.56	cdc2
262	S	0.5	PKC
274	T	0.48	PKC
287	T	0.5	CKII
314	S	0.64	CKII
318	T	0.89	PKC
363	T	0.47	PKC
384	T	0.82	PKC
398	T	0.64	PKC
402	T	0.88	PKC
407	S	0.57	PKA
409	S	0.45	PKG
419	S	0.51	PKC
421	T	0.56	PKC
431	T	0.5	PKC
432	T	0.45	CaM-II
435	S	0.47	PKC
440	T	0.81	PKC
456	T	0.56	PKG
477	S	0.75	PKA
489	T	0.64	CKII
490	S	0.61	CKII
506	T	0.67	PKC
509	S	0.45	CaM-II
518	T	0.45	GSK3

520	S	0.55	PKA
543	T	0.57	CKII
34	Y	0.48	
42	Y	0.39	INSR
89	Y	0.47	
110	Y	0.48	INSR
114	Y	0.59	
125	Y	0.39	SRC
162	Y	0.44	INSR
173	Y	0.42	INSR
188	Y	0.44	INSR
206	Y	0.46	INSR
249	Y	0.45	INSR
255	Y	0.47	SRC
258	Y	0.41	INSR
290	Y	0.39	INSR
299	Y	0.42	SRC
329	Y	0.43	INSR
390	Y	0.47	INSR
420	Y	0.41	INSR
429	Y	0.44	INSR
468	Y	0.41	INSR
496	Y	0.59	
535	Y	0.4	

Specific phosphorylation sites have been predicted and depicted in the table 2.10, which indicates the corresponding residues at specific positions that presumably would undergo phosphorylation by abovementioned specific kinases. The most probable residues for

phosphorylation are Serine, Threonine and Tyrosine, which indicates that the dioxygenase protein belongs to the phosphorylated tyrosine family of proteins (Wu, Hsin-Yi. *et al.* 2010; Li, Haiyu. *et al.* 2009), which implicates in several events under oxidative stress conditions or conditions which mark as low ATP levels such as herbicide degradation, hypoxia and tumor cell proliferation. This also indicates that dioxygenase might belong to the family of α -ketoglutarate dependent dioxygenases, where self hydroxylation reaction mechanism is a common recovery mechanism for enzyme activity under such stress conditions.

2.2.10. Substrate Binding domains:

The whole sequence of the enzyme was 556 amino acid residues.

Then, the whole sequence had been divided into five parts, comprising of 110 residues,

VSN DGIVVPNPKPSKGLT SKLVDLVEKLIVKLMYDSSQPHHYLAGNFTPVIEETPP

TKDLNVIGHLPDCLNGEFVRVGP NPKFAPVAGYHWFDGDGMIHGMR IKDGKAT (1-110)

TYVSRYVKTSRLKQEEYFGGAKFMKIGDLKGLFGLLMVNMQQLRAKLKVV DLSY

GHGTSNTALIYHHGKLLALSEGDKPYVLKVLEDGDLQTVGLLDYDKRLKVEKPHVT (110-220)

TDDHDDVKQKLHSFTAHPKVD PFTGEMFTFGYSHDPPYVMYRVVSKDGF MHD

PVPITVPAPIMMHDFAITENYAIFMDLPLYFKPKEMVKEGKLIFSDETKKARFGVL (220-330)

PRYAKDDLIRWFELPNC FIFHNANAWEEDEEV LITCRLENPDLD MVNGPIK KK

LDTFKNELYEMR FNLTGLATQKKLSESAVDFPRVNESY TGRKQ (330-440)

RFVYGTTLDSIAKVTGIVKFDLHAAPEVGKTKIEVGGNIQGLYDLGPGRFGSE AIFVPRVPGITS

EEDDGYLIFVHDENTGKSAIHVLD AKTMSNDPVAVVELPHRV PYGFHAFVTEEQ LQE QANL (440-560)

All the five sequences were analyzed for batch conserved domains; again the two conserved domains were found that are common i.e. RPE65 superfamily and COG3670

multidomain. Further, all the five sequences were analyzed for model generation, separately. That resulted in different binding domains in each of the sequences.

Batch conserved domain:

RPE65 superfamily

COG3670 multidomain

Seq1-5 RPE65 superfamily

Possible dioxygenase and other protein interactions:

Sq1 common domain, Transport protein, cell adhesion, lipid binding, chorismate mutase, plant protein seed storage, Immunoglobulin like, allergens

Sq2 common domain, Transferase glutamine, phosphoryl transferases, anion transport, anticodon binding protein

Sq3 common domain, Transferase, arylsulphate sulfotransferase, dna binding protein, protein binding kinases, ser/thr kinase, hydrolases, Actin like ATPase, 7 bladed β -propeller, muconate lactonizing enzyme, diphtheria protein, TATA box binding protein, Ribonuclease H like, ubiquitin conjugatine enzyme E2 variant 1

Sq4 common domain, Electron transport, Transferase, magnesium dependent aromatic feny transferase, ser/thr kinase, metallo- β -lactamase

Sq5 common domain, cd1 nitrite reductase, 7- bladed β -propeller, YVTN repeats quinine protein amine dehydrogenase, like domain, 8-bladed β -propeller, nitrite reductase, dicer like protein domain, metallodependent hydrolases, motor protein, metal binding protein, peptide binding protein, calnexin, nitrous oxide reductase, myosin ATPase, Immunoglobulin protein domain.

2.2.11. Inference:

>The sequence has a peptide binding domain in the region from amino acid seq. 436- 560.

> ABA binding could be possible in the region from amino acid seq 1-225

>Active site could be in the region from amino acid seq 220-560, which is confirmed by analyzing the template active site:

(His234, His282, His348, His538) 3npe

(Phe 413, His282, His348, His538) 2biw

>The sequence (first 36 a.a. residues) has a peptide sequence with similarity to allergens like glutenin, gliadin as found in wheat and other allergens as found in animals. Therefore, the dioxygenases must be playing various important roles in strawberry plant defense and at vascular-endothelial and peripheral immune interfaces in mammals.

> The residues which are glycosylated or phosphorylated by specific kinases and proteins are primarily serine, threonine, arginine and tyrosine. As histidine and tyrosine phosphorylation are very common events during oxidative stress and tumor growth, therefore, it can be assumed that dioxygenases might be playing some important roles in tissue morphogenesis and cell proliferation and in facilitating peripheral immune response guided by Helper T-cells (T_h), regulatory T cells (T_{reg}) and dendritic cells. The presence of branched chain amino acid residues leucine, isoleucine and valine indicates the role, dioxygenases presumably play in the innate immunity, which is also clear by the presence of glutamine, glutamate and asparagines and proline in close proximity of such residues. There are certain cysteine residues at specific positions which are probably modified under oxidative conditions by redox molecules such as superoxide radicals, hydroxyl ions and hydrogen peroxide *per se*.

2.2.12. Instability index:

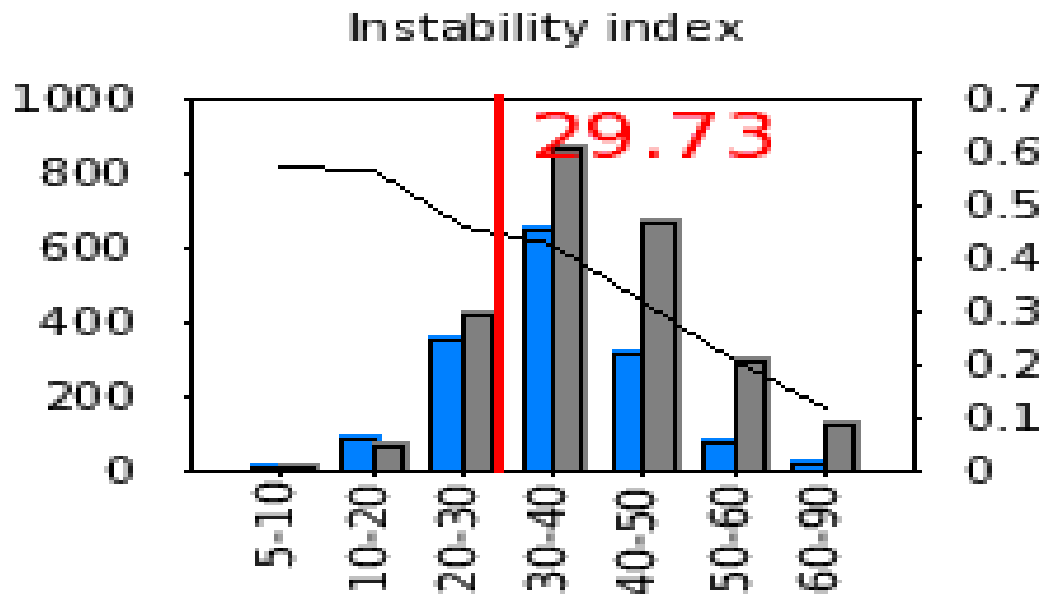
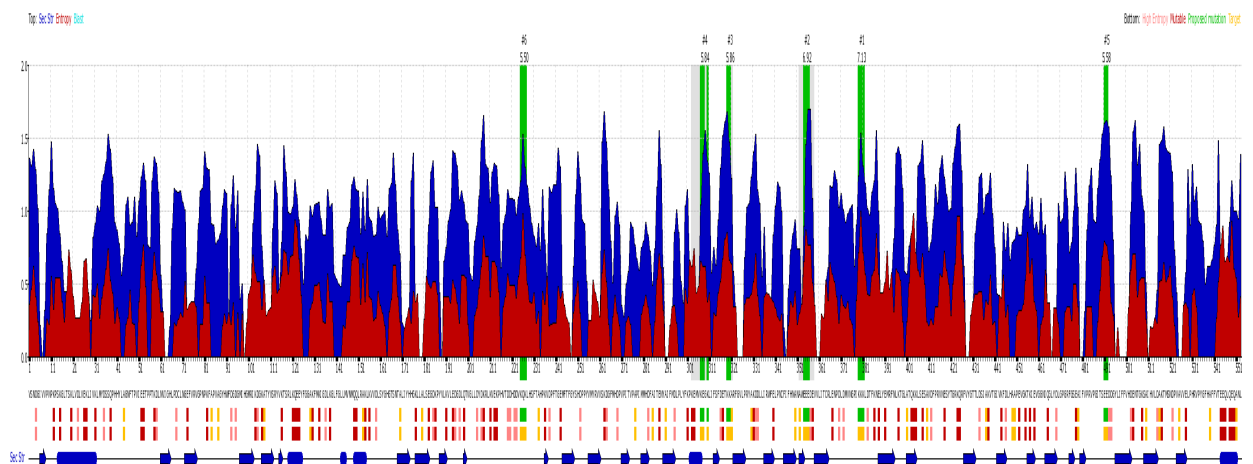


Fig. 2.7. Dioxygenase protein Instability index

2.2.13. Secondary structure:



(a) secondary structure (sec str , entropy, proposed mutation residues)

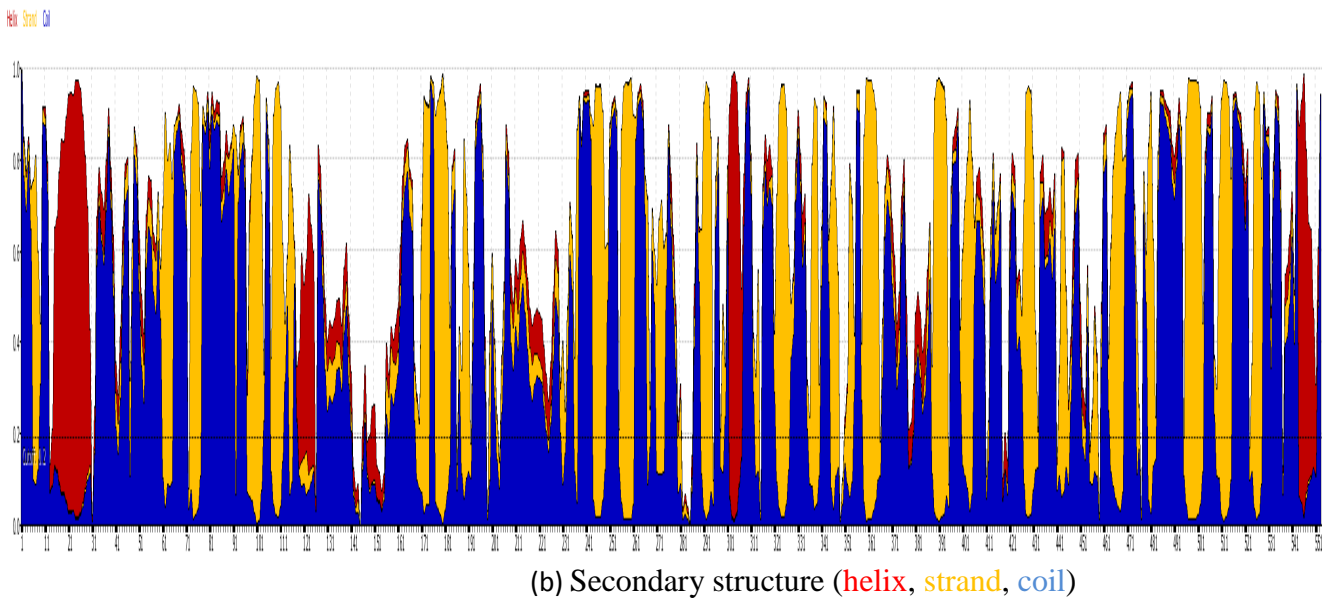


Fig. 2.8. Dioxygenase Secondary Structure

2.2.13.1 Model sequence analysis:

Structural model for the strawberry dioxygenase sequence were generated based on templates as found in the protein BLAST analysis. The template was chosen based on similarity and other parameters such as E-value, MPQs, Z-DOPE and RMSD values.

2.2.13.2. Model sequence:

```

NPKPSKGLTSKLVDLVEKLIVKLMYDSSQPHHYLAGNFTPVIEETPPTKDLNVIGH
LPDCLNGEFVRVGP NPKFAPVAGYHWFDGDGMIHGMR IKDGKATYVSRYVKTSR
LKQEEYFGGAKFMKIGDLKGLFGLLMVNMQQLRAKLVVDLSYGHGTSNTALIY
HHGKLLALSEGDKPYVLKVLEDGDLQTVGLLDYDKRLKVEKPHVTDDHDDVKQ
KLHSFTAHPKVDPFTGEMFTFGYSHDPPYVMYR VVSKDGMHDPVPITVPAPIMM
HDFAITENYAIFMDLPLYFKPKEMVKEGKLIFSDET K KARFGVLPRYAKDDLIR
WFELPNCFIFHNANAWEEEDEVVLITCRLENPD LDMVNGPIKKKLDTFKNELYEM
RFNLKTGLATQKKLSESAVDFPRVNESYTGRKQR FVYGTTLDSIAKVTGIVKFDLH
AAPEVGKTKIEVGGNIQGLYDLGPGRFGSEAIFV PRVPGITSEEDDGYLIFVHDEN
TGKSAIHVLD AKTMSNDPVA VVELPHRVPYGFHAF FVTEEQ LQEQA

```

2.2.13.3. Sequence search:

Table: 2.11 Sequence Similarities

Ref.no.	% similarity	overlap	Name
1. B3FP52_FRAAN	100.000	542	B3FP52 Dioxygenase (Fragment) OS=Fragaria ananassa PE=2 SV=1
2. A9Z0V7_ROSDA	92.620	542	A9Z0V7 Carotenoid cleavage dioxygenase 1 OS=Rosa damascena GN=CCD1
3. Q0H394_CUCME	83.764	542	Q0H394 Carotenoid cleavage dioxygenase (Fragment) OS=Cucumis melo
4. A4URT6_9ROSI	81.734	542	A4URT6 9-cis-epoxycarotenoid dioxygenase OS=Castanea mollissima
5. B5BLW2_MEDTR	82.625	541	B5BLW2 Carotenoid cleavage dioxygenase 1 OS=Medicago truncatula

2.2.13.4. Hydropathy analysis:

12-41 (30) : LVdLVeKLIVKLMYdSSQPhhYLAGNFTPV

DG = 1.5 for segment sequence

DG = 1.39 for most favorable 19AA centered at #32h
(LMYdSSQPhhYLAGNFTPV)

50-68 (19) : dLNVIGHLPdCLNGeFVRV

DG = 0.95 for segment sequence

DG = 0.95 for most favorable 19AA centered at #59d
(DLNVIGHLPdCLNGeFVRV)

74-92 (19) : FAPVAGYhWFdGdGMihGM

DG = 2.32 for segment sequence

DG = 2.32 for most favorable 19AA centered at #83F
(FAPVAGYhWFdGdGMihGM)

122-140 (19) : FMKIGdLKGLFGLLMVNMQ

DG = 0.91 for segment sequence

DG = 0.91 for most favorable 19AA centered at #131L
(FMKIGdLKGLFGLLMVNMQ)

148-166 (19) : VVdLSYGhGTSNTALIYhh

DG = 0.13 for segment sequence

DG = 0.13 for most favorable 19AA centered at #157T
(VVdLSYGhGTSNTALIYhh)

179-197 (19) : YVLKVLedGdLQTVGLLdY

DG = 1.53 for segment sequence

DG = 1.53 for most favorable 19AA centered at #188d
(YVLKVLedGdLQTVGLLdY)

231-249 (19) : FTGeMFTFGYShdPPYVMY

DG = 4.87 for segment sequence

DG = 4.87 for most favorable 19AA centered at #240Y
(FTGeMFTFGYShdPPYVMY)

257-291 (35) : FMhdPVPITVPAPIMMhdFAITeNYAIFMdLPLYF

DG = 13.67 for segment sequence

DG = 8.64 for most favorable 19AA centered at #282A
(HdFAITeNYAIFMdLPLYF)

326-362 (37) : LIRWFeLPNCFIFhNANAWeedeVVLITCRLNPdL

DG = 9.7 for segment sequence

DG = 8.35 for most favorable 19AA centered at #335C
(LIRWFeLPNCFIFhNANAWeedeVVLITCRLNPdL)

455-473 (19) : IQGLYdLGPRFGSeAIFV

DG = 0.51 for segment sequence

DG = 0.51 for most favorable 19AA centered at #464G
(IQGLYdLGPRFGSeAIFV)

476-505 (30) : VPGITSeeddGYLIFFVhdeNTGKSAIhVL

DG = 0.37 for segment sequence

DG = 3.77 for most favorable 19AA centered at #485d
(VPGITSeeddGYLIFFVhdeNTGKSAIhVL)

515-533 (19) : VAVVeLPhRVPYGFhAFFV

DG = 4.82 for segment sequence

DG = 4.82 for most favorable 19AA centered at #524V
(VAVVeLPhRVPYGFhAFFV)

2.2.13.5. β -barrel outer membrane protein analysis:

2.2.13.5.1. Posterior probability plot (model):

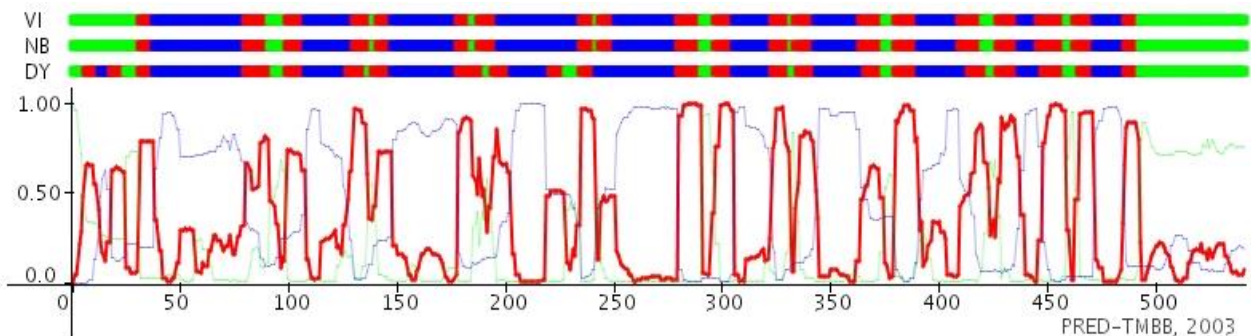


Fig. 2.9. Posterior probability plot (model)

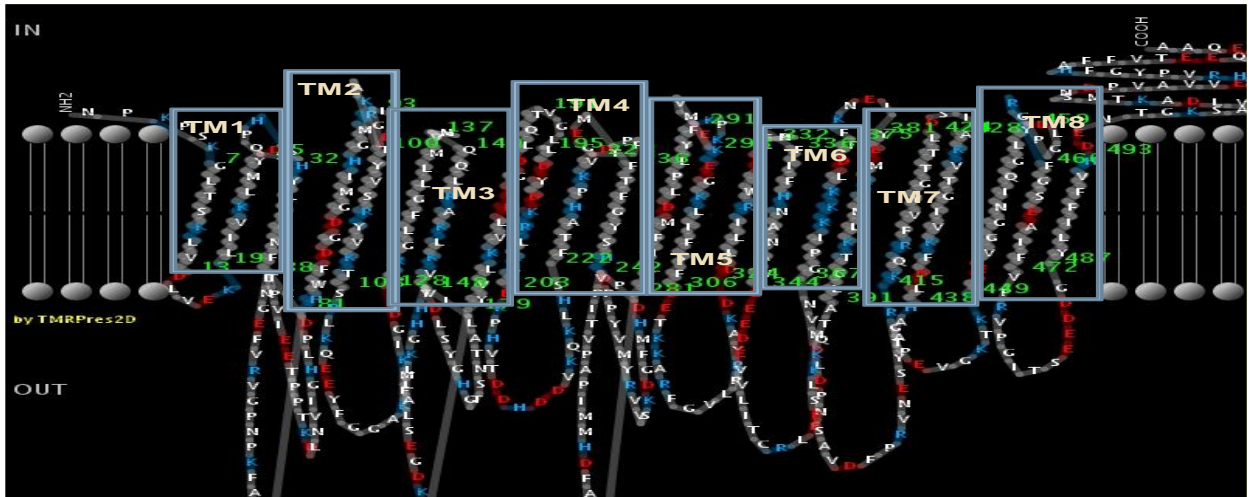


Fig 2.10. 2D-representation (model)

2.2.14. Orientation of Proteins in Membranes:

Table: 2.12. Dioxygenase membrane orientation: hydrophobic thickness, depth and inclination from the membrane surface

Depth/Hydrophobic Thickness	Tilt Angle
$\Delta G_{\text{transfer}} 5.4 \pm 0.9 \text{ \AA} - 9.3 \text{ kcal/mol}$	$46 \pm 3^\circ$

Table: 2.13 Membrane embedded residues

Membrane Embedded Residues (in Hydrocarbon Core)		
Subunits	Tilt	Segments
Embedded_residues:		
B	46	12,15-16,19-23,131,134-135,297-299

The β -barrel outer membrane protein analysis indicates that dioxygenase structural model is consistent with the prediction that it is a membrane integrated protein with some transmembrane sequence segments which precludes the protein as heteroatoms

included peripheral protein (Eyre. Tina A. *et al.*, 2004, Kiser, Philip D. *et al.*, 2009; Kiser, Philip D. *et al.*, 2010).

2.2.15. Model Quality:

coil - coiled coil analysis:

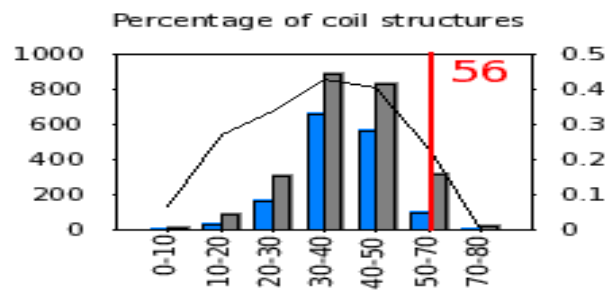


Fig. 2.11. Coil structures

2.2.15.1. Coiled- coil probability:

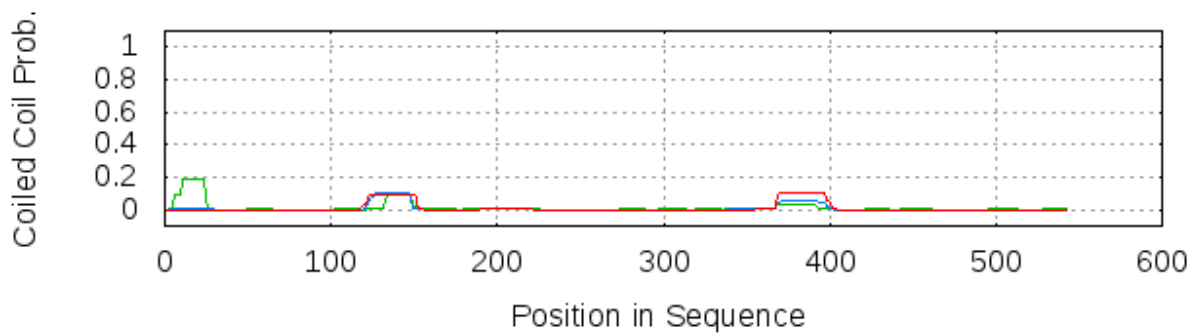


Fig. 2.12. Coiled Coil Probabilities

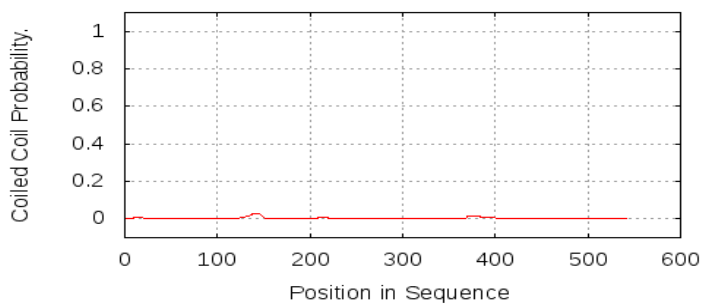


Fig. 2.13. Coiled Coil probability

2.2.15.2. Coiled coil probability:

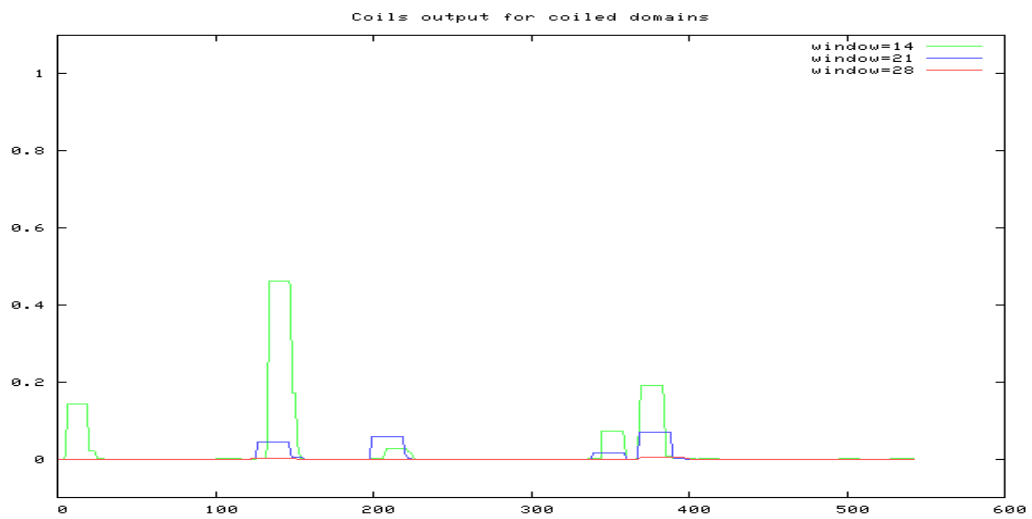


Fig. 2.14. coiled domains

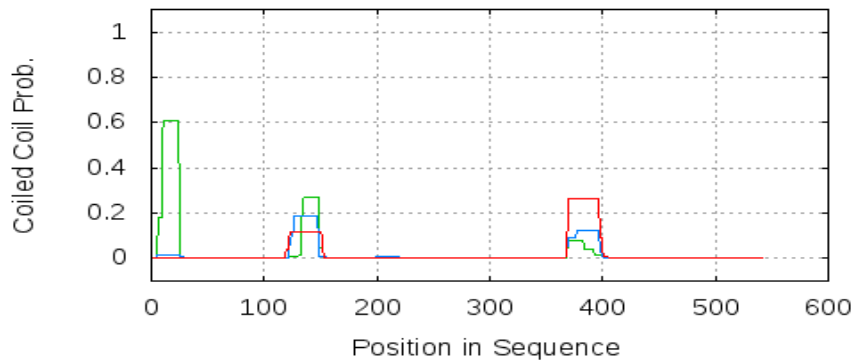


Fig. 2.15. coil-pcoil

Consistent with the coiled coil analysis of the strawberry sequence, the structural model has some coiled-pcoiled structure with very less probability, which indicates the structure model should be further, refined but based on other investigation, it can be exempted for instance the QMEAN z-score values and ERRAT analysis.

2.2.16. Structure model validation and quality assessment:

2.2.16.1. Dioxygenase model Residue Error

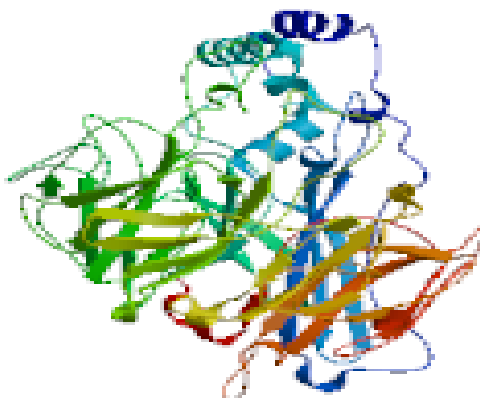


Fig. 2.16. Dioxygenase structure model coloring by residue error

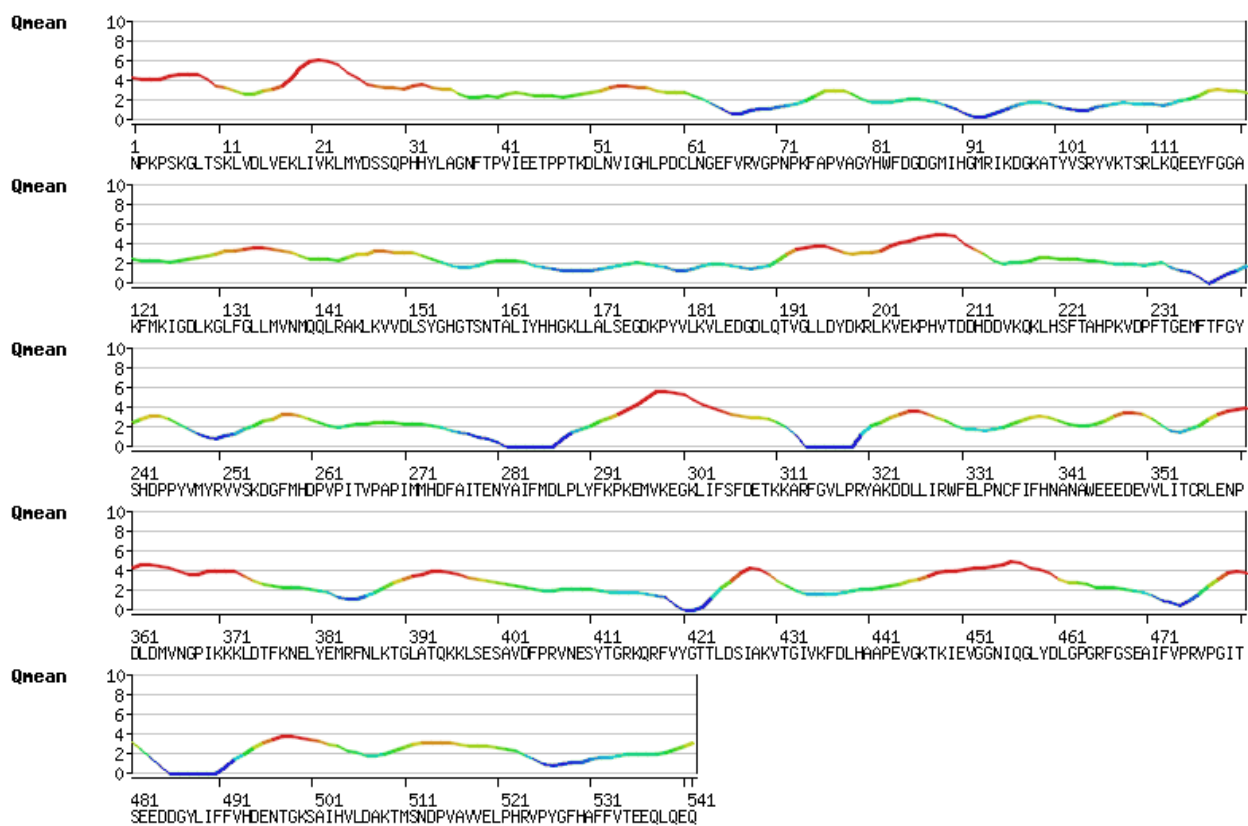


Fig.2.17. QMEAN Plot

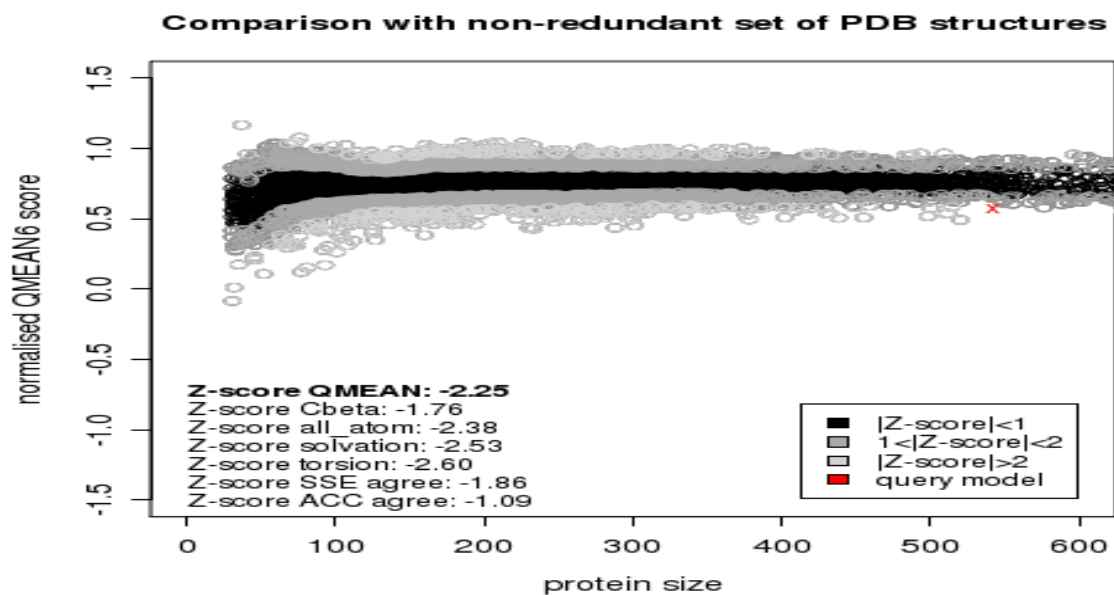


Fig. 2.18. Normalized QMEAN score on comparison with non-redundant set of PDB structures

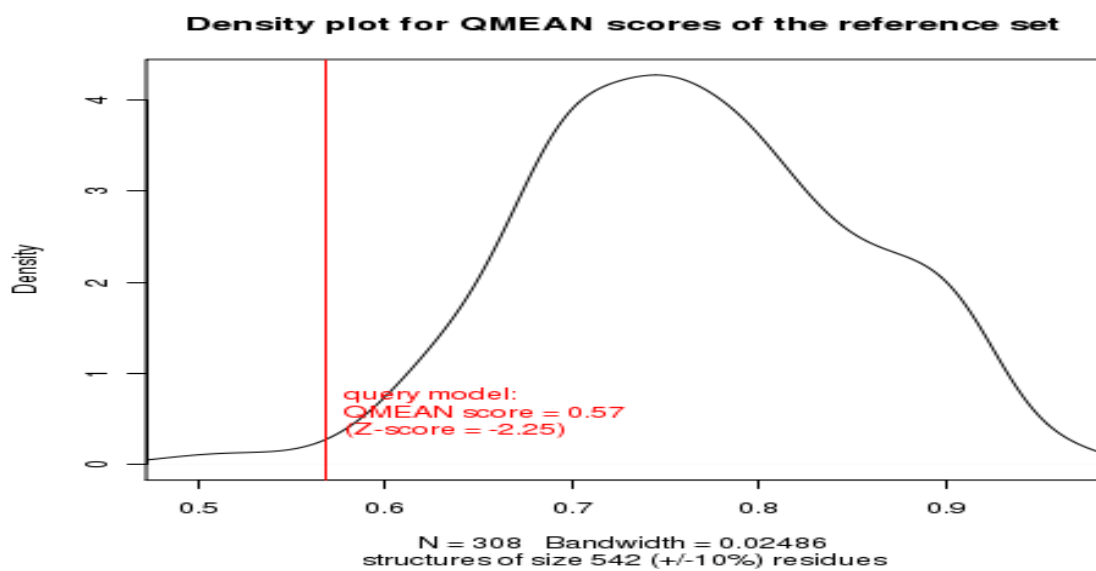


Fig. 2.19. QMEAN score Density Plot

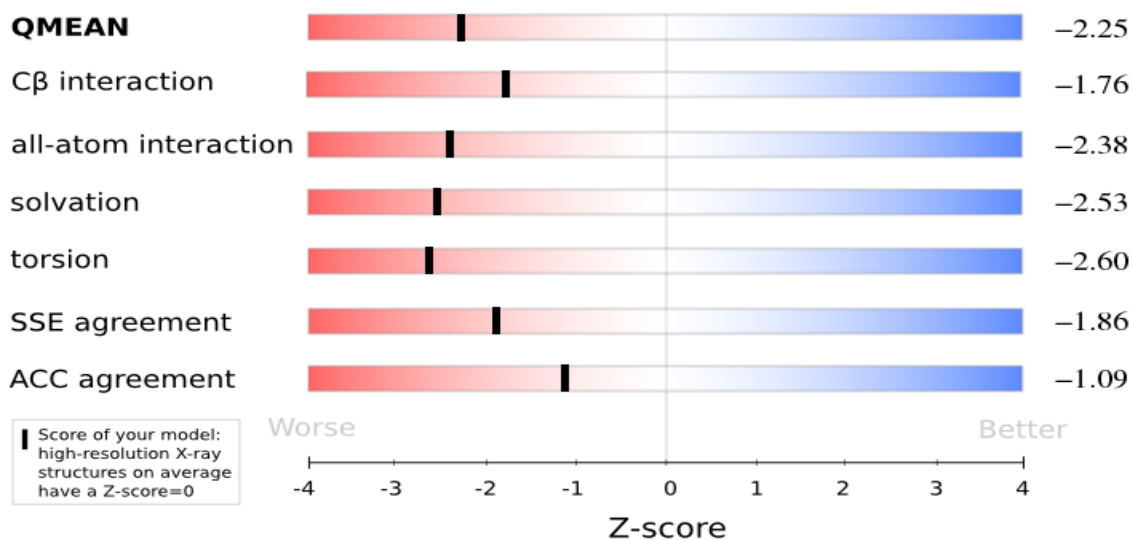


Fig. 2.20. QMEAN Z-score

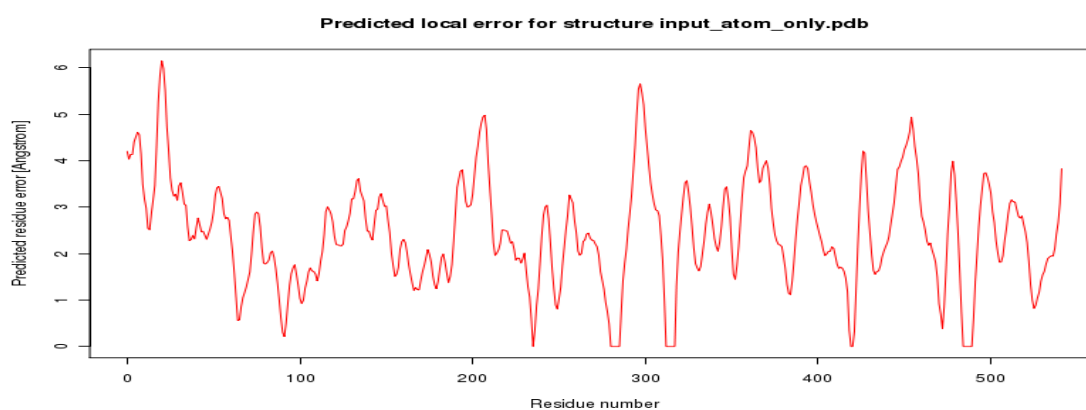


Fig. 2.21. Predicted Local residue Error

Table: 2.14 QMEAN Z-score

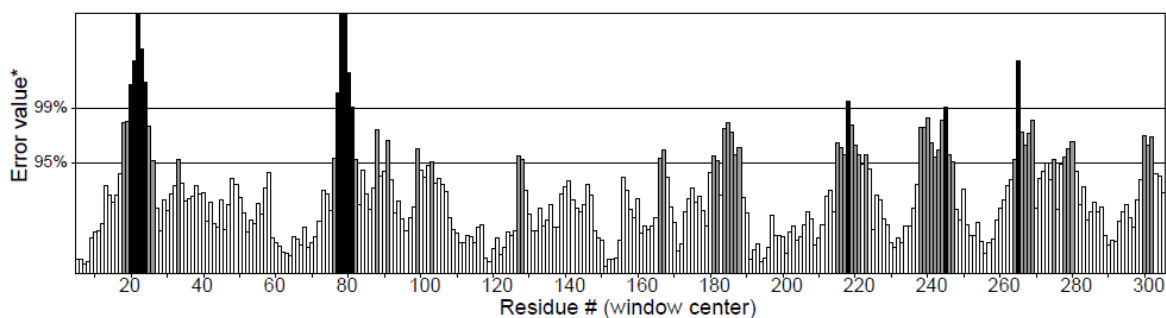
Scoring function term	Raw score	Z-score
C_beta interaction energy	-45.09	-1.76
All-atom pairwise energy	-5204.92	-2.38
Solvation energy	-9.23	-2.53
Torsion angle energy	-60.64	-2.60
Secondary structure agreement	70.1%	-1.86
Solvent accessibility agreement	75.5%	-1.09
QMEAN6 score	0.568	-2.25

A linear combination of 6 terms (for model reliability between 0-1) is given as a composite score known as the QMEAN6 score. The pseudo-energies are represented together with their Z-scores with respect to scores obtained for high-resolution similar size experimental structures solved by X-ray crystallography:

Program: ERRAT2

Chain#:1

Overall quality factor**: 76.779



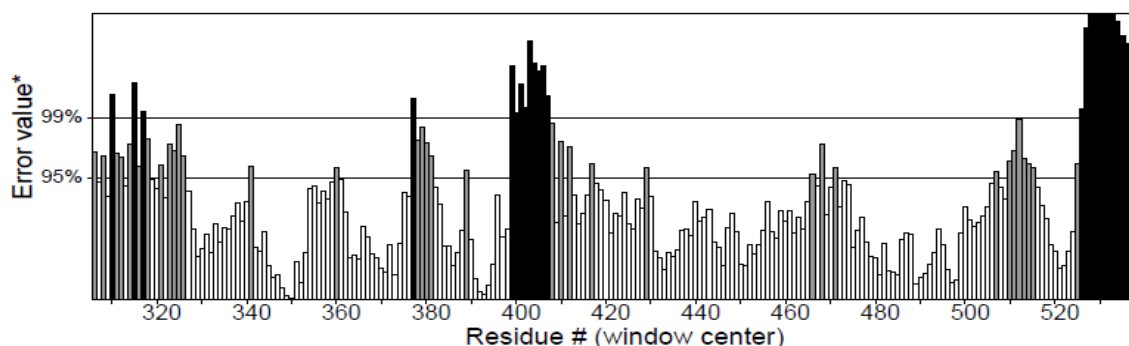
*On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

**Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3Å) the average overall quality factor is around 91%.

Program: ERRAT2

Chain#:1

Overall quality factor**: 76.779



*On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

**Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3Å) the average overall quality factor is around 91%.

Fig. 2.22. ERRAT Error Value

The ERRAT analysis indicates some reservations need to be followed, that is the resolution of good quality crystals; since, homology model structure and substrate binding is being investigated, the rejection limit of the model is permissible with the fact that other parameters are supporting the structure model to be a good quality model. Also, initial residues of the model do not show much higher peaks of error values which are within permissible limits and positions where coiled-coil structure has been predicted are under permissible limit at higher resolution as well as low 2.5-3 Å resolution structure.

2.2.16.2. Ramachandran plot:

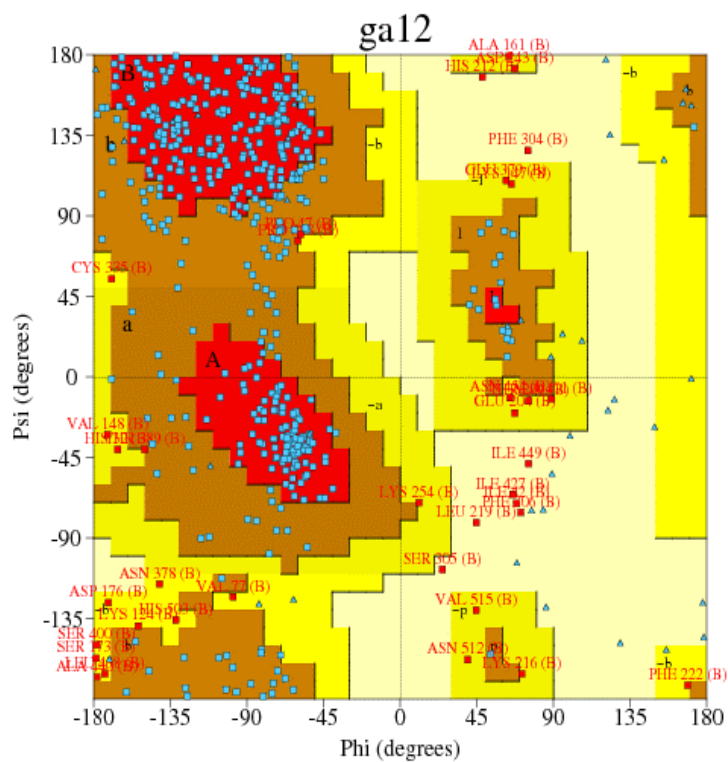


Fig. 2.23. Ramachandran Plot

1. Ramachandran plot statistics:

	No. of residues	Percentage
Most favored regions A,B, L	306	65.9%
Additional allowed regions a, b, l, p	124	26.7%
Generously allowed regions ~a,~b,~l,~p	26	5.6%
Disallowed regions XX	8	1.7%

Non glycine and non proline residues	464	100%
End residues (excl. Gly and Pro)	2	
Glycine residues	42	
Proline residues	34	

Total number of residues	542	

This plot is generated on the basis of an analysis of 118 different structures of minimum resolution of at least 2.0 Å and *R*-factor with a value of no greater than 20.0. However, number of residues in the most favorable region is within permissible limit.

2. G-Factor:

Parameter	Average Score
Dihedral angles	
Phi-psi distribution	-1.29
Chi1-chi2 distribution	-0.66
Chi1only	-0.22
Chi3&chi4	0.24
Omega	0.68

	-0.31
Main chain covalent forces:	
Main chain bond lengths	0.68
Main chain bond angles	0.56
	0.61

Overall average	0.05

Important note: These values are generated by comparing the main-chain bond-lengths and bond angles with the Engh & Huber (1991) ideal values derived from small-molecule data. Therefore, refining of structures using different restraints could cause apparently large deviations from normality.

2.2.17. Fold and Motif analysis:

2.2.17.1. Fold analysis:

Folds: On fold analysis more of 64 folds were found, but 5 of them are mentioned which are of much relevance

	<u>Q-</u> <u>score</u>	<u>Rmsd</u>	<u>No.</u> <u>SSE</u>	<u>Z-</u> <u>score</u>	<u>PDB</u>	<u>Name</u>
1.	0.161	2.94Å	16	6.7	2w18	Crystal structure of the C-terminal wd40 domain of human palb2
2.	0.123	3.98Å	14	2.7	3jrp	Sec13 with nup145c (aa109-179) insertion blade
3.	0.119	3.57Å	15	5.2	3jpx	Eed: a novel histone trimethyllysine binder within the eed-e polycomb complex
4.	0.113	4.14Å	12	2.9	3f3p	Crystal structure of the nucleoporin pair nup85-seh1, space p21212
5.	0.111	4.31Å	13	3.5	3izb	Localization of the small subunit ribosomal proteins into a cryo-em map of saccharomyces cerevisiae translating 80s rib

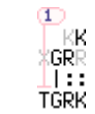



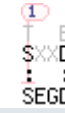
2.2.17.2. Motif analysis:

Table 2.15: Residues and modifications

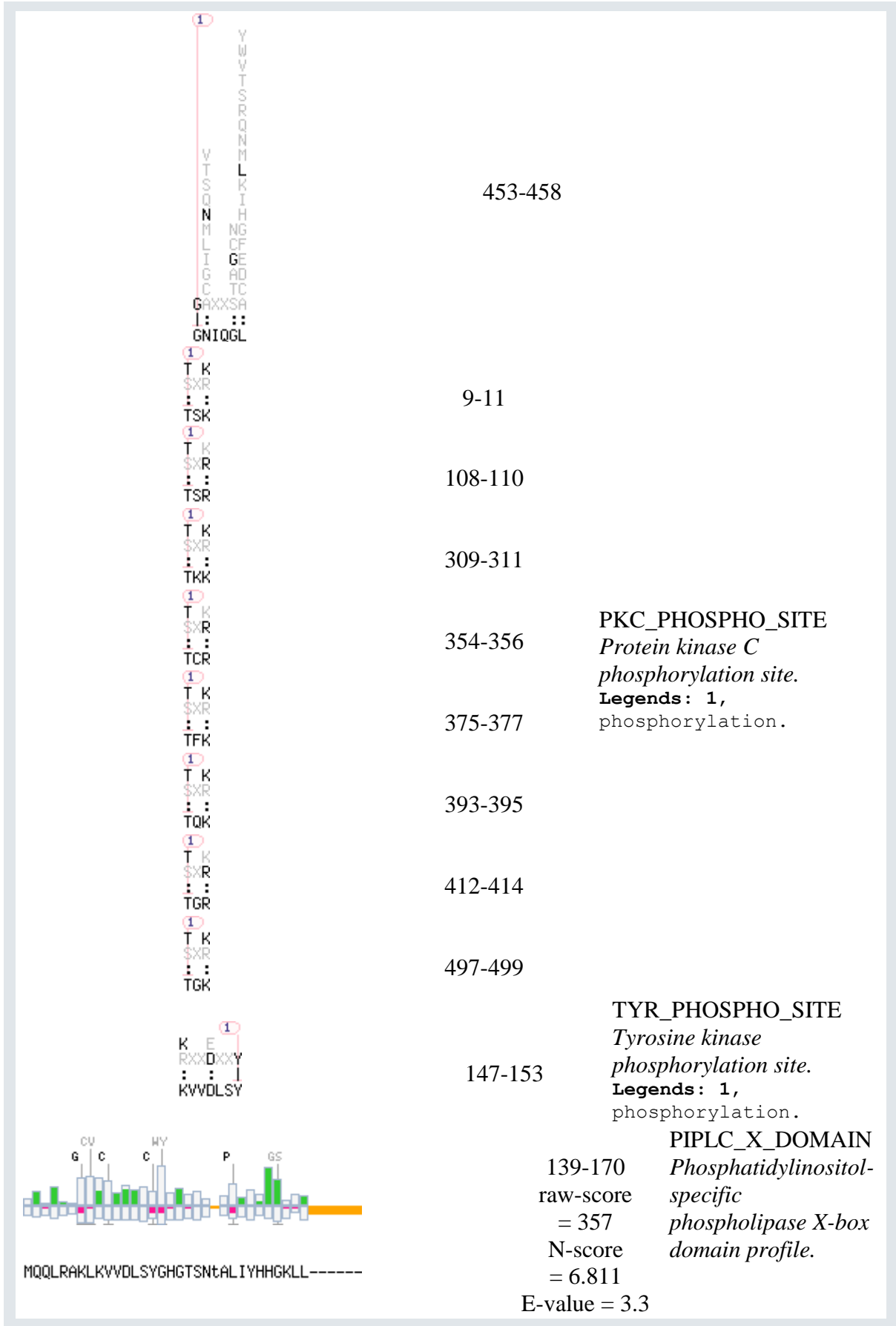
Residue position		Modification type
Start	End	
412	415	AMIDATION
408	411	ASN GLYCOSYLATION
440	447	ATP GTP A
395	398	CAMP PHOSPHO SITE
173	176	CK2 PHOSPHO SITE
305	308	CK2 PHOSPHO SITE
400	403	CK2 PHOSPHO SITE
422	425	CK2 PHOSPHO SITE
447	450	CK2 PHOSPHO SITE
480	483	CK2 PHOSPHO SITE
87	92	MYRISTYL
156	161	MYRISTYL
453	458	MYRISTYL
9	11	PKC PHOSPHO SITE
108	110	PKC PHOSPHO SITE
309	311	PKC PHOSPHO SITE

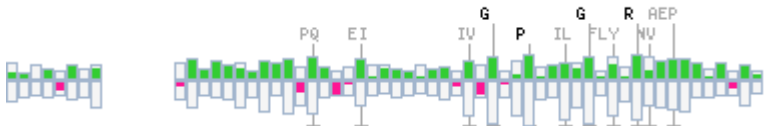
354	356	PKC PHOSPHO SITE
375	377	PKC PHOSPHO SITE
393	395	PKC PHOSPHO SITE
412	414	PKC PHOSPHO SITE
497	499	PKC PHOSPHO SITE
147	153	TYR PHOSPHO SITE
139	170	PIPLC X DOMAIN
21	535	RPE65
21	535	RPE65

Table: 2.16 Motifs with possible Post Translational modifications

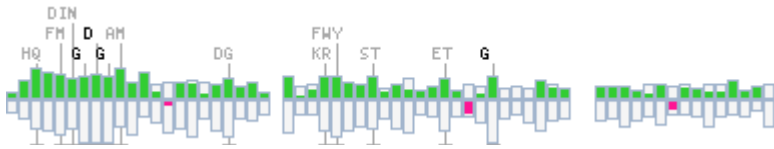
Detail of motifs		
	412-415	AMIDATION <i>Amidation site.</i> Legends: 1, amidation.
	408-411	ASN_GLYCOSYLATION <i>N-glycosylation site.</i> Legends: 1, carbohydrate.
	440-447	ATP_GTP_A <i>ATP/GTP-binding site motif A (P-loop).</i>
	395-398	CAMP_PHOSPHO_SITE <i>cAMP- and cGMP-dependent protein kinase phosphorylation site.</i> Legends: 1, phosphorylation.
	173-176	CK2_PHOSPHO_SITE <i>Casein kinase II</i>

<p>① T E SXXD : SFDE</p> <p>① T E SXXD : SAVD</p> <p>① T E SXXD : TTLD</p> <p>① T E SXXD : TKIE</p> <p>① T E SXXD : TSEE</p>	<p>305-308</p> <p>400-403</p> <p>422-425</p> <p>447-450</p> <p>480-483</p>	<p><i>phosphorylation site.</i> Legends: 1, phosphorylation.</p>
<p>① Y W V T S R R Q N M L K I H V T S Q N M L I G C G A X X S A I : : G M I H G M</p> <p>① Y W V T S R R Q N M L K I H V T S Q N M L I G C G A X X S A I : : G T S N T A</p>	<p>87-92</p> <p>156-161</p>	<p>MYRISTYL <i>N-myristoylation site.</i> Legends: 1, myristyl.</p>

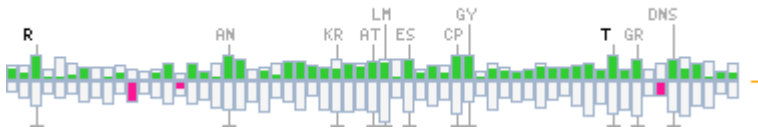




VKLMYDSS-----QPHHYLAGNFTPVIEETPPTKDLNVIGHLPDCLNGEFVRVGPMPKFAFV-



GYHWFDGDMIHGMRIKDGKAT-YVSRVYKTSRLKQEEYFGGAKFMK--IGDLKGLFGLLMVNM



QLRAKLVVDLSYGHGTSNTALIYHHGKLLALSEGDKPYVLKVLDEGDLQTVGLLDYDKRLkvel

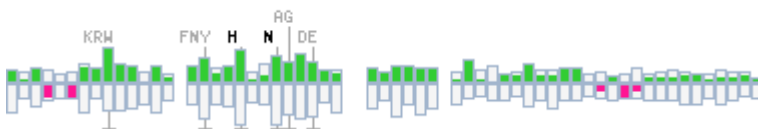
21-535
 raw-score pfam_fs:RPE65
 = 863.7 *Retinal pigment*
 N-score *epithelial*
 = 264.96 *membrane*
 4 *protein*
 E-value
 = 2.3e-
 258



phvtddhddvkQKLHSFTAHPKVDPFT----GEMFTFGYSHDP---PYVMYRVVSKDGFMDHPW



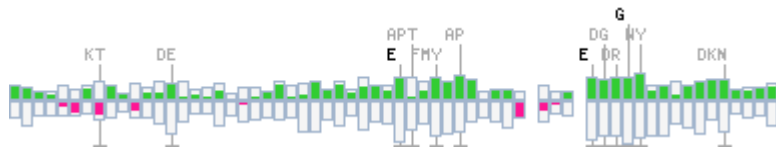
ITVPAPIMMHDAITENYAIFMDLPLYFKP---KEMVKEGKLIFS-----FDETKKARFGVLPK



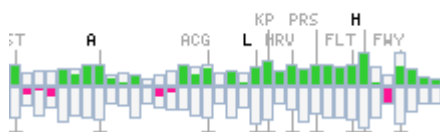
AKDILLIRWFELPN-CFIFHNANAWEEE--DEVVLI-TCRLNPDLDMVNGPIKIKKLDTFKNE-



--LYEMRFNLKTG-LATQKKLSES--AVDFPR--VNESYTGKQRFVYGGTTLDS-IAKVTGIVK



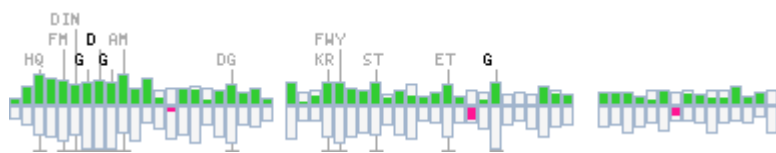
DLHAAPEVVGKTKIEVGGNIQGLYDLGPRFGSEAFVPRVPGI-TSE-EDDGYLIFVHDENTG



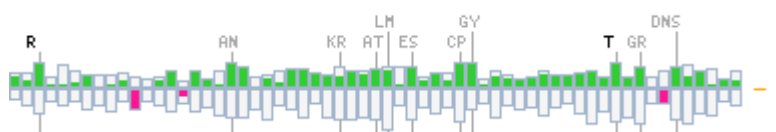
SAIHVLDAKTMSNDPVAVVELPHRVPYGFHAFVTE



VKLMYDSS-----QPHHYLAGNFTPVIEETPPTKDLNIGHLPDCLNGEFVRVGNPKFAPV-

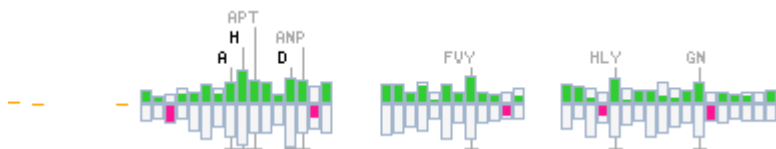


GYHWFDGDMIHGMRIKDGKAT-YVSRVYKTSRLKQEEYFGGAKFMK--IGDLKGLFGLLMVNM



QLRAKLKVVVDSYGHGTSNTALIYHHGKLLALSEGDKPYVLKVLLEDGDLQTVGLLDYDKRLkve

pfam_ls:RPE6
 21-535
 5
 raw-score = 865.5
 N-score = 265.494
 E-value = 6.8e-259
Retinal pigment epithelial membrane protein



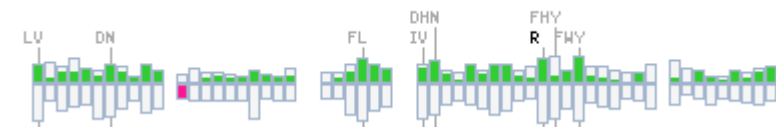
phvtdhdhddvkQKLHSFTAHPKVPDPFT----GEMFTFGYSHDP---PYVMYRVVSKDGFMDHPV



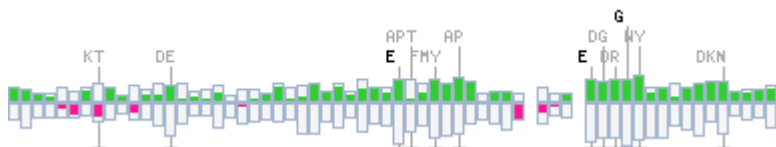
ITVPAPIMMHDFAITENYAIFMDLPLYFKP---KEMVKEGKLIFS-----FDETKKARFGVLP



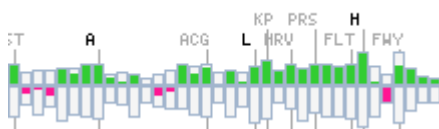
AKDLLIRWFELPN-CFIFHNANAWEEE--DEVVLI-TCRLENPDLOMVNGPIKKKLDTFKNE-



--LYEMRFNLKTG-LATQKKLSES--AVDFPR--VNESYTRKQRFVYGTLLDS-IAKVTGIWK



DLHAAPEVGTKIEVGGNIQGLYDLGPRFGSEAFVPRVPGI-TSE-EDDGYLIFVHDENTG



SAIHWLDAKTMSNDPVAVVWELPHRVPYGFHAFFVTE

Motif analysis for the model sequence indicated same amidation, glycosylation and phosphorylation sites as the strawberry dioxygenase sequence shows for such modifications but with a variation in the position, since the protein undergoes post-translation modifications during its processing through endoplasmic reticulum and golgi complex, and protein folding occurs in the way through its movement from ER to Golgi complex, the sequence depicted what genes have encoded for the dioxygenase protein and in homology modeling, structures are generated based on homology and sequence similarities, thus the model sequence shows positional changes for motifs (Witze, Eric S. *et al.*, 2007).

2.2.17.3. Phosphorylation sites:

Table: 2.17 Net Phosphorylation Sites and Specific Kinases

Position	Residue	Kinase
5	S	PKA
9	T	PKC
10	S	CaM-II
27	S	GSK3
28	S	DNAPK
39	T	GSK3
45	T	CKI
48	T	GSK3
100	T	PKC
103	S	PKC
108	T	PKC
109	S	PKC

152	S	PKA
157	T	GSK3
158	S	PKA
160	T	cdc2
173	S	CKI
191	T	CKII
209	T	CKII
221	S	PKC
223	T	PKG
232	T	GSK3
237	T	PKC
241	S	cdc2
253	S	PKC
265	T	PKC
278	T	CKII
305	S	CKII
309	T	PKC
354	T	PKC
375	T	PKC
389	T	PKC
393	T	PKC
398	S	PKA
400	S	PKG
410	S	PKC
412	T	PKC
422	T	PKC
423	T	CaM-II
426	S	PKC

431	T	PKC
447	T	PKG
468	S	PKA
480	T	CKII
481	S	CKII
497	T	PKC
500	S	CaM-II
509	T	GSK3
511	S	PKA
534	T	CKII
25	Y	
33	Y	INSR
80	Y	
101	Y	INSR
105	Y	
116	Y	SRC
153	Y	INSR
164	Y	INSR
179	Y	INSR
197	Y	INSR
240	Y	INSR
246	Y	SRC
249	Y	INSR
281	Y	INSR
290	Y	SRC
320	Y	INSR
381	Y	INSR
411	Y	INSR

420	Y	INSR
459	Y	INSR
487	Y	
526	Y	

Similar, variations in the phosphorylation at different positions have been shown, which indicates post translational modifications the dioxygenase might undergo upon activation or in terms of activation process by specific kinases or endogenous environmental cues, which could trigger specific kinases for initiating phosphorylation of specific amino acid residues (Witze, Eric S. *et al.*, 2007).

2.2.17.4. Glycosylation sites:

2.2.17.4.1. NetCGlyc Prediction:

Table: 2.18 Net C-Glycosylation Sites

Feature	Start	Stop
C-manno	82	82
C-manno	329	329
C-manno	344	344

The C-glycosylation in the model sequence shows some variation in the corresponding positions from the glycosylation sites in the strawberry protein sequence, that indicates the possible modification changes in the protein during protein folding and processing in the endoplasmic reticulum (ER) and Golgi complex. The structure model sequence shows glycosylations at positions 82, 329 and 344, while the strawberry protein sequence indicated C-glycosylations at positions 91, 338 and 353, which could be due to homology between different dioxygenases and templates of different similar proteins.

2.2.17. 4.2 NetNGlyc Prediction:

Table: 2.19 Net N Glycosylation sites

Position	Potential
37	NFTP
408	NESY

Similarly, N-glycosylation in the model sequence shows some variations in the positions from the glycosylation positions in the strawberry sequence. The structure model sequence shows glycosylations at positions 37 and 408, while the strawberry sequence indicated glycosylations at 46 and 417; this could be due to homology modeling. These variations in the N-glycosylation positions may also indicate the possibility for modification changes of specific amino acid residues during protein processing and protein folding.

2.2.17.4.3. NetOGlyc Prediction:

Table: 2.20 Net O Glycosylation sites

S/T	Position	Y/N
S	5	.
T	9	.
S	10	.
S	27	.
S	28	.
T	39	.
T	45	.
T	48	.
T	100	.

S	103	.
T	108	.
S	109	.
S	152	.
T	157	.
S	158	.
T	160	.
S	173	.
T	191	.
T	209	.
S	221	.
T	223	.
T	232	.
T	237	.
S	241	.
S	253	.
T	265	T
T	278	.
S	305	.
T	309	.
T	354	.
T	375	.
T	389	.
T	393	.
S	398	.
S	400	.
S	410	.
T	412	.

T	422	.
T	423	.
S	426	.
T	431	.
T	447	.
S	468	.
T	480	.
S	481	.
T	497	.
S	500	.
T	509	.
S	511	.
T	534	.

The O-glycosylation in the model sequence shows some variation in the positions, as the strawberry sequence indicated most possible glycosylation at Thr 274, but the model sequence shows O-glycosylation of Thr residue at position 265, other Ser and Thr residues also have variation in O-glycosylation at different positions. This could be due to homology modeling or these changes indicate possible modifications during protein processing in the ER and Golgi complex, suggesting further modifications in the protein folding (Witze, Eric S. *et al.*, 2007).

2.2.18. Reverse template comparison vs structures in PDB.

1. [3npe Structure of vp14 in complex with oxygen](#)
2. [2biw Crystal structure of native enzyme of synechocystis apocarotenoid cleavage oxygenase](#)
3. [3kvc Crystal structure of rpe65 from bovine at 1.9 Å⁰ resolution](#)
4. [1pby Structure of the quinoxinone amine dehydrogenase adduct phenylhydrazine from paracoccus denitrificans at 1.7 Å⁰ resolution](#)
5. [3hx6 Crystal structure of pily1 c-terminal domain from pseudomonas aeruginosa](#)

2.2.19. Existing PDB structures:

(PDB code)	(E-value)	% similarity	overlap	Name
1. 3npe(A)	2.1e-28	38.636	528	Structure of vp14 in complex with oxygen
2. 2biw(A)	6e-06	27.567	526	Crystal structure of apocarotenoid cleavage oxygenase from synechocystis, native enzyme
3. 2bix(A)	6e-06	27.567	526	Crystal structure of apocarotenoid cleavage oxygenase from synechocystis, fe-free apoenzyme
4. 4f2z(A)	1.6e-05	21.507	544	Crystal structure of rpe65 in a lipid environment
5. 3fsn(A)	2.6e-05	21.228	570	Crystal structure of rpe65 at 2.14 angstrom resolution

2.2.20. Model:

Accession	<i>undefined</i>
Species	<i>Strawberry</i>
Length	542 aa
Isoelectric Point	0
Charge at pH7	-4.8
Molecular Weight	71,017 Da

2.2.20.1. Model secondary structure:

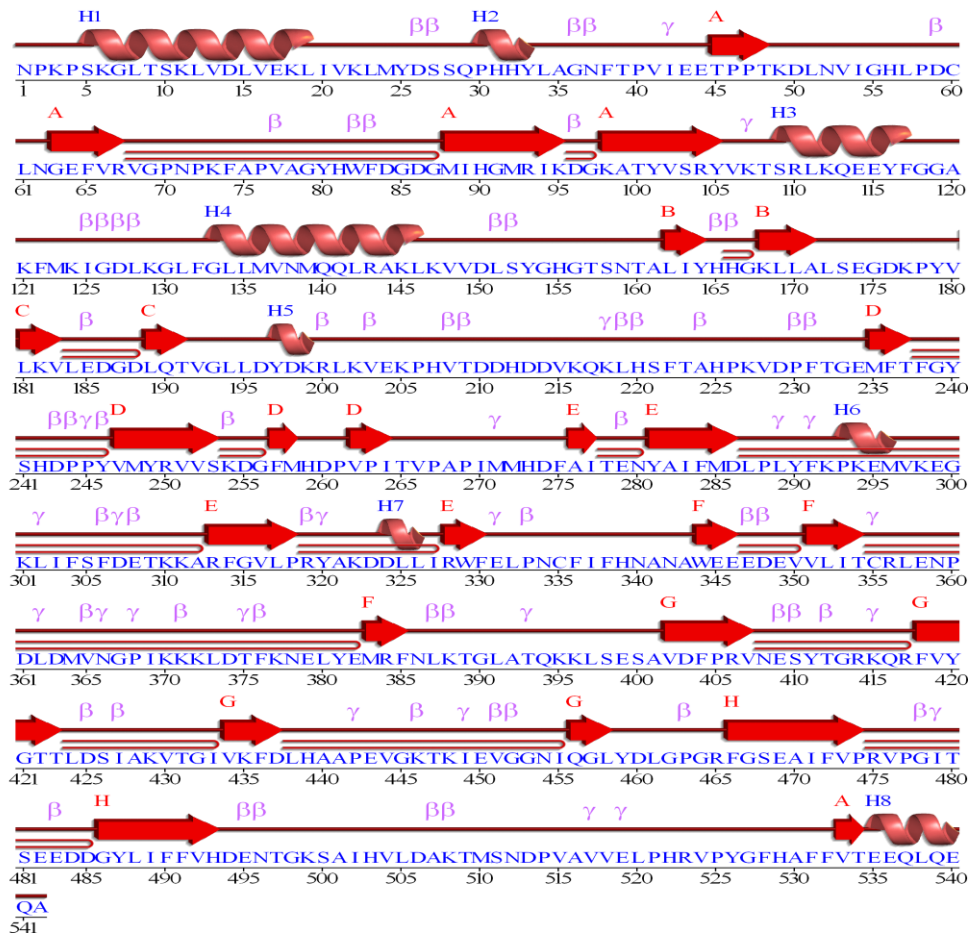




Fig. 2.24. Dioxygenase model secondary structure

Key:

- Sec. struc:  Helices labelled H1, H2, ... and strands by their sheets A, B, ...
 Helix Strand
 Motifs: β beta turn γ gamma turn  beta hairpin

2.2.20.2. Model secondary structure topology:

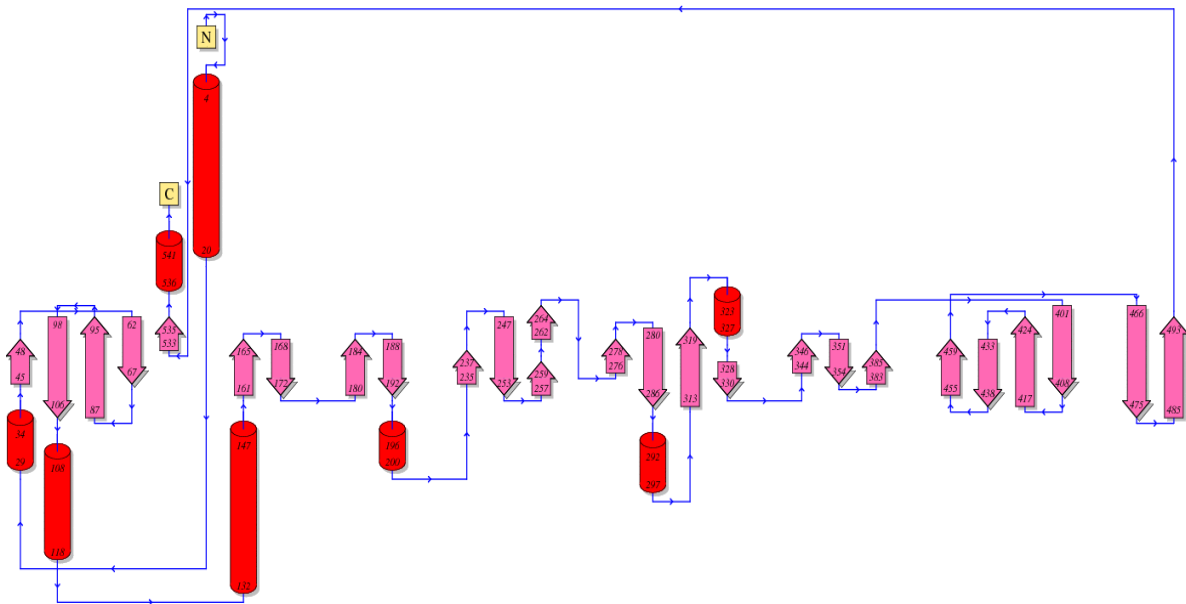


Fig.2.25. Dioxygenase model secondary structure topology

2.2.21. Dioxygenase protein structure:

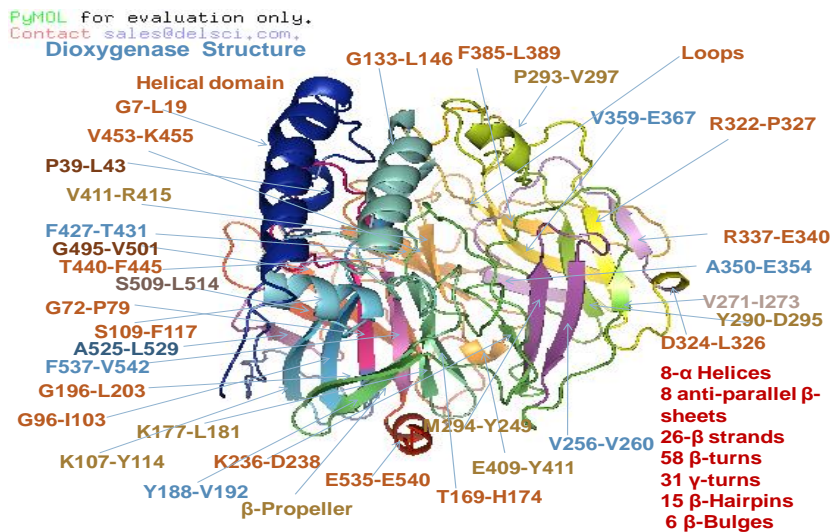


Fig.2.26. Dioxygenase Structure

PyMOL for evaluation only.
Contact sales@delsci.com.

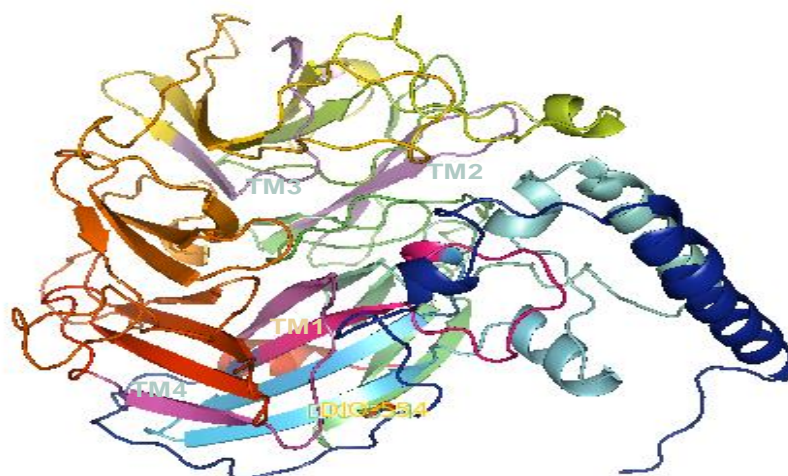


Fig. 2.27. Dioxygenase-dioxygen

2.2.21.1.1. Dioxygenase membrane orientation:

PyMOL for evaluation only.
Contact sales@delsci.com.

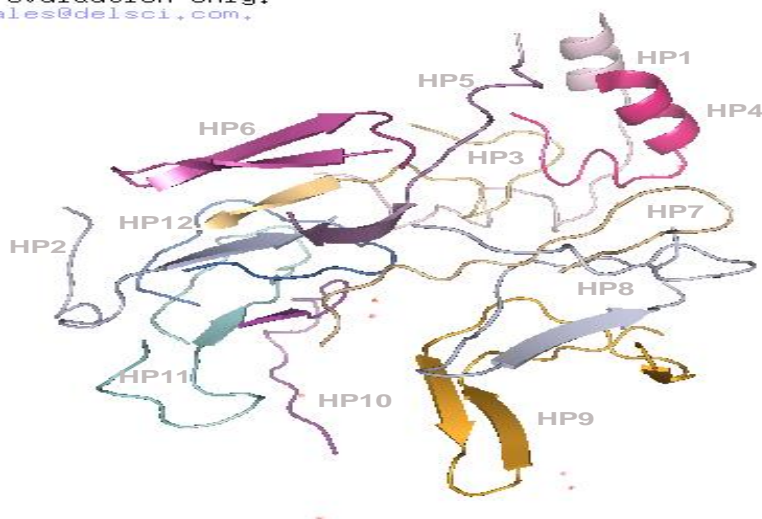


Fig. 2.28. Dioxygenase hydrophathy segments (HP1-HP12) when His, Asp and Glu residues are neutral and the partition is from bilayer to water.



Fig. 2.29. Transmembrane Segments

2.2.21.1.2. Transmembrane segments

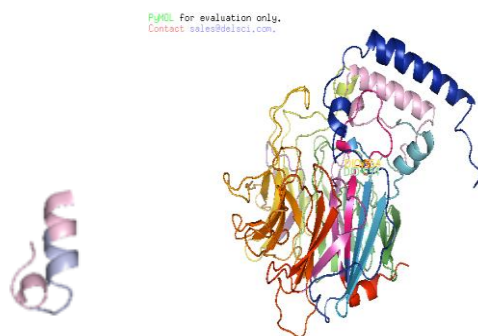


Fig. 2.30 Transmembrane (TM1)

(1) FVRVGNPKFAPVAGYhWF

(2) MFTFGYShDPPYVMYRVVS

(3) LIRWFELPNCFIFhNANAW

(4) VAVVELPhRVPYGFhAFFV

(5) Lys130, Phe131, Met132, Lys133, Ile134, Gly135, Asp136, Leu137, Lys138, Gly139, Leu140, Phe141, Gly142, Leu143, Leu144, Met145, Val146, Asn147, Met148, Gln149, Gln150, Leu151, Arg152, Ala153, Lys154, Leu155, Lys156, Val157, Val158, Asp159, Leu160, Ser161, Tyr162

Analysis of surface properties of α -helix and other residues indicates a hydrophobic patch formed by around 25 residues which are able to penetrate the membrane to the membrane phospholipids head groups and in the membrane interior interact with the fatty acid.

2.2.21.1.3. Orientation of dioxygenase in membranes:

Depth/Hydrophobic Thickness	Tilt Angle
$\Delta G_{\text{transfer}} 5.4 \pm 0.9 \text{ \AA}-9.3 \text{ kcal/mol}$	$46. \pm 3.^\circ$

Pymol for evaluation only.
Contact sales@deisc1.com.

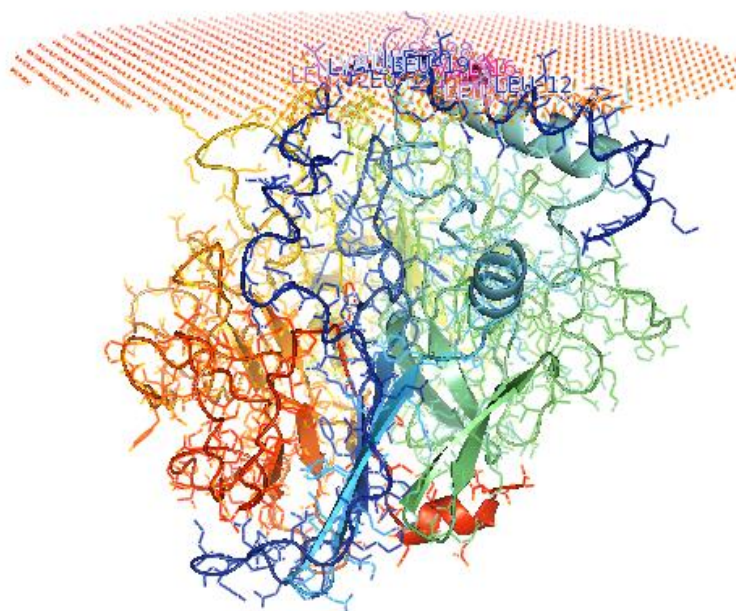


Fig. 2.31. Dioxygenase membrane orientation and membrane interacting residues

[Leu 12, Leu15, Val16, Leu19, Leu23, Leu131, Leu134, Leu135, Val297, Lys298, Glu299]

The membrane orientation of the dioxygenase indicates mainly branched chain residues i.e. leu, ile and val that are interacting with the membrane surface, which in the homology model structure is represented by a circular plane. In this representation the hydrophobic α -helix and the helical domain are interacting with the membrane surface (Messing *et al.*, 2010).

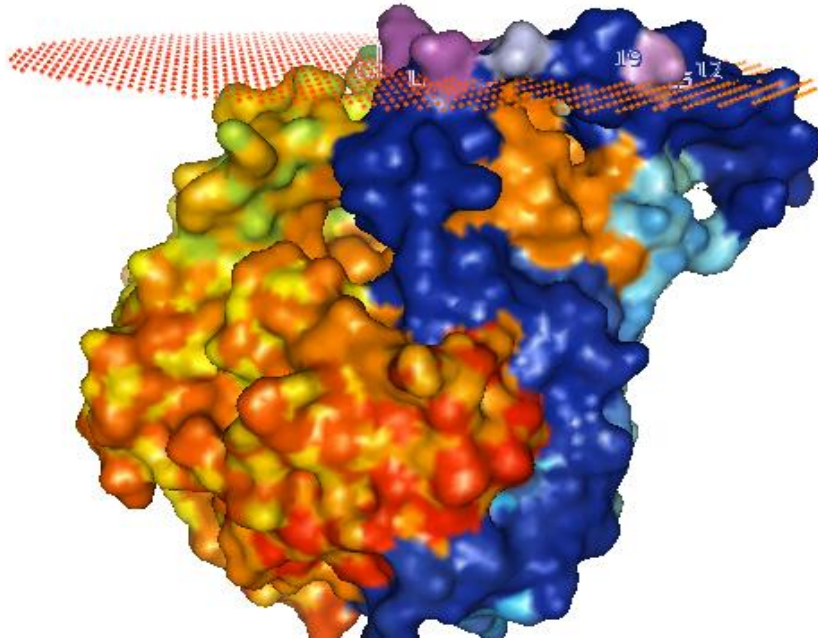


Fig. 2.32. Dioxygenase membrane orientation, Hydrophobic Pore line patches (shown in orange)

Pro30,Pro70,Phe74,Ala75,Pro76,Val77,Ala78,Gly79,Tyr80,Trp82,Phe83,Asp84,Gly85,Asp86,Met88,Ser109,Arg110,Glu114,Ala120,Lys121,Phe122,Met123,Lys124,Ile125,Asp127,Leu128,Lys129,Phe132,Met136,Gln140,Val149,His155,Thr157,Asn159,Thr160,Ala161,Ile163,His165,Leu172,Ser173,Glu174,Gly175,Asp176,Glu204,His220,Thr223,Ala224,His225,Pro226,Lys227,Val228,Asp229,Met235,Tyr240,Met271,His273,Asp274,Phe275,Ile277,Met285,Val297,Lys298,Glu299,Leu302,Ile303,Phe338,His339,Asn340,Ala341,Glu347,Leu357,Pro368,Phe376,Lys377,Leu380,Phe466,Ser468,Glu469,Phe472,Val473,Pro474,Arg475,Val476,Pro477,Glu482,Glu483,Asp484,Asp485,Gly486,Tyr487,Leu488,Glu495,Ala507,Thr509,Tyr526,Gly527,Phe528,His529,Ala530,Phe531,Phe532,Glu536,Gln537

In this position, the dioxygenase structure model indicates that a group of neutral, positive and negatively charged residues interact directly with charged surface of the membrane or through contacts mediated by water (Messing *et al.*, 2010). Consistent with the previous studies this α -helix plays key role in the membrane interactions as the removal of this helical domain disrupts the hydrophobic patches and exposes the positive surface charge and also, this explains that the removal of this helical domain would result in the loss of the

enzyme-membrane interaction and the loss of abscisic acid biosynthesis, which have been observed in many other biochemical and genetic studies.

2.2.21.2. Electrostatic patches:

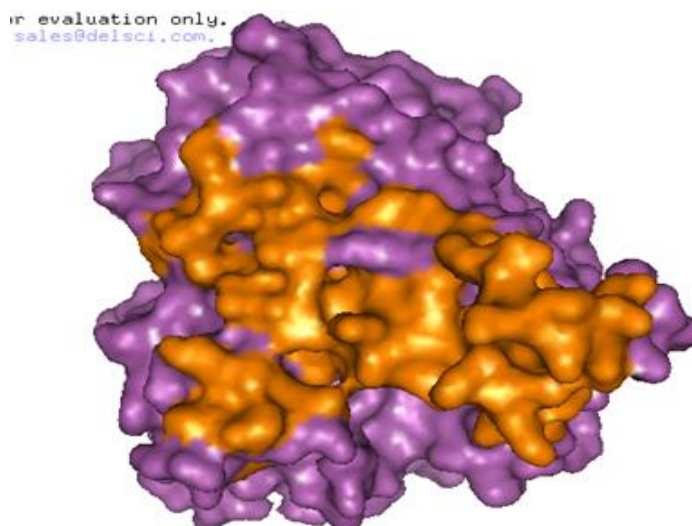


Fig.2.33. Dioxygenase Electrostatic Patches

Largest Positive Patch:

 PRO206 GLY7 PRO2 GLY193 ASN1 VAL180 TYR116 TYR179 ALA144 GLN141 LYS147
 GLN190 LYS182 LEU194 VAL192 THR108 GLU115 LYS168 LYS218 LYS202 LYS112
 SER109 THR191 LYS145 ILE89 PRO178 LYS177 LEU151 VAL148 SER10 GLN113
 ASP150 ARG104 SER152 LEU189 LYS11 LEU142 HIS207 GLY167 LYS107 GLN541
 VAL149 LYS6 VAL203 ALA542 LYS3 GLU185 ASN138 GLU540 THR9 LEU146 LYS199
 TYR153 VAL137 LEU8 PRO4 PHE117 LEU195 SER5

2.2.21.3. Cysteine disulphide bonds:

Table: 2.21 Cysteine Disulphide bonds

Disulfide bond scores			
Cysteine sequence position	Distance	Bond	Score
60 - 335	275	GHLPDCLNGEF-FELPNCFIFHN	0.01043
60 - 355	295	GHLPDCLNGEF-VVLITCRLENP	0.01948
335 - 355	20	FELPNCFIFHN-VVLITCRLENP	0.01043

Predicted bonds

60 - 355	GHLPDCLNGEF - VVLITCRLENP
Predicted connectivity	
1-3	

2.2.21.4. Structure pore lining residues:

PyMOL for evaluation only.
Contact sales@delsci.com.

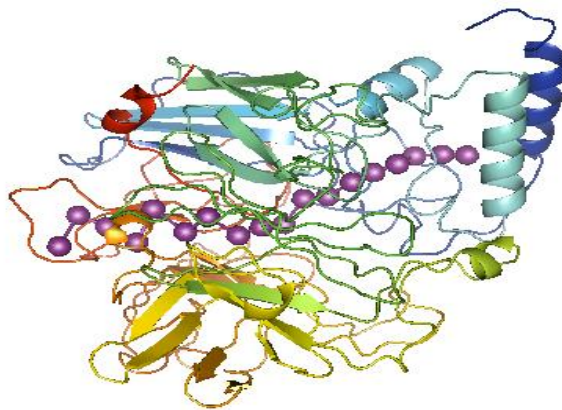


Fig.2.34. Dioxxygenase Structure Pore or Tunnel

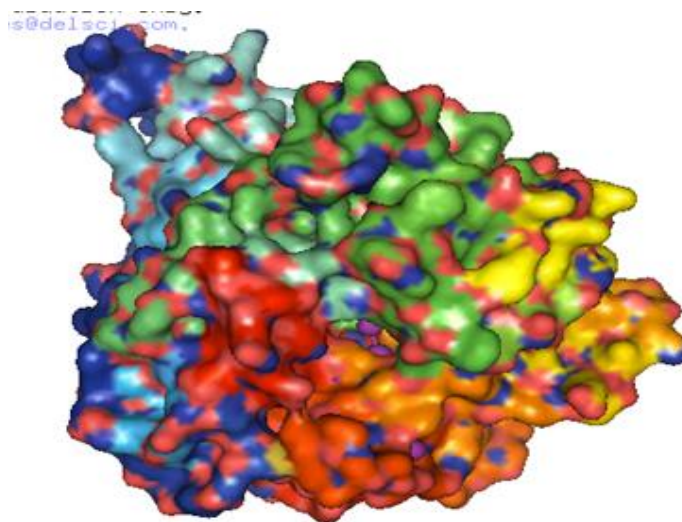


Fig.2.35. Dioxxygenase Structure Tunnel

Pore lining residues:

Pro30,Pro70,Phe74,Ala75,Pro76,Val77,Ala78,Gly79,Tyr80,Trp82,Phe83,Asp84,Gly85,Asp86,Met 88,
Ser109,Arg110,Glu114,Ala120,Lys121,Phe122,Met123,Lys124,Ile125,Asp127,Leu128,Lys129,Ph

e132, Met136, Gln140, Val149, His155, Thr157, Asn159, Thr160, Ala161, Ile163, His165, Leu172, Ser173, Glu174, Gly175, Asp176, Glu204, His220, Thr223, Ala224, **His225**, Pro226, Lys227, Val228, Asp229, Met235, Tyr240, Met271, **His273**, Asp274, Phe275, Ile277, Met285, Val297, Lys298, Glu299, Leu302, Ile303, Phe338, **His339**, Asn340, Ala341, Glu347, Leu357, Pro368, Phe376, Lys377, Leu380, Phe466, Ser468, Glu469, Phe472, Val473, Pro474, Arg475, Val476, Pro477, Glu482, Glu483, Asp484, Asp485, Gly486, Tyr487, Leu488, Glu495, Ala507, Thr509, Tyr526, Gly527, Phe528, **His529**, Ala530, Phe531, Phe532, Glu536, Gln53

2.2.21.4.1. Pore shape:



Fig.2.36. Dioxygenase Structure Pore Shape

Identified shape: UDSUD

Shape string legend:

D = conical frustum with decreasing diameter

U = conical frustum with increasing diameter

S = cylinder

Commonly recognized shapes:

DU = Hourglass

UD = Diamond

UDU/UDU = Hourglass-Diamond-Complex

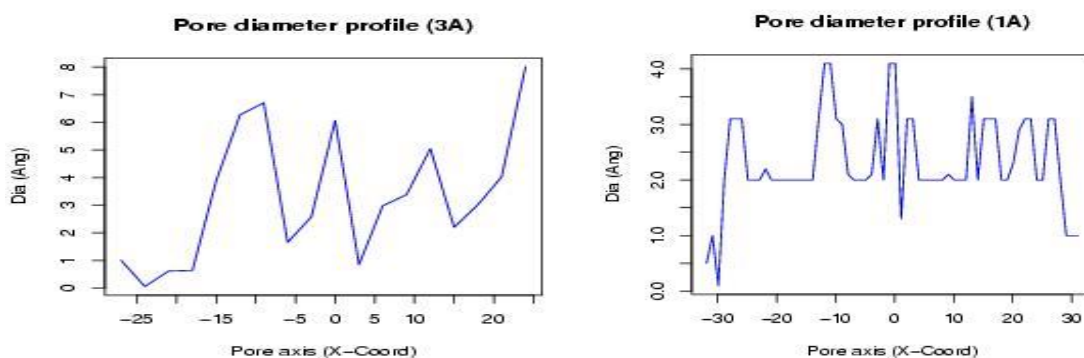


Fig. 2.37. Pore diameter

2.2.21.4.2. Pore visualization:

2.2.21.4.2.1. Protein orientation:

- along the pore axis (= x-axis) lowest coordinate at the bottom;
- pore centres are represented by red spheres at 3 Angstrom steps, and their radii are proportional and represent the corresponding measured diameters;
- Orange and blue areas show pore-lining atoms and corresponding residues, respectively.

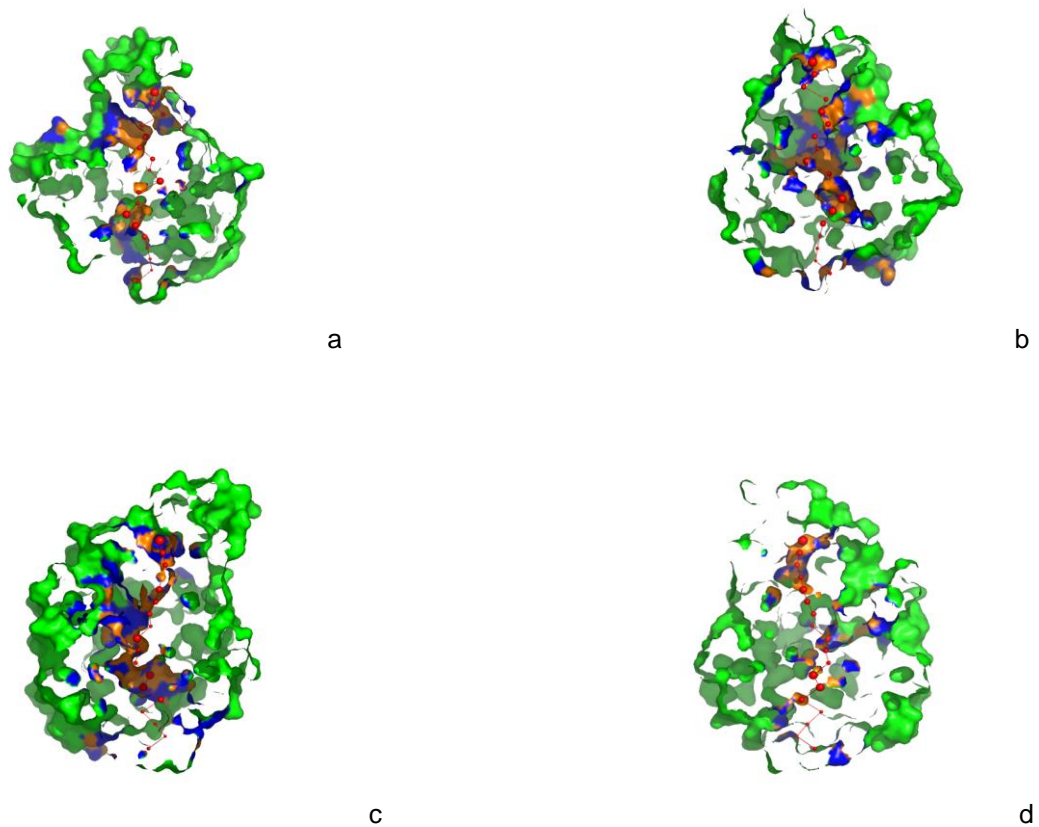


Fig.2.38. From left to right: a. XZ-plane section, $Y > 0$ coordinates only; b. XZ-plane section, $Y < 0$ coordinates only; c. XY-plane section, $Z > 0$ coordinates only; d. XY-plane section, $Z < 0$ coordinates only.



Fig.2.39. Identified pore-lining atoms only.

2.2. 21.4.3. Features of the cavity:

2.2.21.4.3.1. Protein orientation:

- along the pore axis (=x-axis) lowest coordinate at the bottom:
- pore centres are represent by red spheres at 1 Angstrom step along the pore axis.

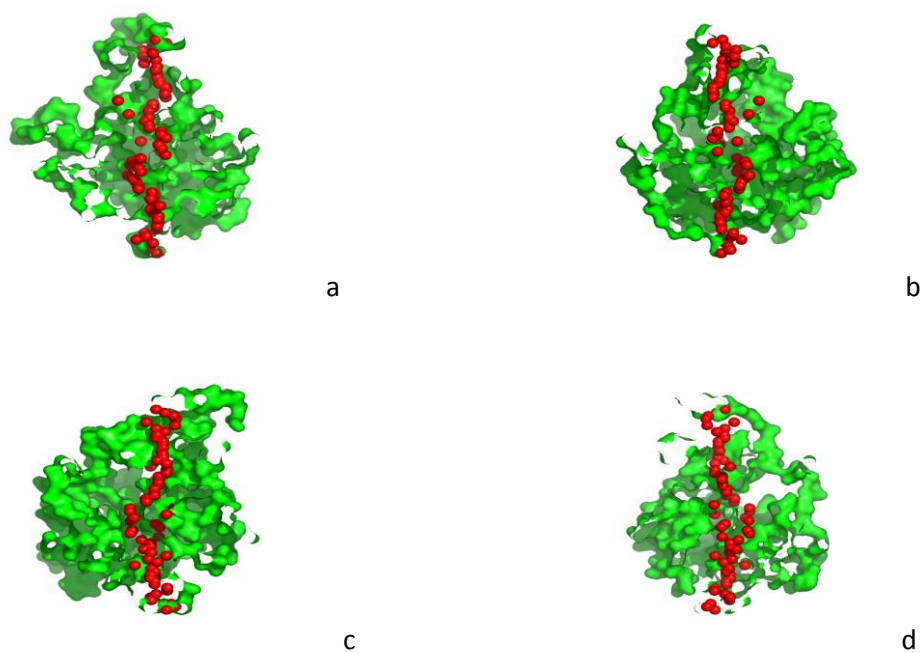


Fig. 2.40. a. Top left: XZ-plane section, $Y > 0$ coordinates only;
 b. Top right: XZ-plane section, $Y < 0$ coordinates only;
 c. Bottom left: XY-plane section, $Z > 0$ coordinates only;
 d. Bottom right: XY-plane section, $Z < 0$ coordinates only.



Fig, 2.41. Pore Shape

Number of found lines: 3

Points in straight areas: 26.92

Total RMSD: 3.71

2.2.21.5. Dioxygenase protein structure:

The dioxygenase structure contains various folds and transmembrane segments with both N- and C- terminals located in the intracellular side. The protein structure constitutes two distinct domains the N and C domains. The transmembrane segments consist of both α -helical domain and β -sheets (Messing *et al.*, 2010). The α -helical domain seems to be interacting with the membrane surface, while β -strands with medium size loops are discontinuous which can facilitate conformational changes during substrate transport as dioxygenases belong to the group of enzymes where substrate binding is a dynamic process which means, the substrates are channelized after binding to the dioxygenase enzyme for further processes. The overall structure of dioxygenase has a conformation facing outward and the substrates are possibly trapped within centre of the transmembrane domain completely or minimally occluded from the intracellular side through a channel or pore of 8-40Å diameter, while solvent is accessible from the extracellular side and it is not too wide for the substrate to escape. The structure model of dioxygenase with and without

substrate could add up template models to the collection of dioxygenase conformations (Messing *et al.*, 2010).

Consistent with other dioxygenase structure of known type, the strawberry dioxygenase structure comprises a β -propeller containing 7-8 antiparallel β -sheets and 8 α -helices two of them are interacting at N and C terminal, while one α -helix interact with the membrane surface. In the intracellular side mainly branched chain residues are present without forming a loop but are interacting with the transmembrane cytosolic helical domain by extensive polar interactions. The residues forming the intracellular helical domain are highly conserved in dioxygenases. Particularly, residues mediating inter-domain interaction are common in strawberry dioxygenase and maize VP14 (Messing *et al.*, 2010). On the basis of this analysis it can be predicted, that similar helical domain interacting with membrane surface and other cytoplasmic domains might exist in other dioxygenases as well and may be interacting with the transmembrane portion of the dioxygenases as observed in the strawberry dioxygenase structure.

2.2.21.5.1. Structure outline:

The Ramachandran plot analysis of strawberry dioxygenase structure model is also consistent with the maize VP14 (Messing *et al.*, 2010) and other α -ketoglutarate dependent dioxygenases, with somehow lower percentage of residues in most favored region, which is expected for a good quality structures of resolution of at least 2 Å with a R-factor value no greater than 20.0 and 26.7% in the additional allowed region, whereas 1.7% residues are in disallowed regions. The strawberry dioxygenase structure model is within acceptable geometry for a minimum 3.0 Å structure, with 65.9% residues in most favored regions, 26.7% residues in additional allowed regions, 5.6% residues in generously allowed region but very less residues 1.7% in disallowed regions of the Ramachandran plot. The number of residues such as Glycine is 42 and Proline is 34 and the end-terminal residues excluding

Glycine and Proline is 2. Also, in the structure model there are Glycine-Glycine repeats before and after Cysteine residues, indicative of hinge property of the structure.

Primary sequence alignment of dioxygenases shows that they share a highly conserved catalytic core from residues Asp150-Arg250 and Lys370-Thr480 and N- and C- terminal segments with divergent length and sequences. Other crystal structures of various dioxygenases have shown this conserved catalytic core domain which contains the full length of protein (Messing *et al.*, 2010; Fritze *et al.*, 2004). The overall structures of these conserved core domains are highly similar, displaying specific folds including the C terminal wd40 domain of human palb2 , Sec 13 with nup145c insertion blade, histone trimethyl lysine binder within the polycomb complex, nucleoporin pair nup85-seh1, small subunit ribosome localization, Ras like domain, Integrin receptor carboxy terminus like domains etc. The structure also displays a Rossmann fold domain for NAD⁺ binding and a domain for zinc-binding including several other domains. The N-terminal domain (Lys3-Val149) has apparently no catalytic function but there are specific substrate binding domains including Allergen and lipid binding domains, α -epoxycholesterol binding domain, Chorismate mutase binding domain, Focal adhesion kinase (FAK) binding domain and also Abscisic acid (ABA) binding domain etc. and C-terminal domain (Ser481-Gln539) and the core domain (Asp150-Arg250 and Lys370-Thr480) harbor the iron binding site (His225, His273, His339, His529) at the centre of the β -propeller and the catalytic residues in vicinity and the α -helix that presumably functions as a gate regulating access of the substrate to the active site. The β -propeller like structure consists of 8 antiparallel β -sheets containing 2-5 β -strands among total 26 β -strands in the structure. The first β -sheet contains 5 β -strands of topology 3-1-13, the second and third β -sheet contains 2 β strands each of topology 1, the fourth and the fifth β -sheet contains 4 β -strands each of topologies 1-12X and -1-1-1, the sixth β -sheet contains 3 β -strands and its topology is 1 1,

the seventh β -sheet contains 4 β -strands of topology 1 1 1 and the eighth β -sheet contains 2 β -strands with topology 1 depicted as letter A-H. Consistent with other structures containing β -propeller, the structure contains a long tunnel or pore surrounded by hydrophobic residues running across the centre of the structure from one end to the other. (Messing *et al.*, 2010) The pore centres at 1Å^o step at the pore axis, which is displayed in the pore-diameter plot and the pore representation, indicating that the pore has a definite shape identified as UDSUD, which signifies a combination of shapes such as U= conical frustum with increasing diameter, D=conical frustum with decreasing diameter, S=cylinder, whereas the commonly recognized shapes are DU=Hourglass, UD-Diamond and UDU/UDU= Hourglass-Diamond-Complex. The pore line residues are usually hydrophobic and interact with the other neutral, positive and negative residues in the transmembrane portion and in the membrane. The hydrophobic patches and the electrostatic patch residues are important both structurally as well as functionally because the dioxygenase activity which requires Fe²⁺ is located inside the tunnel or pore on the central axis of β -propeller (Messing *et al.*, 2010).

2.2.21.5.2. α -Helices:

Table:2.22 8 helices

No.	Start	End	Type	No. resid	Residue s per turn	Sequence
*1.	Ser5	Leu19	H	15	3.72	SKGLTSKLVDLVEK L
*2.	Pro30	Tyr33	H	4	4.39	PHHY
*3.	Ser109	Phe117	H	9	3.60	SRLKQEEYF
*4.	Gly133	Leu146	H	14	3.61	GLLMVNMQQQLRAK L
5.	Tyr197	Lys199	G	3	-	YDK
6.	Pro293	Met296	H	4	3.91	PKEM
7.	Asp324	Leu326	G	3	-	DLL
8.	Glu535	Glu540	G	6	3.12	EEQLQE

The structure consists of 8 α -helices, depicted as **H₁₋₈**. The first α -helix H₁ contains 15 residues (Ser5-Leu19), the second α -helix H₂ contains 4 residues (Pro30-Thr33), the third α -helix H₃ contains 9 residues (Ser109-Phe117), the fourth α -helix H₄ contains 14 residues (Gly133-Leu146), the fifth α -helix H₅ contains 3 residues (Tyr197-Lys199), the sixth α -helix H₆ contains 4 residues (Pro293-Met296), the seventh α -helix H₇ contains 3 residues (Asp324-Leu326), the eighth α -helix H₈ contains 6 residues (Glu535-Glu540). There is also helix-helix interaction between Helix 1 and Helix2 with 1-1 and 6-4 residues respectively.

2.2.21.5.2.1. Helix-helix interaction:

Table: 2.23 Helix-Helix Interactions

No.	Helices	Helix types	Distance (Å)	Angle (°)	Interaction		No. interacting residues		
					type		Total	Helix 1	Helix 2
1.	B1 B3	H H	14.2	-101.1	I	C	1	1	1
2.	B1 B4	H H	10.0	-166.6	N	C	10	6	4

2.2.21.5.2.2 β -sheets:

Table: 2.24 8 beta sheets

Sheet	No. of strands	Type	Barrel	Topology
A	5	Antiparallel	No	3 -1-13
B	2	Antiparallel	No	1
C	2	Antiparallel	No	1
D	4	Antiparallel	No	1-12X
E	4	Antiparallel	No	-1-1-1
F	3	Antiparallel	No	11
G	4	Antiparallel	No	111
H	2	Antiparallel	No	1

2.2.21.5.3.1. β -strands:

Table:2.25 26 Beta strands

No.	Start	End	Sheet	No. resid	Edge	Sequence
1.	Thr45	Thr48	A	4	Yes	TPPT
2.	Gly63	Arg67	A	5	No	GEFVR
3.	Met88	Lys95	A	8	No	MIHGMRK
4.	Lys98	Tyr105	A	8	No	KATYVSRY
5.	Leu162	His165	B	4	Yes	LIYH
6.	Lys168	Ala171	B	4	Yes	KLLA
7.	Leu181	Val183	C	3	Yes	LKV
8.	Leu189	Thr191	C	3	Yes	LQT
9.	Met235	Thr237	D	3	Yes	MFT
10.	Val247	Ser253	D	7	No	VMYRVVS
11.	Phe257	Met258	D	2	Yes	FM
12.	Val262	Ile264	D	3	Yes	VPI
13.	Ala276	Ile277	E	2	Yes	AI
14.	Tyr281	Asp286	E	6	No	YAIFMD
15.	Arg313	Pro318	E	6	No	RFGVLP
16.	Arg328	Phe330	E	3	Yes	RWF
17.	Trp344	Glu346	F	3	Yes	WEE
18.	Val351	Thr354	F	4	No	VLIT
19.	Met383	Phe385	F	3	Yes	MRF
20.	Val402	Val407	G	6	Yes	VDFPRV
21.	Phe418	Thr423	G	6	No	FVYGTT
22.	Val434	Asp437	G	4	No	VKFD

23.	Gln456	Leu458	G	3	Yes	QGL
24.	Phe466	Pro474	H	9	Yes	FGSEAlFVP
25.	Gly486	His493	H	8	Yes	GYLIFFVH
26.	Val533	Thr534	A	2	Yes	VT

The structure consists of 26 β -strands, comprising 8 β -sheets depicted as (A-H) in the secondary structure. The first β -sheet (A) contains 5 β -strands (1, 2, 3, 4, 26 of 4, 5, 8, 8 and 2) residues, while the first β -strand (Thr45-Thr48) and 26th β -strand (Val533-Thr534) have edges but other β -strands 2nd β -strand (Gly63-Arg67), 3rd β -strand (Met88-Lys95) and 4th β -strands (Lys98-Tyr105) are without edges. The second β -sheet (B) contains 2 β -strands (5, 6 of 4) residues each, both the 5th β -strand (Leu162-His165) and 6th β -strand (Lys168-Ala171) have edges. The third β -sheet (C) contains 2 β -strands (7, 8 of 3) residues each, both the 7th β -strand (Leu181-Val183) and 8th β -strand (Leu189-Thr191) have edges. The fourth β -sheet (D) contains 4 β -strands (9,10,11,12 of 3,7,2,3) residues, the 9th β -strand (Met235-Thr237), 11th β -strand (Phe257-Met258), 12th β -strand (Val262-Ile264) have edges but the 10th β -strand (Val247-Ser253) is without edge. The fifth β -sheet (E) contains 4 β -strands (13, 14, 15, 16 of 2, 6, 6, 3) residues, the 13th β -strand (Ala276-Ile277), and 16th β -strand (Arg328-Phe330) have edges but 14th β -strand (Tyr281-Asp286) and 15th β -strands (Arg313-Pro318) are without edges. The sixth β -sheet (F) contains 3 β -strands (17, 18, 19 of 3, 4, 3) residues; both the 17th β -strand (Trp344-Glu346) and 19th β -strand (Met383-Phe385) have edges, whereas the 18th β -strand (Val351-Thr354) is without edge. The seventh β -sheet (G) contains 4 β -strands (20,21,22,23 of 6,6,4,3) residues, both the 20th β -strand (Val402-Val407) and 23rd β -strand (Gln456-Leu458) have edges, but the 21st β -strand (Phe418-Thr423) and 22nd β -strands (Val434-Asp437) are without edges. The eighth β -sheet (H) contains 2 β -strands (24, 25 of 9, 8) residues and both the 24th β -strand (Phe466-Pro474) and 25th β -strands (Gly486-His493) have edges.

2.2.21.5.3.2. β -hairpins:

Table:2.26 15 beta hairpins

No.	Strand 1			Strand 2			Hairpin class
	Start	End	Length	Start	End	Length	
* 1.	Gly63	Arg67	5	Met88	Lys95	8	22:22
* 2.	Met88	Lys95	8	Lys98	Tyr105	8	2:2 IP
* 3.	Leu162	His165	4	Lys168	Ala171	4	2:2 IP
* 4.	Leu181	Val183	3	Leu189	Thr191	3	3:5
* 5.	Met235	Thr237	3	Val247	Ser253	7	12:12
6.	Val247	Ser253	7	Phe257	Met258	2	3:3
7.	Ala276	Ile277	2	Tyr281	Asp286	6	2:4
8.	Tyr281	Asp286	6	Arg313	Pro318	6	26:26
9.	Arg313	Pro318	6	Arg328	Phe330	3	11:11
10.	Trp344	Glu346	3	Val351	Thr354	4	4:4 I
11.	Val351	Thr354	4	Met383	Phe385	3	28:28
12.	Val402	Val407	6	Phe418	Thr423	6	9:11
13.	Phe418	Thr423	6	Val434	Asp437	4	16:16
14.	Val434	Asp437	4	Gln456	Leu458	3	17:19
15.	Phe466	Pro474	9	Gly486	His493	8	9:11

The dioxygenase structure contains around 15 β -hairpin segments of which five β -hairpins are important. The first β -hairpin is of class 22:22 between strand 1 (Gly63-Arg65) and strand2 (Met88-Lys95), the second β -hairpin is of class 2:2 between strand1 (Met88-Lys95) and strand2 (Lys98-Tyr105), the third β -hairpin is of class 2:2 between strand1 (Leu162-His165) and strand2 (Lys168-Ala171), the fourth β -hairpin is of class 3:5 between strand1(Leu181-Val183) and strand2 (Leu189-Thr191) and the fifth β -hairpin is of class 12:12 between strand1 (Met235-Thr237) and strand2 (Val247-Ser253).

2.2.21.5.3.3. β -turn:

Table:2.27 58 beta turns

No.	Turn	Sequence*	Turn type	H-bond
* 1.	Tyr25-Ser28	YDSS	I	Yes
* 2.	Asp26-Gln29	DSSQ	IV	
* 3.	Ala35-Phe38	AGNF	IV	
* 4.	Gly36-Thr39	GNFT	I	Yes
* 5.	Pro58-Leu61	PDCL	I	Yes
6.	Pro76-Gly79	PVAG	IV	
7.	His81-Asp84	HWFD	I	Yes
8.	Trp82-Gly85	WFDG	I	
9.	Lys95-Lys98	KDGK	I'	Yes

10.	Lys124-Asp127	KIGD	IV	
11.	Ile125-Leu128	IGDL	IV	
12.	Gly126-Lys129	GDLK	IV	
13.	Asp127-Gly130	DLKG	I	Yes
14.	Asp150-Tyr153	DLSY	IV	
15.	Leu151-Gly154	LSYG	IV	
16.	Tyr164-Gly167	YHHG	IV	
17.	His165-Lys168	HHGK	I'	Yes
18.	Leu184-Gly187	LEDG	I	Yes
19.	Lys199-Lys202	KRLK	IV	
20.	Lys202-Lys205	KVEK	II	
21.	His207-Asp210	HVTD	IV	Yes
22.	Val208-Asp211	VTDD	IV	
23.	Lys218-Ser221	KLHS	II'	Yes
24.	Leu219-Phe222	LHSF	VIII	
25.	Thr223-Pro226	TAHP	IV	
26.	Asp229-Thr232	DPFT	IV	
27.	Pro230-Gly233	PFTG	I	
28.	His242-Pro245	HDPP	IV	
29.	Asp243-Tyr246	DPPY	IV	
30.	Pro245-Met248	PYVM	VIII	
31.	Ser253-Gly256	SKDG	IV	Yes
32.	Thr278-Tyr281	TENY	I	
33.	Ser305-Glu308	SFDE	IV	
34.	Asp307-Lys310	DETK	I	Yes
35.	Pro318-Ala321	PRYA	IV	Yes
36.	Leu332-Cys335	LPNC	IV	
37.	Glu346-Glu349	EEDF	IV	
38.	Glu347-Val350	EDEV	I	Yes
39.	Met364-Gly367	MVNG	VIII	
40.	Lys370-Leu373	KKKL	IV	Yes
41.	Thr375-Asn378	TFKN	IV	Yes
42.	Asn386-Thr389	NLKT	I	Yes
43.	Leu387-Gly390	LKTG	IV	
44.	Asn408-Tyr411	NESY	I	
45.	Glu409-Thr412	ESYT	I	Yes
46.	Tyr411-Arg414	YTGR	II	Yes
47.	Leu424-Ile427	LDSI	I	
48.	Ser426-Lys429	SIK	IV	
49.	Gly445-Lys448	GKTK	IV	
50.	Glu450-Gly453	EVGG	IV	
51.	Val451-Asn454	VGGN	IV	
52.	Gly462-Arg465	GPGR	IV	Yes
53.	Pro477-Thr480	PGIT	IV	
54.	Glu482-Asp485	EEDD	I	
55.	Asp494-Thr497	DENT	IV	
56.	Glu495-Gly498	ENTG	IV	Yes
57.	Asp506-Thr509	DAKT	I	
58.	Ala507-Met510	AKTM	I	

The dioxygenase structure contains 58 β -turns (table) of which first 5 β -turns, which are type I and type IV β -turns with hydrogen bonding between involved residues. The first β -turn (Tyr25-Ser28) is type I, the second β -turn (Asp26-Gln29) is type-IV β -turn, the third β -

turn(Ala35-Phe38) is also type IV β -turn, and both the fourth β -turn (Gly36-Thr39) and the fifth β -turn (Pro58-Leu61) are type I β -turn.

2.2.21.5.3.4. β -bulges:

Table: 2.28 6 beta bulges

Bulge type	Res X	Res 1	Res 2
Antiparallel wide	Val102B	Pro46B	Pro47B
Antiparallel G1	Leu172B	Thr160B	Ala161B
Antiparallel G1	Leu184B	Gly187B	Asp188B
Antiparallel G1	Ser253B	Gly256B	Phe257B
Antiparallel classic	Thr422B	Asp403B	Phe404B
Antiparallel classic	Phe491B	Ser468B	Glu469B

There are three kinds of β -bulges in the dioxygenase structure- classic, G1, and wide; but all the 6 β -bulges are anti-parallel. The first β -bulge (Val102,Pro46,Pro47) is antiparallel wide, three β -bulges which are of anti-parallel G1 type are the second β -bulge (Leu172,Thr160,Ala161), the third β -bulge (Leu184,Gly187,Asp188) and the forth β -bulge (Ser253,Gly256,Phe257), whereas the fifth β -bulge (Thr422,Asp403,Phe404) and the sixth β -bulge (Phe491,Ser468,Glu469) are of type anti-parallel classic.

2.2.21.5.3.5. γ -(gamma) turn:

31 gamma turns:

Table:2.29 31 gamma turns

No.	Start	End	Sequence*	Turn type
* 1.	Val41	Glu43	V I E	CLASSIC
* 2.	Val106	Thr108	V K T	INVERSE
* 3.	Ile125	Asp127	I G D	CLASSIC
* 4.	Gln217	Leu219	Q K L	INVERSE
* 5.	Thr223	His225	T A H	INVERSE
6.	Pro244	Tyr246	P P Y	INVERSE
7.	Ile270	Met272	I M M	INVERSE
8.	Pro288	Tyr290	P L Y	INVERSE
9.	Tyr290	Lys292	Y F K	INVERSE

10.	Lys301	Ile303	K L I	INVERSE
11.	Ser305	Asp307	S F D	CLASSIC
12.	Phe306	Glu308	F D E	INVERSE
13.	Arg319	Ala321	R Y A	INVERSE
14.	Phe330	Leu332	F E L	INVERSE
15.	Thr354	Arg356	T C R	INVERSE
16.	Asp361	Asp363	D L D	INVERSE
17.	Val365	Gly367	V N G	INVERSE
18.	Gly367	Ile369	G P I	INVERSE
19.	Asp374	Phe376	D T F	INVERSE
20.	Thr375	Lys377	T F K	INVERSE
21.	Ala392	Gln394	A T Q	INVERSE
22.	Arg414	Gln416	R K Q	INVERSE
23.	Ser426	Ala428	S I A	CLASSIC
24.	Ala441	Glu443	A P E	INVERSE
25.	Lys448	Glu450	K I E	CLASSIC
26.	Val451	Gly453	V G G	CLASSIC
27.	Gly462	Gly464	G P G	INVERSE
28.	Pro477	Ile479	P G I	CLASSIC
29.	Gly478	Thr480	G I T	INVERSE
30.	Ala516	Val518	A V V	INVERSE
31.	Val518	Leu520	V E L	INVERSE

There are 31 γ -turns in the dioxygenase structure, of which the first 5 γ -turns are classic or inverse, which seem to be important. The first (Val41-Glu43) and third γ -turns (Ile125-Asp127) are classic, whereas the second (Val106-Thr108), fourth (Gln217-Leu219), fifth (Thr223-His225) are of inverse type. The other γ -turn the seventh γ -turn (Ile270-Met272) and the fifth γ -turn (Thr223-His225) are important since both these γ -turns end with the residues His273 and His225 respectively that are involved in the iron bound coordination site.

2.2.21.5.4. Loops and Kinks:

2.2.21.5.4.1. Large loop modeling:

The dioxygenase structure has unusually large loops (Adhikari *et al.*, 2011) (more than 25 residues long stretch) that links α -helix and β -strand (α - β loop), β -strand and α -helix (β - α loop) and two β -strands (β - β loop) and this is structurally consistent with other

dioxygenase structures (Messing *et al.*, 2010) as most of them are transmembrane containing hydrophobic residues which interact with other neutral, positive and negative amino acids, which is apparent from the electrostatic and hydrophobic surface patches on the dioxygenase structure. Energetically loops of less than 10-15 residues are favorable but longer loops of 25 amino acid residues are functionally representative, which could be solvent exposed or buried partially and linking various structure portions (Jamroz *et al.*, 2010; Adhikari *et al.*, 2011). These loops may interact with each other, which can have some influence on the performance of the enzyme activity when a substrate is incorporated into the active site and substrate binding site as they carry some flexible residues which shift to a small distance or distally depending on the size of the substrate. In the human dioxygenases e.g. tryptophan-1,3-dioxygenase and indoleamine-2,3-dioxygenase (Zhang *et al.*, 2007), such sifting is observed on substrate binding, which are comparatively larger than that of plant dioxygenases. These large loops can carry residues involved in holding the Fe²⁺ required for dioxygenase activity, which lie at the centre of the long tunnel in such kinds of dioxygenases such as maize-VP14 (Messing *et al.*, 2010). Since, the presence of longer loops in the dioxygenase structure is consistent with the structure of maize dioxygenase maize-VP14 (Messing *et al.*, 2010), it is possible to assume that the longer loops present in the structure is particularly important as loops often regulate protein functions, for example, loops function as molecular recognitions of binding sites and are involved in executing protein functions. The active sites and binding pockets mediate the specificity of protein interaction because of the local conformations of protein structure. The loops end at correct anchor positions, however, energy functions for the conformational modifications would be inadequate as the crystal structure itself may not show electron density for certain amino acid residues or differentiating O from NH₂ and also after aligning the structure without the loop regions and for the loop residues RMSD is

calculated. The best orientation of the model of the crystal in the unit cell has a Z-score of value 0.65 (which is a raw score) and -2.25. The Z-score value is more to the positive side as positive Z-score is considered as reliable for good quality models, however for phase modifications Z-score value <7 are unrefinable while >7 are considered as refinable (Adhikari *et al.*, 2011; Shehu *et al.*, 2012). Thus, structures refined using different constraints could indicate apparently large deviations from normal structures. Therefore, longer loops such as Val297-Arg313 (α - β), Asp494-Glu535 (β - α) and Thr423-Arg465 (β - β), which also contain β -hairpins, β and γ -turns are predicted to be important for folding pathways in these loops, which can be improved by larger RMSD for better resolution up to 9.2\AA to 6.2\AA , however overall template based structure model seem to be good for $3\text{-}5\text{\AA}$ structure. The standard $\beta\alpha\beta\alpha$ rule is followed in the region $\beta 8\alpha 5\beta 9\alpha 6$ [Tyr281-Asp286, Pro293-Val297, Arg313-Pro318, Asp 324-Leu326] otherwise $\beta\beta\alpha$, $\alpha\alpha\beta$, and $\beta\alpha\beta\beta$, $\alpha\alpha\beta\beta$ patterns are normally present in the structure.

2.2.21.5.4.2. Kinked α -helix:

There is a kink in the 7th α -helix, (Glu409-Tyr411), because of the serine residue (Ser410), which is also important as serine, threonine and cysteine residues bend α -helices in the $\mathbf{X}_1=\mathbf{g}^-$ conformations and this is particularly significant for membrane proteins, in the transmembrane α -helix insertion of this small bending angle at one side of the cell membrane causes a displacement of the residues significantly from their positions at the other side of the membrane. Thus, it is hypothesized that Ser and Thr residues because of their local alterations (rotamer configurations) can cause throughout significant conformational changes across transmembrane helices thus these residues play important roles in the molecular mechanisms of transmembrane signaling which are underlying. On the structural basis these observations can provide input to understand the possible impact of amino acid residues such as Ser on the conformational equilibrium between inactive and

active state structures of the dioxygenase enzyme, similar effects have been observed in neurotransmitter family of G- protein coupled receptors (GPCRs); therefore significance of the α -helix kink created by Ser residue can be further investigated by implementing the fact on various substrate interacting with dioxygenase (Ballesteros *et al.*, 2000).

2.2.21.5.5. Coordination:

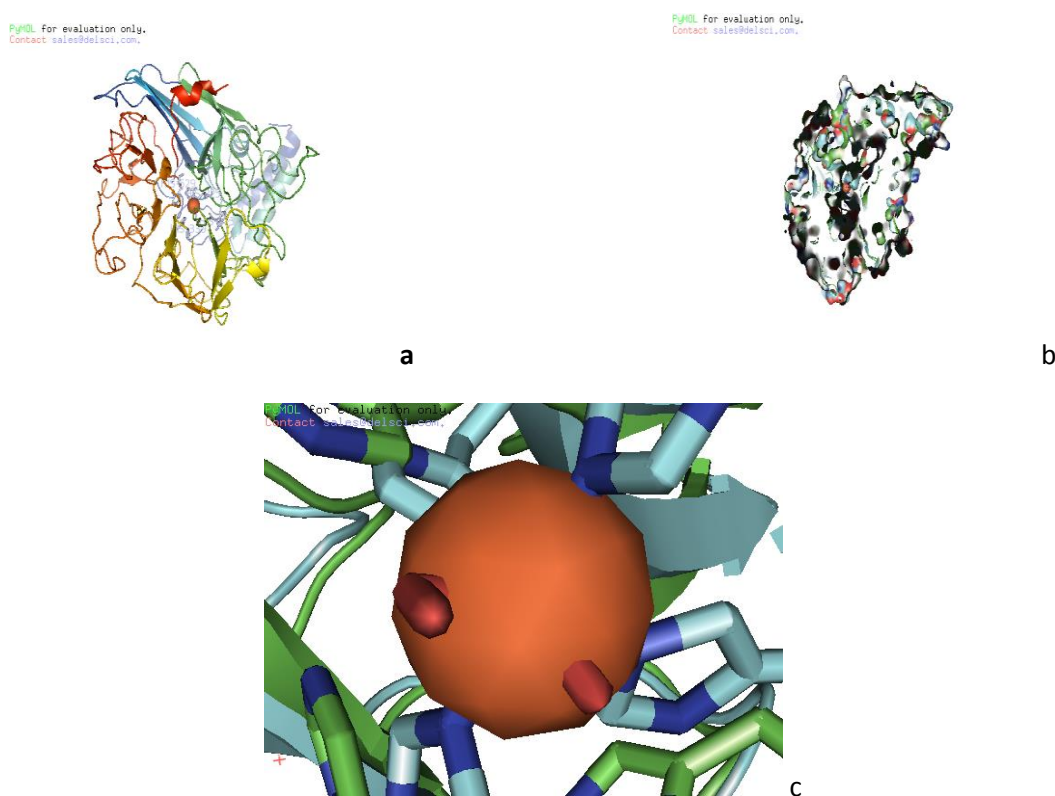


Fig.2.42. Dioxygenase Structure Coordination

The structure contains a long tunnel surrounded by β -sheets runs from one end to the other crossing the centre, which is consistent with other β -propeller structures. The dioxygenase activity requires a Fe²⁺ ion which is located inside the tunnel on the axis central to the β -propeller; this metal ion is octahedrally bound by four His residues that are positioned on the long loops. The active site is located in the C-terminal domain, including the core domain comprised of transmembrane segments (2), α -helices (3) and β -sheets (4-5), that surrounds a pronounced cavity or tunnel or pore which is between 8Å^o-40Å^o in width at

3Å° and 1Å° resolution respectively. A water molecule and a molecule of dioxygen, which occupy positions below and above the plane which is occupied by water molecules and other four His residues. The four His residues His225, His273, His339 and His529 are located on loops which are connecting α -helices, and β -strands and bound to the Fe²⁺ octahedrally (Messing *et al.*, 2010). The α -helices that are surrounding the active site includes H₁,H₄ and H₈, while the anti-parallel β -sheets which are surrounding the active site are A,B,D,E,F,G and H which includes β -strands β ₁₂, β ₁₃, β ₁₄, β ₁₇, β ₂₀, β ₂₁ and β ₂₆. The anti-parallel β -sheets, which comprises 3-8 β -strands form β -propeller like structure, surrounding a tunnel, where Fe²⁺ is located by four His residues, which are positioned at the innermost side of the long loops and the positions of His residues are supported by other residues on β -strands and α -helix. The C-terminal α -helix H₈ and other α -helices H₁ and H₄ and H₅ jointly form an α -helical domain on top of the β -propeller, the α -helix H₄ which is also a transmembrane, is interacting with the membrane surface and other hydrophobic and positively or negatively charged residues in the membrane. The N-terminal provides the outermost β -strands (β ₁₋₄) and the C-terminal provides innermost β -strand (β ₂₆) to the β -sheet A, whereas β -sheets B, D, E, F and G are in the core domain, which contains the length of the protein structure. These long loops which hold the Fe²⁺ at its position contains β -turns, γ -turns and β -hairpins, including β -bulges. The active site residue His225 is located on a loop between α -helix H₅ and β -strand β ₉ (β -sheet D) containing β -turn β ₂₅ (Thr223-Pro226) and γ -turn γ ₅ (Thr223-His225), surrounded by β -strands β ₂₀, β ₂₁ and loops between β -sheets E, F, F, G and G, G, which also contains β -hairpins BH11(Val351-Thr354 and Met383, Phe385), BH14 (Val434-Asp437 and Gln456-Leu458) β -turns β ₄₉ (Gly445-Lys448) and γ -turns γ ₂₄ (Ala441-Glu443). The residue His273 is located on loop between β -sheets D and E which comprises β -strands β ₁₂ and β ₁₃, the loop contains γ -turn γ ₇ (Ile270-Met272) the γ -turn ends with the His273 which is

bound to the Fe^{2+} . The active site residue His 339 is located on the loop connecting β -strands $\beta 16$ and $\beta 17$, and β -sheets E and F, surrounded by β -strands $\beta 14$, $\beta 20$ and $\beta 21$ and other loops which contain β -turn BT36 (Leu332-Cys335), BT37(Glu346-Glu349) and γ -turn $\gamma 14$ (Phe330-Leu332). The fourth active site residue His529 is located on the loop between β -sheets H and A, which connects β -strands $\beta 25$ and $\beta 26$, the loop on which His529 is positioned contains 4- β -turns (BT55-58, Asp494-Thr497, Glu495-Gly498, Asp506-Thr509, Ala507-Met510) and γ -turns $\gamma 30$ (Ala516-Val518) and $\gamma 31$ (Val518-Leu520), surrounded by β -sheets B, β -strands $\beta 5$, $\beta 6$, $\beta 20$, $\beta 21$ and α -helices H_1 , H_4 and H_8 .

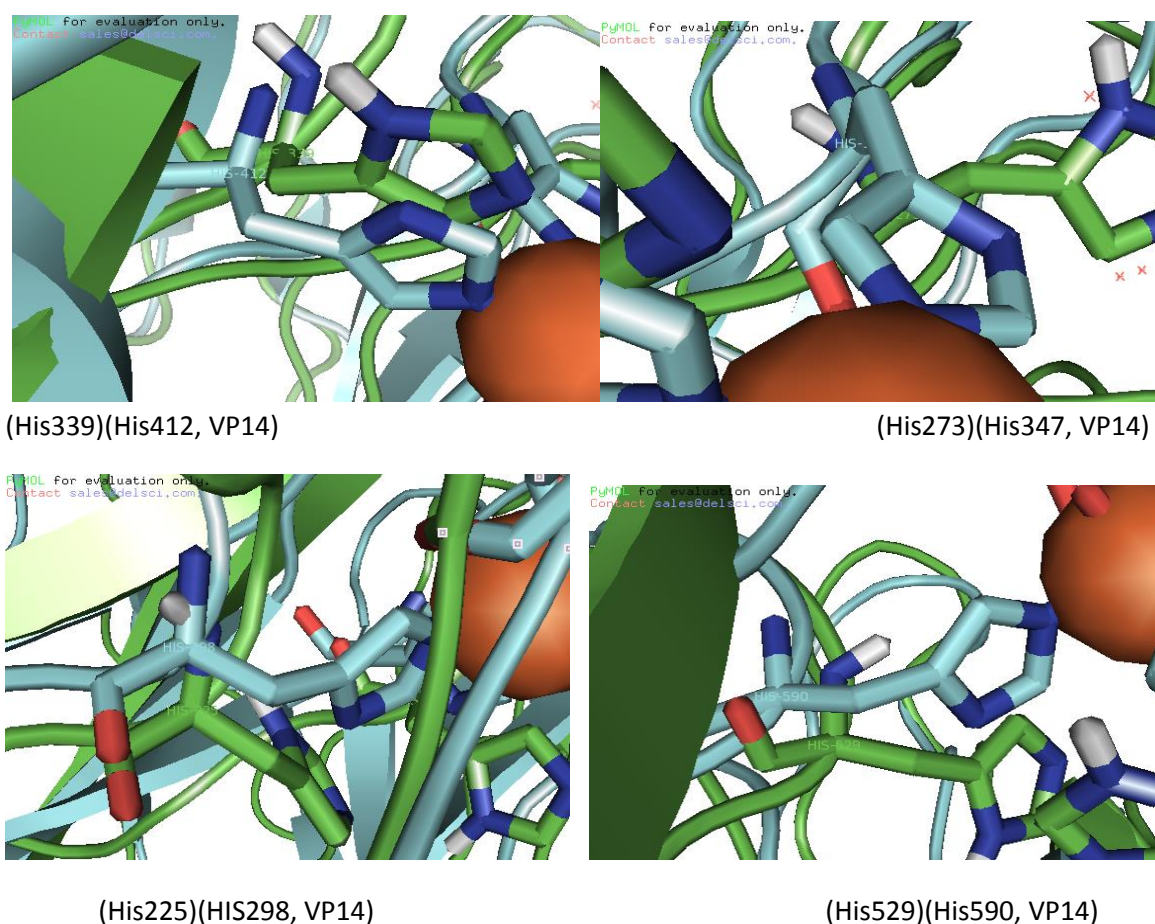


Fig.2.43. Dioxygenase Active site residues: **a.** His339; **b.** His273; **c.** His225; **d.** His529

Only a few enzymes for example superoxide reductase, photosystem II and fumarate reductase and have this type of Fe^{2+} coordination (Messing *et al.*, 2010).

2.2.21.5.6. Comparison:

Human indoleamine-2,3-dioxygenase, tryptophan-2,3-dioxygenase and plant dioxygenase maize-VP14 and strawberry dioxygenase active sites:

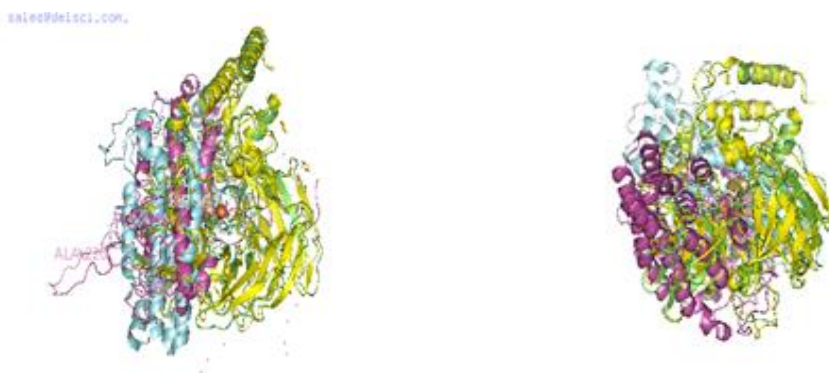


Fig. 2.44. Strawberry dioxygenase-maize VP14, Human tryptophan-2,3-dioxygenase and indoleamine 2, 3-dioxygenase

Initial comparison of plant dioxygenases such as strawberry dioxygenase and maize VP14 (Messing *et al.*, 2010) indicates that both types of dioxygenases are structurally similar; however there are sequence variation, which does not seem to be identical. Human tryptophan-2,3-dioxygenase and indoleamine-2,3-dioxygenase (Zhang *et al.*, 2007) are membrane proteins and have 6-7 transmembrane domains, the superimposed structure models of strawberry dioxygenase and maize VP-14 indicate that they are almost structurally identical with conserved active site residues, that includes four His residues octahedrally bound with the Fe^{2+} , which is required for enzyme activity. The active site residues in the maize-VP14 are His298, His347, His 412 and His590, while strawberry dioxygenase has coordination site involving His225, His273, His339 and His529. These His residues in the strawberry dioxygenase are located on long loops, which are supported by other residues on β -sheets and α -helices, as in maize-VP14, the β -propeller structure is similar comprising 7-8 β -blades and the His residues located on the innermost strands of different blades. Both the strawberry dioxygenase and maize-VP14 structures contain

along with the catalytic iron (Fe^{2+}), a water molecule and a molecule of dioxygen which could be represented by an elongated density.

Both the structures contain transmembrane segments and α -helical domain interacting with membrane surfaces and hydrophobic patches comprising residues which interact with other neutral, positive and negative residues in the membrane. The oxidative degradation of Trp to N-formyl kynurenine is catalyzed by human enzymes e.g. tryptophan-2,3-dioxygenase and indoleamine-2,3-dioxygenase, the superimposed structures show high similarity whereas the mechanisms for dioxygenase activity is different as some conserved residues such as His and Trp bind to the heme bound ligand in different ways. Overall, the superimposed structures of strawberry dioxygenase, maize VP14, human tryptophan-2,3-dioxygenase and indoleamine-2,3-dioxygenase show that the β -propeller structure of strawberry dioxygenase and maize VP14 are interacting with the α -helical transmembrane domains of human tryptophan-2,3-dioxygenase and indoleamine-2,3-dioxygenase indicating that strawberry dioxygenase is a membrane integral protein and the α -helical domains of strawberry dioxygenase and maize VP14 interact with the membrane surface and act as a gate for the substrate for their entry to the active site.



a

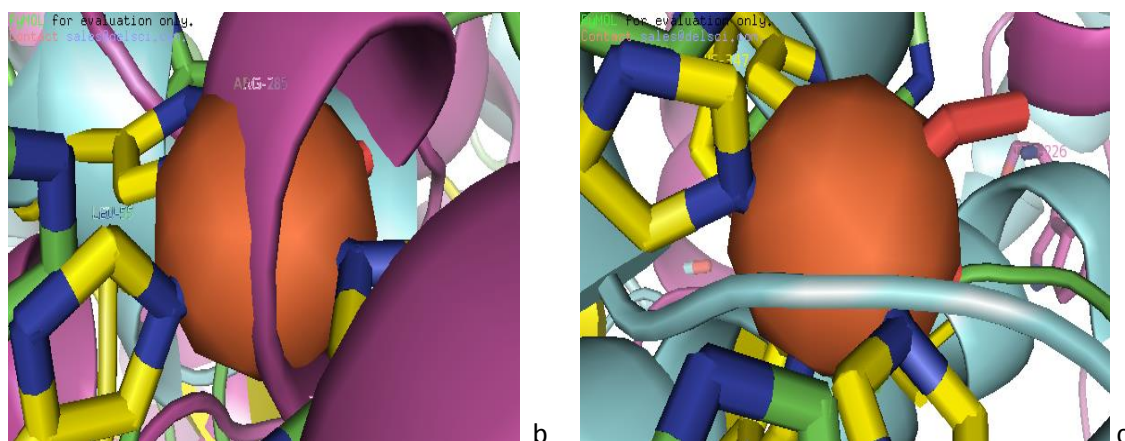


Fig.2.45. Active site comparison of Strawberry Dioxygenase, Maize VP14, Human Tryptophan-2,3- dioxygenase and Indoleamine-2,3-dioxygenase

The plant dioxygenases and human dioxygenases vary in the coordination as the coordination site is different presumably because of the structural difference as plant dioxygenases, strawberry dioxygenase and maize VP14 active sites are located on the central axis of a long tunnel that runs from one end to the other within protein structure, whereas the human dioxygenases, tryptophan-2,3-dioxygenase and indoleamine-2,3-dioxygenases are membrane proteins and that their structures consist of 6-7 transmembranes and the active sites lie in between transmembrane α -helical domains, such that Trp226 is located on the loop connecting α -helix (α_1), Gly275 on α -helix (α_2), Thr340 on α -helix (α_3), in the Indoleamine-2,3-dioxygenase (IDO) whereas, Tyr324, and Thr328 are located on α -helix (H_3), Tyr149 and Phe158 on loop (L_1), His192 on loop (L_2), Gln101 on α -helix (H_2), in the human tryptophan-2,3-dioxygenase(TDO) and the surrounding residues Lys366, Leu367 and Glu368 on a loop between β -strand (β_1) and α -helix (α_a) in the maize-VP14 which also support the His residues bound to the Fe^{2+} , and the His204 on α -Helix (H_1) in the Indoleamine-2,3-dioxygenase (IDO) supports the coordination site. The human Tryptophan-2, 3-dioxygenase α -helix (α_2) residue Arg285 coincides the iron sphere and the Indoleamine-2, 3-dioxygenase loop (L_1) residue Tyr149 interacts with the

Tryptophan-2,3-dioxygenase α -helix (α_2) residue Arg285, similarly Leu55 of TDO α -helix H₃ interact with Met295 of maize VP14 located on the loop (L₁) that ends on His298, the active site residue in the maize VP14. The Asn 229 located on loop of IDO interacts with Asn115 located on α -helix H₄ of TDO, similarly Phe117 located on α -helix H₄ of TDO interacts with Leu232 located on the loop that connects Trp226 with the α -helix (α_1). These are certain interactions, which support that the coordination sites of plant and human dioxygenases are very similar but shifted and located on the axis of α -helical domain in the case of human tryptophan-2,3-dioxygenase and indoleamine-2,3-dioxygenase, which is different from plant dioxygenases such as strawberry dioxygenase and maize VP14 as these are structures consists of a β -propeller structure and two α -helical domains over the top of the β -propeller, which may interact with the membrane surface while the α -helical domain acts as a gate for substrates regulating their entry to the active site.



Fig.2.46. Active Site Residues: Strawberry dioxygenase-maize VP14, human tryptophan-2,3-dioxygenase and indoleamine-2,3-dioxygenase

The active site residues among strawberry dioxygenase, maize VP14, human tryptophan-2,3-dioxygenase and indoleamine 2,3-dioxygenase are conserved in the sense that the catalytic Fe²⁺ is bound to four His residues in the plant dioxygenases, strawberry dioxygenase (His225, His273, His339, His529) and maize VP14 (His298, His347, His412,

His590), surrounded by Glu, Asp, Trp, Tyr, Asn and Gln residues but in the human indoleamine-2,3-dioxygenase (IDO), the His residues are replaced by Trp226, Thr340 and Gly275, whereas by Gln101, His192, Tyr324, Thr328, Tyr149 and Phe158 in the tryptophan-2,3-dioxygenase (TDO) (Zhang *et al.*, 2007).

Moreover, the superimposed structures indicate that the β -propeller structures of plant dioxygenases and the transmembrane domains of human tryptophan-2,3-dioxygenase and indoleamine -2,3-dioxygenase have some of the domains in common, which is consistent with the currently available structures of this group of proteins (Messing *et al.*, 2010; Zhang *et al.*, 2007). However, the superposition also proves that the plant dioxygenase e.g. strawberry dioxygenase is a membrane integral protein, which could be monotopic as well as membrane integral with a declination of 46.6° from the membrane surface and as these groups of enzyme continuously channelize the substrates, the orientation of dioxygenase on the membrane surface is playing a significant part in the dynamic process of substrate binding with the strawberry dioxygenase. The transmembrane α -helical domains of human tryptophan-2, 3-dioxygenase and indoleamine-2, 3-dioxygenases are also dynamic in the membrane (Zhang *et al.*, 2007). The possible interactions with other proteins depend on the type of domains present on the dioxygenase structure as the superimposed structures of strawberry dioxygenase, maize-VP14, human tryptophan-2,3-dioxygenase and indoleamine-2,3-dioxygenase indicates that strawberry dioxygenase presumably interacts with other membrane bound proteins and certain receptors and receptor binding proteins which are involved in specific signal transduction mechanisms coordinating specific signaling pathways.

2.2.21.5.7. Substrate binding pockets:



Fig.2.47. Dioxigenase-substrate binding pockets

There are 10 substrate binding pockets identified in the dioxigenase structure that comprise amino acid residues located on α -helices, β -sheets and long loops. The substrate binding pockets in the dioxigenase structure are different sites in the structure formed by the secondary structures including α -helices, β -strands, β -turns and β -hairpins; the surface structure shows them as grooves and pores surrounded by specific residues. The substrate binds to amino acid residues in such binding pockets and the substrate binding pocket rearranges accordingly to accommodate larger or smaller substrates.

The dioxigenase substrate binding pockets presumably fit to accommodate different substrates containing aromatic rings such as phenyl rings or dioxan rings and alkyl or exposed epoxy side chain compounds.

(1) SBP1 : LMYDSS(23-28)HYLA(32-35)F(38)P(58)CLN(60-62)G(79)HWF(81-83)IG(125-126)LKGL(128-131)GL(133-134)V(137)L(151)T(157)TALIY(160-164)SEGDK(173-177)KVEK(202-205)DD(213-214)K(218)HSF(220-222)AHPKVDP(224-230)G(233)GYSH(239-242)M(248)R(250)PIMMHD(269-274)M(285)PLYF(288-291)P(293)MV(296-297)EGKLIFS(299-306)E(308)N(334)FIFHNAN(336-342)RL(356-357)NPDLD(359-363)VNGPI(365-369)LDTFKNE(373-379)SA(400-401)DFPRVNES(403-410)TG(412-413)L(424)IAKV(427-430),SE(468-469)VPRVP(473-477)Y(487)I(489)V(504)D(506)VA(515-516)YGFHAFF(526-532)T(534)EQ(536-537)

The first substrate binding pocket comprises:

The α -helices H₂, H₄, H₈

β -turns 1,3,5,6,7,11,12,15,20,23,25,26,33,37,39,41,44,48,57; β -sheets B,D,G,H,A

β -strands 5,10,20,24,25,26 ; β -hairpins 3,6,12,15 and γ -turns 8,18,19.

- (2) SBP2: ASN(37), VAL(41) , PHE(65), VAL(66), ARG (67),MET (88), HIS (90), GLY (91), MET (92), TYR (101),VAL(102),SER(103),SER(468),GLU(469),ILE(471),ILE(489),PHE(490),PHE(491),VAL(492),HIS(493),ASP(494),THR(497),GLY(498),LYS(499),SER(500),ILE(502),VAL(515),ALA(516),VAL(518),LEU(520),PRO(521),HIS (522),ARG(523),VAL(524),PRO(525),TYR(526),GLY (527),PHE (528),HIS(529), ALA(530).

The second substrate binding pocket comprises, β -hairpins 1,2,15, β -strands 2,3,4,24,25, β -sheets A,H, β -turns 3,4,55,56; γ -turn 1,30,31; β -bulge 6.

- (3) SBP3:GLU (17) ILE (20),VAL(21),LYS(22),PRO(72), PHE (74), ALA (75), PRO (76), VAL (77), ALA (78), TYR(80),ASP(84),ASP(86),ARG(110),GLU(114),ALA(120),LYS(121),PHE(122),MET(123),LYS(124),ASP(127),PHE(132),LEU(135),MET(136),MET(139),GLN(140),ARG(143),VAL(149),HIS(155).

The third substrate binding pocket comprises: α -helices 1,3,4, β -turns 6,7,8,10,11,12,13,14,15; γ -turn 3.

- (4) SBP4:HIS (165),HIS(166),LYS(168),LEU(170),TYR(179),VAL(180),LEU(195),ASP(196),TYR(197),LEU (201) GLN (217).

The fourth substrate binding pocket comprises: β -turns 16,17,19,20, β -strands 5, 6; β -sheets B, γ -turn 4, α -helix 5.

- (5) SBP5:ALA(276),ILE(277),THR(278),TYR(281),ALA(282),ILE(283),PHE(314),GLY(315),VAL(316),PHE(330),ALA (343),TRP(344),GLU(345),LEU(352),ILE(353),THR(354),PHE(385).

The fifth substrate binding pocket comprises: β -turn 32, β -strands 13,14,15,16,17,18,19, β -sheets E, F.

(6) SBP6:THR(423),THR(431),GLY(432),ILE(433),VAL(434),LEU(458),TYR(459),ASP(460),LEU(461),GLY(462) GLY(464),ARG(465),PHE(466), GLY(467),PHE(490).

The sixth substrate binding pocket comprises: β -turn 52, β -strands 21, 22, 24 β -sheets G, H, β -hairpins 13, 14, 15; γ -turn 27.

(7) SBP7:TYR(420),GLY(421),ILE(433),VAL(434),LYS(435),LEU(461),ALA(470),LEU(488),PHE(490),LEU(505),MET(510),PRO(514).

The seventh substrate binding pocket comprises: β -strands 21, 22, 24, 25; β -sheets G, H; β -turn 58.

(8) SBP8:TYR(197),ARG(200),LEU(201),VAL(215),LYS(216),GLN(217),VAL(252),GLY(256),MET(258).

The eighth substrate binding pocket comprises: α helix 5, γ -turn 4, β -turns 19, 31; β -strand 11; β -sheet D, β -bulge 4.

(9) SBP9: GLU
(64),VAL(66),GLY(91),MET(92),ARG(93),THR(100),VAL(183),LEU(184),GLY(187),LEU(189),VAL(533).

The ninth substrate binding pocket comprises: β -strands 2, 7, 8, 26; β -sheets A, C; β -hairpins 1, 2, 4, β -bulge 3, β -turn 18.

(10) SBP10: ASN(1),PRO(2),LYS(3),LYS(6),GLY(7),SER(10),VAL(148).

The tenth substrate binding pocket comprises: α helix 1.

2.2.21.5.8. Substrate specificity and enantio-selectivity:

Aforementioned crystallographic studies and additional spectroscopic evidences could provide clues that the protein structures contain specific amino acid residues involved in the formation of salt bridges and hydrogen bonds with the 5-carboxylate and C-1 carboxylate groups of the dioxygenase substrate α -ketoglutarate. These conserved residues are also involved in substrate binding or surrounding, specifically with regard to the structural basis of enantio-specificity. The phenoxy ring of substrates can interact with the

amide nitrogen of Ser152, Ser158 and hydroxyl groups of Tyr153, Tyr420 or the guanidino groups of Arg406, His165 and His225. The tyrosines, presumably, Tyr153, Tyr164, Tyr420 which are supposed to lie around or near oxygen atom, could be playing very important roles in enantio-selectivity. In addition, residues lining the hydrophobic substrate binding pockets including Val 407, Val419, Val228, Val533, Leu172, Leu219, Ile283 along with additional residues near polar carboxylate that includes His225 and His273 and the carboxyl oxygen of Glu469, Glu174, if protonated or bridged by the proton of the water molecule, within the interaction limit may confer specificity to S-enantiomers by forming a hydrogen bond with the oxygen atom of the substrate. Therefore, different enantiomers are assumed to bind with remarkably similar geometries with the carboxyl group and aromatic rings nearly overlapping but at different angles. Such residues could also form π -interactions, with the hydrophobic ring of the substrate and also with the positive side chain of highly flexible residues e.g. Arg406, Phe403, Phe360. Site directed variants (substitutions) can be used to test the subset of proposed residues for their role in controlling enantio-selectivity. Some residues including Glu could hinder binding of incorrect enantiomer from binding and inhibiting the enzyme functions, while, Phe hinders binding of correct (R-) enantiomer and helping the (S-) enantiomer to exclude the active site, so that the expanded active site in the Phe variant more readily binds its substrate. Other amino acid residues besides His339, His439, His529 and Glu409 are also identified that are likely to bind or interact with the (R-) and (S-) enantiomers of the substrate, including several that are likely to contribute to the enantio-selectivity or enantio-specificities of the enzyme.

2.2.21.5.9. Self Hydroxylation:

The dioxygenase enzymes undergo a kind of a self preservation mechanism (Koehntop *et al.*, 2006) in the absence of prime substrate; the solvent retention in the 6th co-ordination position of the metal ion causes such reactions to occur, this further leads to an increase in the activation energy, that facilitates attack by O₂ after displacing the substrate. These reactions are also known as self hydroxylation reactions, which are considerably slower than the conversion of primary substrate in to products, which controls the destructive effects of these nonselective reactions. Structural models of different α -KG dependent dioxygenases combined with ligand or substrates, serve the basis to hypothesize that in the absence of substrate, the conversion of α -ketoglutarate (α -KG) into succinate, utilizes an outer sphere tyrosine residue to act as the metal ligand, which presumably prevent or significantly limit high valent intermediates from the exposure of alternative oxidation reactions. These self hydroxylation reactions could prevent non-selective oxidations which are more destructive, for example, radical induced peptide chain-cleavage, which could result in loss of enzymatic activity permanently. In fact, self hydroxylation of Tyr and Phe residues allow the recovery of catalytic functions of enzyme, since, these residues perform very important structural and functional role. The dioxygenase structural model provides us clues to make an assumption that Phe, Tyr and Trp are some amino acid residues, which are usually involved in this kind of hydroxylations and to consider the possibility that among non-heme iron enzymes such 'uncoupled' self hydroxylation reactions are not an uncommon side chain reactions. (Koehntop *et al.* 2006) The dioxygenase structure possesses multiple essential His, Tyr, Cys, Phe, Lys and Ala residues, that are supposed to be involved in these reactions and based on that; the participation of two mechanistically important pathways in the self hydroxylation of non-heme iron enzymes are being proposed, which could efficiently demonstrate the versatility of mono-nuclear iron centres

in context of activating dioxygen with bidentate coordination of co-substrate α -KG. The charge transfer transition and the enhanced ligand vibrations in the presence and absence of prime substrate by resonance raman spectroscopy or the product ion spectra searched against theoretical digest of the enzyme complex *in silico* can provide the accurate determination of the signatory peak of the corresponding modified compound which can be further traced by nanoscale capillary LC-MS/MS analysis (Koehtop *et al.*, 2006). This MS/MS analysis could indicate the amino acid hydroxylations such as, Trp-OH or Tyr-OH. Moreover, it also shows the surrounding or nearby residues such as His, is acting as a metal ligand. But, this would be beyond the scope of the study, so, the present study has been focused on identifying those specific Tyr, Trp and Phe residues, which are coordinated to the Fe^{2+} or are positioned in close proximity to the active site, which might involve charge transfer transition from metal to ligand or ligand to metal. Since, these reactions have biochemical significance and occur to protect the enzyme from a more destructive, the irreversible oxidation, and the rates of uncoupled reactions relative to coupled reactions; the opportunity to investigate yet unresolved intermediate steps in the mechanism of this fascinating class of dioxygenase structurally has been well achieved.

CHAPTER 3

DIOXYGENASE SUBSTRATE BINDING DOMAINS

DIOXYGENASE SUBSTRATE BINDING DOMAINS

3.1 Introduction:

Every protein sequence has some specific characteristic features to be analyzed in detail, such as motifs, domains etc. The protein structure contains specific substrate binding domains that involve selecting protein partners to interact and based on interaction surface area; determination of intensity of interaction that corresponds to electrostatic contact surface and hot spot regions within the structure becomes possible (Buckingham *et al.*, 2004). Hot spots are regions where most of the interacting residues are available. The interacting amino acid residues and their discriminate binding to the substrate chemical groups can disrupt the contact surfaces between proteins which they interact in normal conditions. This disruption of contact surfaces by substrate serves as a basis for enzyme inhibition (Sergeant *et al.*, 2009). The substrate binding domains of a protein can be a useful tool for developing targeted therapeutics as the analysis of substrate binding pockets which harbor various substrates could provide required information about the mechanism of substrate entry into the binding pocket and the mechanism of their exit. When the mechanism of substrate access to the substrate binding pocket is solved; it is possible to predict the superficial or high intensity interaction between substrate chemical groups and involved amino acid residues and this is possible by homology modeling. Molecular simulation studies can provide information about the nature of substrate binding pocket before and after substrate removal. However, the architecture of substrate binding pocket and involved residues in forming the binding pocket can provide a little information about the hydrophobicity of the substrate binding pocket after charge and energy minimization and substrate removal. The docking prep study depends on substrate interaction surface area, energy and charge minimization or force field.

In homology modeling the substrate chemical groups and interacting residues both in close proximity of the binding substrate and also that are shifted distally in accommodating the substrate in the substrate binding pocket, provide mechanistic information about the reactions that could occur upon substrate binding. The feasibility of biochemical reactions and underlying signaling pathways largely depend on modifications on the residue side chains or aromatic ring. These modifications can affect chemical homeostasis of the local microenvironment and also bring about conformational changes that implies in varying interaction intensity of residue and substrate groups. These varying interaction intensities serve the basis of interactions such as weak vanderwaal forces, hydrophobic to strong polar interactions. The hydrogen bond formations between C-H, N-C, C-O and catch-slip bond interactions between carbonyl carbon of Glutamine and Lysine side chain are certain examples of the varying interaction intensity between residues and substrate groups. The interaction pattern has predictive value that is characteristics of any enzyme-substrate interaction study.

3.2. Dioxygenase sequence and substrate binding domains:

VSNDGIVVPNPKPSKGLTSLVLDLVEKLVKLMYDSSQPHHYLAGNFTPVIEETPPT
 KDLNVIGHLPDCLNGEFVVRVGNPKFAPVAGYHWFDGDGMIHGMRIKDGKATYV
 SRYVKTSRLKQEEYFGGAKFMKIGDLKGLFGLLMVNMQQLRAKLVVDLSYGH
 GTSNTALIYHHGKLLALSEGDKPYVLKVLEDGDLQTVGLLDYDKRLKVEKPHVT
 DDHDDVKQKLHSFTAHPKVDPFTGEMFTFGYSHDPPYVMYRVVSKDGFMHDPVP
 ITVPAPIMMHDFAITENYAIFMDLPLYFKPKEMVKEGKLIFSDETCKARFGVLP
 RYAKDDLIRWFELPNCFIHNAWEEDEEVVLITCRLENPDLDMVNGPIKKKLDTF
 KNELYEMRFNLKTGLATQKKLSESAVDFPRVNESYTGRKQRFVYGTTLDSIAKVT
 GIVKFDLHAAPEVGKTKIEVGGNIQGLYDLGPGRFGSEAFVPRVPGITSEEDDGYL
 IFFVHDENTGKSIAHVLDAKTMSNDPVAVVELPHRVPYGFHAFVTEEQLQEQL
 L

The dioxygenase sequence analysis indicates some substrate binding domains in the protein sequence that are novel in the sense that these substrate binding domains provide

reasons to speculate about the dynamic functional importance of dioxygenase family of enzymes varying from plant to animals and humans.

The group of α -ketoglutarate dependent dioxygenases comprises of different plant dioxygenases including 4-Hydroxyphenylpyruvate dioxygenase, Catechol dioxygenase, Lipoxygenases, Prolyl hydroxylases, Tryptophan hydroxylases, Tyrosine hydroxylases, human Tryptophan-2,3-dioxygenases and Indoleamine-2,3-dioxygenases that undergo very similar chemical reactions mechanistically and have substrates to act on which are common to various signaling pathways. These enzymes are involved in carotenoid biosynthesis, fatty acid metabolism, amino acid metabolism, aromatic compound degradation, rhodopsin biosynthesis, flavanol and quinone biosynthesis, succinate biosynthesis and other plant metabolites, whereas human counterparts perform innate or adaptive immune response functions, that is coordinated by several molecules that are common to the underlying signaling cascade, facilitating cell-to-cell communications. These binding domains also provide reasons to assume that dioxygenases play a very central role in the signaling pathways that connect catabolic processes to the plant defense response triggered by many allergens, also to the activated peripheral immune response during diseased conditions in humans. Also, these binding domains are imperative that it can be hypothesized that plant hormone (Kende and Zeevaart *et al.*, 1997) interactions with dioxygenases in plants could pave the way to find potential new inhibitors or herbicides and also explain herbicide degradation, which are observed in the presence of some herbicides under oxidative conditions. These binding domains can provide new insights to the structure aided drug designing and could serve the basis of predicting various drug substrates which can bind to the dioxygenases and would undergo dioxygenase mediated signal transduction processes which are implicated in the various clinical trial results of many drugs pertaining to the treatments of numerous diseases and related conditions. There are myriads of signaling

cascade reactions and these drugs follow some of them while some signaling pathways are uninterrupted or less interfered by certain drugs that leads to side effects or poor clinical trial results, therefore, we imply these binding domains for developing understanding for potential suitable drug combinations that might prove beneficial for combinatorial drug therapy. In this study, some novel substrate binding domains in the dioxygenase structure have been identified; that enhances further implications of dioxygenase structure in the research orientated studies as well as finding other beneficial roles that dioxygenases play in numerous cellular processes.

Other highlighted sequences depict transmembrane sequence.

The sequence analysis indicates following substrate binding domains:

Substrate binding domain 1

- (1) Allergen and Lipid binding
- (2) Epoxycholesterol
- (3) Cell adhesion

Substrate binding domain 2

- (1) Ser/Thr kinase binding:
- (2) Actin like ATPase

Substrate binding domain 3

- (1) Dicer like protein
- (2) Peptide binding
- (3) Myosin ATPase

(4) Immunoglobulin protein

3.3. Dioxygen:

On earth, oxygen is the common allotrope of elemental oxygen O_2 but also called as dioxygen or molecular oxygen. Dioxygen can occur in the singlet state oxygen, which is a metastable form; whereas the ground state is known as triplet oxygen. The bond length of O_2 is 121 pm and has a bond energy of 498 kJ/mol. Oxygen is a colorless gas which has a boiling point of $-183\text{ }^\circ\text{C}$ (90 K or $-297\text{ }^\circ\text{F}$). By cooling with liquid nitrogen it can be condensed out of air, which has a boiling point of $-196\text{ }^\circ\text{C}$ (77 K or $-321\text{ }^\circ\text{F}$). Liquid oxygen is paramagnetic i.e. liquid oxygen contained in a flask suspended by a string is attracted towards a magnet. It has a pale blue color. The molecular oxygen has quite unusual electronic configuration that it has the only stable diatomic homonuclear molecule with a paramagnetic ground state. Between two oxygen atoms, there is simple electron-pair bonding while enough for predicting a double bond but require transitions to rationalize the triplet ground state.

The dioxygen, has a molecular orbital description of $1s^2 2s^2 2p^4$ which is the configurations of two conjugated oxygen atoms and it leads to a $(1\sigma_g)^2(1\sigma_u^*)^2(2\sigma_g)^2(2\sigma_u^*)^2(2p\sigma_g)^2(2p\pi_u)^4(2p\pi_g^*)^2$ molecular configuration for the $^3\Sigma_g^-$ ground state. The 5 degenerate π^* orbitals contain two remaining unpaired electrons. Above the ground state, the first excited state of O_2 is $^1\Delta_g$ singlet with an energy of 22.5 kcal/mol. The superoxide ions are generated by one electron reduction of O_2 and the peroxide ion O_2^{2-} is generated by two electrons reduction. Since, in the superoxide ion electrons are added to the π_g^* orbital thus its bond length increases to about 1.34 \AA and in the peroxide ion to approximately 1.49 \AA . This is comparative to 1.21 \AA which is the

bond length for molecular oxygen showing the expected bond order trend: $2(\text{O}_2)$, $1.5(\text{O}_2^-)$ and $1 (\text{O}_2^{2-})$.

Dioxygen complexes: The dioxygen binding has stoichiometry of 1:1 or 1:2; dioxygen complexes have possibly three structural types, which have been characterized. One electron transfer to dioxygen forms a complex which is formally recognized as the first type of complexes, while the complexes 2 and 3 include the oxidation states reflecting two-electron transfer. Besides this, two other structural types involving heme have been characterized, i.e. in hemocyanin, the $\mu^2\text{-}\eta^2$ binding of dioxygen to binuclear copper and the oxo-diiron centre of hemerythrin hydroperoxide complex (Stenkamp *et al.*, 1994).

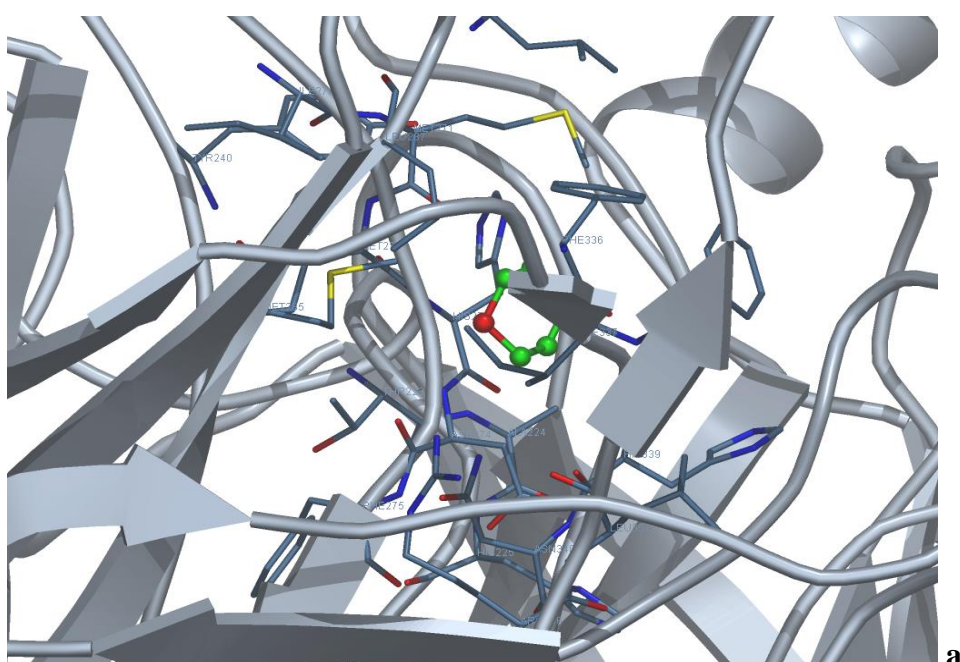
Natural oxygen carriers: The hemoproteins (myoglobin and hemoglobin) carrying oxygen contains an active site constituted by an iron (II) protoporphyrin IX (the heme), between the imidazole group of the proximal histidine residues and the iron atom bound to the protein by a single coordinate bond. The heme is hold by various residues from the protein through hydrophobic interactions.

The coordination site of the iron can also bind to several smaller ligands such as CO, NO alkyl isocyanides, the distal aromatic residues around it are involved in controlling the immediate environment. The polar, hydrophobic or steric interactions could be induced by these residues which participate in regulating the affinities of the bound ligands. Dioxygenases are other enzymes, which incorporate molecular oxygen by cleaving carbon-carbon double bonds and playing very important roles in the biosynthesis of plant hormone Abscisic acid, carotenoid cleavage, aromatic compound degradation, retinoid biosynthesis, carbohydrate and amino acid metabolisms and immune regulation. Therefore, a major preoccupation of this study is to understanding the chemical and structural basis of these remote impacts on binding affinities of ligands.

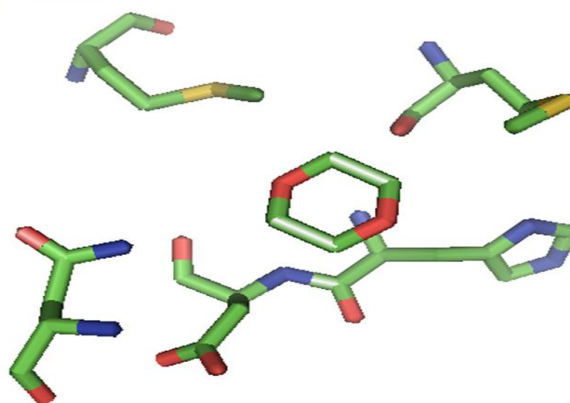
3.3.1. Dioxygenase and dioxygen binding:

Table 3.1 Energy table

Rank	Est. Free Energy of Binding	Est. Inhibition Constant, Ki	vdW+Hbond+desolv Energy	Electrostatic Energy	Total Intermolec. Energy	Frequency	Interaction Surface
1.	-3.58 kcal/mol	2.39 mM	-3.52 kcal/mol	-0.05 kcal/mol	-3.58 kcal/mol	100%	249.961



for evaluation only.
act_sales@delsci.com.



b

Fig. 3.1 dioxygenase-dioxygen

Interactions

271: MET

273: HIS (active site residue)

274: ASP

285: MET

340: ASN

3.3.2. Dioxygenase and Dioxygen interactions:

Table 3.2 Interaction table: Dioxygenase-Dioxygen

hydrophobic		other	
C	MET271 [3.35] (CE)	O	HIS273 [3.60] (CB)
C	HIS273 [3.51] (CB)	C	ASP274 [3.20] (CB)
C	MET285 [3.43] (CE)	O	MET285 [3.41] (CE)
		C	ASN340 [3.49] (ND2)

The dioxygen, which is the main substrate for dioxygenases, indicated the binding residues Met, Asn, Asp and His, that are involved in oxygen mediated oxidation of various substrates, which is a typical function of this group of enzymes. Enzymes that have functions in the biosynthesis of small molecule including plant hormones and bacterial antibiotics belong to this family, which undergo certain modifications e.g. in the hypoxia inducible factor (HIF), hydroxylation of amino acid side chains i.e. proline and asparagines (Flaus *et al.*, 2006; Kenneth *et al.*, 2009); oxidative removal of methyl group from both methylated histone proteins and alkylated nucleic acids and in the DNA, hydroxylation of methyl-cytosine. Structural scenario suggest the dynamic binding of molecular oxygen in terms of reaching to the active site; since, active site residue His 273, that is bound to the

dioxygen is supported by nearby Glu, Asp and Asn residues, which may provide carbonyl and nitro groups required for facilitating the binding to take place. The active site in all forms of dioxygenases is located inside a tunnel running through the structure (Messing *et al.*, 2010), which could differ in shape and diameter, thus serves the basis for this dynamic binding as dioxygen has to reach to the active site and activated by the iron centre upon binding to effect various oxidative transformations. Functionally, His residues have remarkable property of scavenging oxidative molecules and oxidative reactions result in the accumulation of redox molecules, such as superoxide ions, hydrogen peroxide; which could modify other Cys residues. The structure contains His, Phe and Tyr residues, which are possible targets for hydroxylations and phosphorylations under such oxidative conditions to recover the enzymatic functions from permanent inactivation mediated by destructive or non-selective oxidation reactions. This coupled or uncoupled self recovery reactions are not an uncommon event in this class of α -ketoglutarate (α -KG) dependent dioxygenases (Koehtop *et al.*, 2006).

3.4. Substrates binding domains:

3.4.1. Substrate binding domain 1

3.4.1.1 Allergen and Lipids:

Many major allergens are lipid-binding proteins which indicate that the lipid molecules may be responsible for their allergenicity. For example, the house dust mite allergens Der p 2 and Der f 2 because of the structural similarity with LPS-binding protein, bind to bacterial lipopolysaccharide (LPS). Another house dust mite allergen, Der p 7 shows less affinity binding with the bacterial lipopeptide polymyxin B but does not bind LPS. Also, the structure of primary amino acid from group 14 allergens in house dust mite clearly indicates that these are lipid-binding proteins. Two major allergens in *Parietaria judaica*,

Par j 1 and Par j 2 are other respiratory allergens which can also bind to lipid; more specifically these allergens preferentially bind to monoacylated negative lipids. In southern Europe, rosaceae fruits in birch-independent areas have allergens which belong to the group of nonspecific lipid transfer proteins (nsLTP) were named because of the broad lipid-binding specificity they possess which is closely related to their 3D structure. Similarly, the 3D structure of the strawberry dioxygenase indicated a solvent-accessible cavity that runs across the core of the molecule. In this cavity, dioxygenase can potentially bind to ligands e.g. fatty acids, flavonoids and cytokinins. By employing computational approach, the qualitative and quantitative data on the ligand-binding properties of a major allergen i.e. Nematode allergen and lipid have been obtained and analyzed. Moreover, the observed lipid binding could be a very common characteristic of many important groups of allergens. The allergen bound bacterial lipopeptides may trigger the Th2 immunity by activating the Toll like receptor (TLR) and TNF α pathways. In fact, various bacterial and mycobacterial lipopeptides by interacting with TLR2 can induce the release of cytokines which function in coordination with TLR1 or TLR6. TLR4 recognizes the LPS from Gram-negative bacteria and transmit a strong danger signal for the immune responses in the human body. Commonly, the induction of Th1-like responses is up regulated by the accumulation of LPS, whereas low amounts support the induction of Th2-responses. Recently, another study reported that TLR4 signaling could also be involved in the induction of Th2 responses to inhaled antigens. Interestingly, in the study, it has been demonstrated that in a humid milieu such as mucosal surfaces of the human body pollen grains in addition to allergens release pollen-associated lipid mediators. Pollen-derived E (1) phytosteranes are modulating the function of human dendritic cells in a manner favoring the polarization of Th2 cells. Although, the mechanism of these responses were poorly understood but it can be speculated that pollen allergens with lipid-binding capacity

may also bind pollen-associated lipid mediators that can contribute to their allergenicity. Based, on these binding domains, the study includes lipid binding and cholesterol binding to the carotenoid binding sites of dioxygenases, while similar results have been reported recently, in the photosystem II enzymes (Kalman *et al.*, 2012).

3.4.1.2 Allergen-lipid binding:Nematode Allergen-Lipid Binding

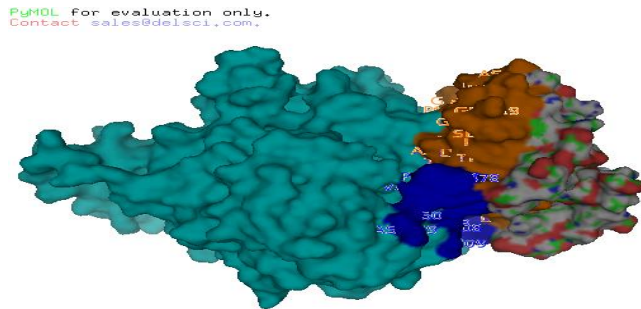


Fig. 3.2 Dioxygenase-Nematode allergen and lipid

Dioxygenase (cyan) and allergen-lipid (green) contact surfaces

Dioxygenase-Allergen –Lipid (orange-blue):

Dioxygenase: K(11)K(18) IE(42-43)P(46)L(51)FA(74-75)G(118)NT(496-497)P(521).

Allergen-lipid: A(39)H(43)E(47)EE(91-92)E(99)I(113)K(121)K(122)S(129).



Fig. 3.3 Dioxygenase (spectrum)-Allergen-lipid (blue) binding residues.

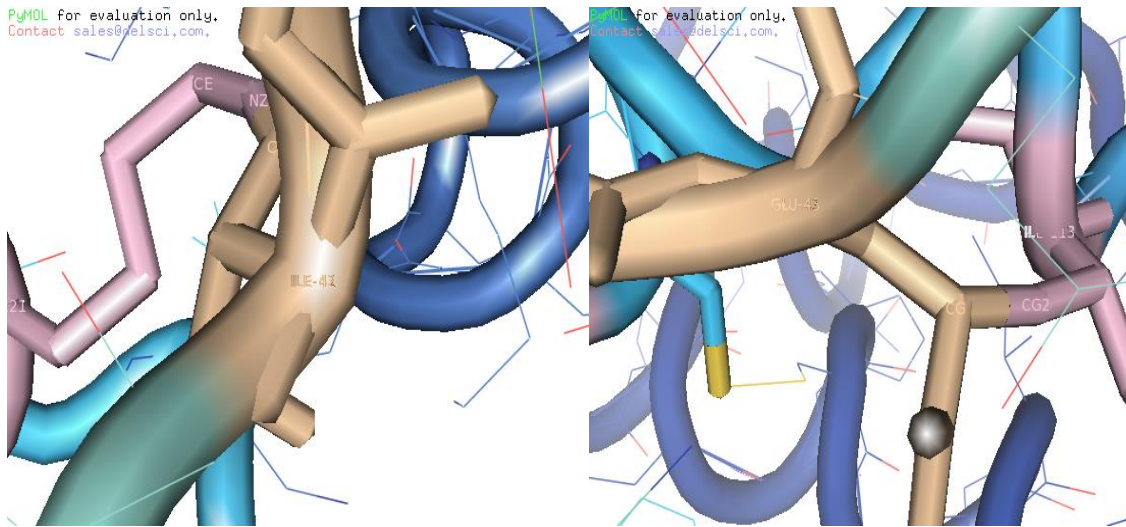


Fig. 3.4 Dioxxygenase (cyan) and allergen-lipid (blue) binding residues, Dioxxygenase (wheat), Allergen-lipid (light pink)

Interactions:

Dioxxygenase	Allergen-lipid
Ile42	Lys121
(CG2)	(NZ)
Glu43	Ile113
(CG)	(CG2)
Lys11	Glu92
Asn1	Glu99
Gly118	Glu91
Ser129	Lys122
Ala39	Leu51
His43	Pro521
Glu 47	Asn496

3.4.2. Alpha-epoxycholesterol:

There are extensive evidences indicating that oxidized lipoproteins could be involved in the complex and multifactorial etiology of atherosclerosis. The mechanism by which lipoproteins are oxidized and the location site and are not well studied and also the location of lipid oxidation is not known i.e. whether the lipoproteins are oxidized and formed locally in the artery wall and/or the uptake of circulating oxidized lipoproteins causes the sequestration of oxidized lipids in atherosclerotic lesions (Chapman *et al.*, 2011). The present study has been focusing on demonstrating structurally, oxidized lipids in the diet could be responsible for generating such circulating potentially atherogenic oxidized lipoproteins.

Thus, the aim of the present study has been to describe whether oxidized dietary fats such as oxidized cholesterol and fatty acids are absorbed and contribute to the oxidized lipids repertoire in circulating lipoproteins (Andersen *et al.*, 1979). In animals such feeding experiments show that when oxidized cholesterol or fatty acids were fed to rabbits, it increased the fatty streak lesions in the aorta (Andersen *et al.*, 1979). Moreover, the aortic lesions in apo-E and LDL receptor-deficient mice are increased by the dietary oxidized cholesterol. A typical western diet rich in oxidized fats could contribute to the increased arterial atherosclerosis in their population. Other studies by Andersen *et al.*, 1979 reported that, when normal subjects were fed with a meal rich in oxidized linoleic acid, only the postprandial chylomicron/chylomicron remnants (CM/RM) contained the oxidized fatty acids and were removed from circulation within 8 h. High density lipoprotein (HDL) or Low density lipoprotein (LDL) fractions did not contain oxidized fatty acids in at any time. However, when human subjects were fed with alpha-epoxy cholesterol, the CM/RM contained the alpha-epoxy cholesterol in serum and also the endogenous very low density lipoprotein, LDL and HDL contained them which sustained in the circulation for 72 h.

When CM/RM fraction rich in alpha-epoxy cholesterol with human HDL and LDL and without alpha-epoxy cholesterol was incubated in *vitro*, this resulted in a rapid transport of oxidized cholesterol to both HDL and LDL from CM/RM. Thus, the present study suggests that cholesteryl ester transfer protein could be mediating this transfer. Therefore, it is assumed based on the structure model that alpha-epoxy cholesterol in the diet is incorporated into CM/RM fraction and then transferred to LDL and HDL facilitating lipoprotein oxidation. It is being hypothesized that oxidized fatty acids derived from diet accumulated in chylomicron remnants and oxidized cholesterol was deposited in remnants and LDL, both by increasing oxidized lipid contents in chylomicron remnants and circulating LDL enhance chances of atherosclerosis. The structure model for α -epoxycholesterol-dioxygenase complex supported this hypothesis.

3.4.2.1. Alpha-epoxycholesterol binding:

3.4.2.2 Energy table:

Table3.3: Energy Table (Dioxygenase- α -epoxycholesterol)

Rank	Free energy of binding	Inhibition constant, K_i	vdW+Hbond+ desolv Energy	Electrostatic Energy	Total Intermol. Energy	Frequency	Interact. surface
1	+6.42kcal/mol		+3.88 kcal/mol	-0.05 kcal/mol	+3.83 kcal/mol	50%	894.202
2	+8.25kcal/mol		+5.11 kcal/mol	-0.01 kcal/mol	+5.10 kcal/mol	50%	847.138

PyMOL for evaluation only.
Contact sales@delsci.com.

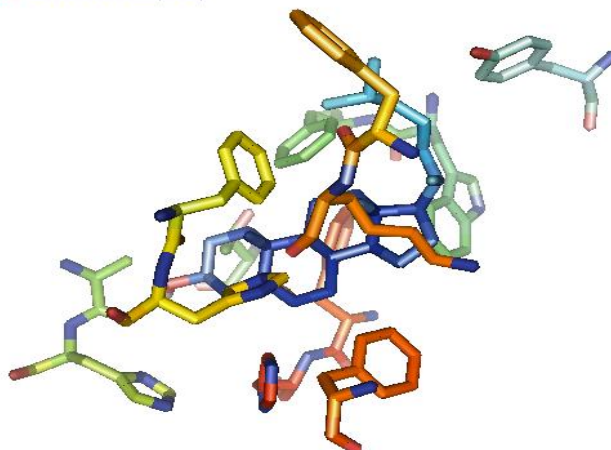


Fig.3.5 Dioxxygenase- α -epoxycholesterol

Interactions:

Tyr33, Trp82, Phe83,

Thr160, Ala224, His225,

Phe338, His339, Phe376,

Lys377, Phe404, Phe528,

His529.

3.4.2.3. Dioxxygenase and Alpha-epoxycholesterol interactions:

Table 3.4: Interaction table (Dioxxygenase- α -epoxycholesterol)

Hydrogen bonds	Polar	Hydrophobic	Pi-Pi	Cation-Pi	other
none	H1- His225 [3.89] (ND1)	C11- Trp82 [3.67] (CE3,CZ3)	C1- His339 [3.22] (CD2)	none	C23- Tyr33 [3.71] (OH)
	O1- His339 [3.90] (NE2)	C13- Trp82 [3.26] (CE3,CZ3)	C5- His339 [3.53] (CD2,CE1)		C24- Tyr33 [3.90] (OH)
		C6- Trp82 [3.44] (CZ3)	C5- Phe404 [3.82] (CD1)		
		C7- Trp82	C1- Phe528		C12-

		[3.79] (CZ3) C14- Trp82 [3.86] (CZ3) C16- Trp82 [3.58] (CZ3) C9- Phe83 [2.57] (CE1,CE2,CZ) C11- Phe83 [3.01] (CE1,CZ) C12- Phe83 [3.55] (CE2) C3- Phe83 [3.86] (CZ)	[3.77] (CD1) C5- Phe528 [3.87] (CD1) C5- Phe529 [2.65] (CD2,CG) C1- Phe529 [3.43] (CD2)		Thr160 [3.00] (CG2) C17- Thr160 [2.84] (CG2) C18- Thr160 [2.66] (CG2) O2- Thr160 [3.59] (CG2) H1- Thr160 [3.46] (CG2) O2- Ala224 [2.64] (CB) H1- Ala224 [2.34] (CB) C1- His339 [2.73] (NE2)
--	--	---	---	--	--

The active site coordination in the structure contains iron bound octahedrally to four His residues, consistent with other dioxygenase structures (Messing *et al.*, 2010), the shape of the active site coordination may vary depending on the environmental conditions, such as pH and hydroxyl or nitrate compounds. The shape of coordination varies from octahedral, distorted octahedral to bipyramidal in these class of dioxygenase under different local environment. This type of coordination has been reported in fumeratase and photosystem II enzymes (Messing *et al.*, 2010; Zhang *et al.*, 2007). Recently, one study reported a lipid binding site in photosynthetic reaction centre (Kalman *et al.*, 2011), here, the present study identifies a novel Allergen-lipid binding domain in the strawberry dioxygenase that confirms the hypothesis that dioxygenases play very important role in lowering cholesterol, cardiovascular diseases and in the peripheral immune responses. The dioxygenase and nematode allergen-lipid complex indicated binding of dioxygenase Ile42 and Glu43 residues with Lys121 and Ile113 residues of allergen-lipid respectively and other possible interaction between dioxygenase residue Ser129 and Lys122 of allergen-lipid. This interaction between Glu and other branched chain residues e.g. Ile indicates the possible role of dioxygenase in innate immunity, as Ile has been found to activate defensins in the intestinal epithelium (Yoneda *et al.*, 2009). Moreover, Glu residues are also structurally and functionally essential for the activation of ubiquitin-E3 ligase and ubiquitin mediated proteasomal degradation of proteins pertaining to angiotensin system (Hashimoto *et al.*, 2012). The angiotensin converting enzyme, ACE2 is remarkably responsible for interacting gut microbiota and involved in severe intestinal bleeding and other intestinal epithelium dysfunctions (Hashimoto *et al.*, 2012). This also shows that the strawberry contains natural compounds that can lower cholesterol and allergens, which may have been observed as anaphylactoid reactions on consumption of strawberries. Transcription factors such as Nrf2 have been reported in the strawberry fruit extract, which plays important role

in diabetes and cardiovascular diseases (Thornalley *et al.*, 2012); thus, it is anticipated that dioxygenase and allergen-lipid complex may solve some issues in the potential role of dioxygenases in the gut microbiota and cardiovascular system. Further, as dioxygenases also participate in fatty acid metabolism, dioxygenase binding with α -epoxycholesterol has been also studied which could be an important substrate for dioxygenases along with 9-cis-epoxycarotenoid. There are extensive evidences that a key role is played by oxidized lipoproteins in the etiology of atherosclerosis. The dioxygenase and α -epoxycholesterol complex and the strawberry dioxygenase-lipid complex can provide required insights that diets with oxidized lipids donate such oxidized lipoproteins in circulation are potentially atherogenic and on consumption the oxidized dietary fats for example oxidized cholesterol and fatty acids are absorbed by the body which accumulate and add up to the oxidized lipids pool content in circulating lipoproteins. Moreover, the dioxygenase- α -epoxycholesterol structure complex indicated active site polar residues His225 and His339 intracting with H1 and O1 atoms of the substrate respectively, further, hydrophobically interacting Trp82 and Phe83 residues interact with carbon atom of the α -epoxycholesterol and other residues such as Tyr33, Thr160 and Ala224 interact with hydrogen, oxygen and carbon atoms of the substrate along with His339 and Phe529, Phe404 and Phe528 which form π -stacking interactions with the C1 and C5 atoms of the α -epoxycholesterol, this binding interaction could provide not only structural insights into determining the site of lipoprotein oxidation but also the functional mechanism, that how the oxidized lipoproteins are formed locally and sequestered in the site of atherosclerotic lesions or in the amyeloid plaques following the uptake of circulating oxidized lipoproteins (Chapman *et al.*, 2011; Andersen *et al.*, 1979). Based on this structure complex and the substrate binding domain, cholesterol lowering drugs such as ‘statins’ and other anti-platelet, anti-arrhythmic therapeutic drugs as drug substrates could

be studied which could resolve the multifactorial complex etiology of different arterial, coronary and neuronal plaque formations.

3.4.3. Cell adhesion proteins:

The process of binding of a cell to any surface i.e. extracellular matrix or another cell is called cell adhesion by using molecules for cell adhesion e.g. integrins, selectins and cadherins. In maintaining multicellular structure correct cellular adhesion is essential. Cellular adhesion links the cytoplasm of cells to activate receptor mediated signal transduction process.

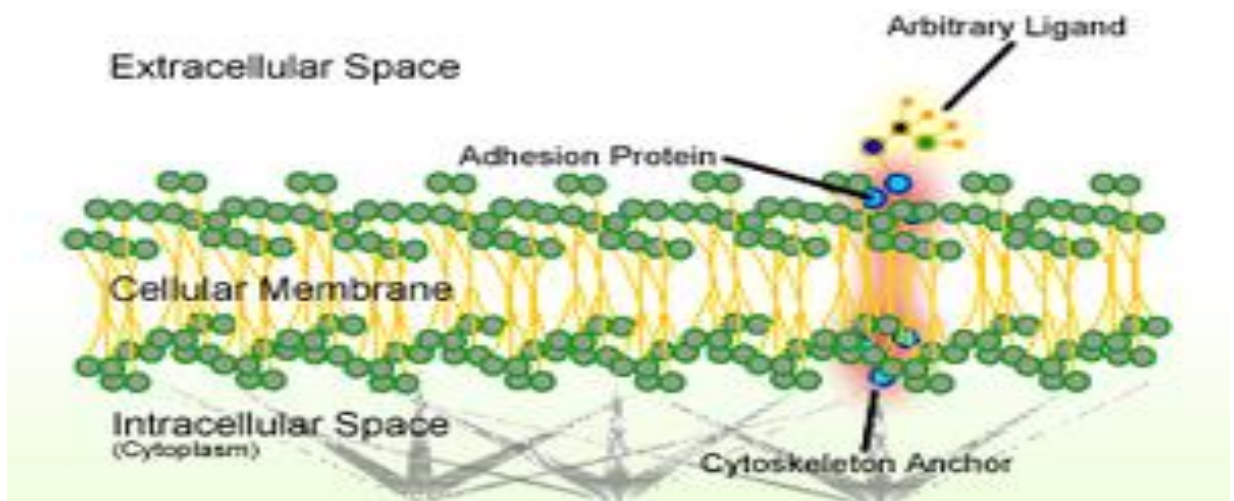


Fig. 3.6 Adhesion proteins, cytoskeleton anchor

In the process extracellular enzymes first hydrolyze the cell adhesion molecules. The cell adhesion process is directly related to protein adsorption. Multiple adhesion molecules are expressed in eukaryotic protozoan, such as malaria parasite uses one adhesion molecule called the circumsporozoite protein for binding to liver cells. The pathogenic protozoan (*Plasmodium falciparum*), uses another adhesion molecule called the merozoite surface protein to bind red blood cells. There are many different types of adhesion molecules present in human cells; the major classes include integrins, cadherins, selectins and Ig superfamily members. Different ligands are recognized by these adhesion molecules and individually have different functions. The defects in expression of adhesion molecules can

relate to the defects in cell adhesion. The selectivity of cell–cell adhesion is involved in the assembling of vertebrate cells into ordered tissues; the selective recognition is allowed by hemophilic attachment that similar type of cells stick together whereas different type of cells stay segregated.

3.4.3.1. Human Genetic Diseases

The dysfunction in expressing a specific adhesion molecule causes human genetic diseases. The leukocyte adhesion deficiency-I (LAD-I) is an example, in which β 2-integrin subunit precursor is not expressed in the patients. During inflammation this β 2-integrin is required for leukocytes adhesion to the blood vessel wall to activate the signaling during infection. In LAD-I patients the leukocytes fail to adhere to the surface and patients show symptoms of infections which could be severe.

3.4.3.2. Tumor adhesion

The circulatory tumors use the circulatory system to spread through various tissues and use cell adhesion mechanisms. The release of epoxyicosatrienoic acid increases this propensity.

3.4.3.3. Prokaryotes

Adhesion molecules such as ‘adhesins’ are mainly found in prokaryotes. Adhesins occur on flagellae, pili (fimbriae) and on the cell surface. Bacterial colonization includes adhesion as the first step and regulates tropism (cell or tissue-specific interactions).

3.4.3.4. Viruses

The viral binding to host cells includes adhesion molecules present in them. For example, influenza virus contains a surface hemagglutinin which is recognizes sugar sialic acid on cell surface molecules of the host. Moreover, in lymphocytes an adhesion molecule i.e. gp120 for HIV is expressed and it binds to its ligand CD4.

3.4.3.5. Focal adhesion kinase (FAK):

Protein tyrosine kinase 2 (PTK2) is also known as Focal Adhesion Kinase (FAK) which is encoded by the *PTK2* gene in humans. The focal adhesion-associated protein kinase PTK2 is believed to be involved in cellular adhesion process such as cellular sticking and cell proliferation. Other studies have shown that blocking FAK, because of decreased mobility the breast cancer cells became less metastatic. The engagement of integrin receptor, growth factor stimulation and the mitogenic neuropeptides action could cause phosphorylation of FAK.

Integrin receptors belong to the family of heterodimeric transmembrane glycoprotein receptors that cluster upon engagement of ECM leading to phosphorylation of FAK and recruitment of focal adhesions (Puklin-Faucher and Vogel *et al.*, 2009; De Mali *et al.*, 2003; Lietha *et al.*, 2008; Pandey *et al.*, 2012). In human endothelial cells, a cleavage of FAK by caspase 3 at Asp-772 activates early apoptotic signaling, which generates two fragments of FAK of approx. 90 and 130 kDa in mass. The smaller fragment of FAK is termed "killer FAT domain" and it is the domain associated with death signaling. Throughout the process of apoptosis, FAK is an important protein responsible for rounding of the cell, loss of focal contacts and membrane changes during apoptosis such as blebbing that includes cortical actin ring contraction followed by condensation of chromatin and nuclear fragmentation. The FAK overexpression can lead to apoptosis inhibition and an increase in the metastatic tumors sustainability. The cytosolic protein tyrosine kinase is found abundantly in the focal adhesion cells attached to extracellular matrix constituents. The encoded protein belongs to the FAK subfamily of protein tyrosine kinases which include PYK2 but having less sequence similarity to other kinases from different subfamilies. Most cells exempting certain types of blood cells are expressing FAK. FAK tyrosine kinase on activation plays an important role in the early step of cell

migration. FAK activity turn on intracellular signal transduction promoting loss of cell contacts with the extracellular matrix, thus facilitating cell migration. The development process also requires FAK and the loss of FAK activity results in lethality. It seems that for cell migration FAK is not absolutely required and is involved in other cellular processes, including the regulation of p53, the tumor suppressor. This gene is encoding four isoforms as at least four transcript variant have been found, but only two of them have been determined and characterized (Schaller *et al.*, 2010; Eleniste *et al.*, 2012; Lietha *et al.*, 2008; Pandey *et al.*, 2012).

In the cells the focal adhesion dynamics include the recruitment of FAK which is a protein of 125 kDa that participate in cell adhesion among different type of cells and has important role in cell motility and survival. FAK was initially identified as an oncogene protein tyrosine kinase v-src target and it is a highly conserved, non-receptor tyrosine kinase. This cytosolic kinase is playing significant role in various cellular functions including locomotion of cells, response to mitogens and cell survival. Typically, FAK is recruited at focal adhesions structures; these are multi-protein structures connecting the extracellular matrix (ECM) to the cytoplasmic cytoskeleton (Eleniste *et al.*, 2012; Dasari *et al.*, 2010). Additional components which interact with FAK in terms of activating and forming focal adhesion network include talin, actin, filamin, vinculin, paxillin and tensin.

3.4.3.6. Focal adhesion kinase (FAK) binding:

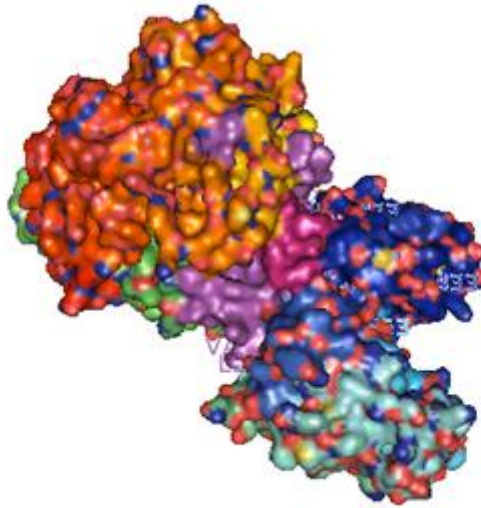


Fig. 3.7. Dioxxygenase-Focal Adhesion Kinase (FAK)

Contact surface residues 1 (shown in magenta)

[SKGLTSKLVDLVEKLIVKLMYDSSQ (5-29)

IDGLKGLFGLLMVNMQQLRAKLVVDSYGH (125-155)

VESYSHDPPYVPAIPNCFIF] (203,204,221,240-247,267,268,270,333-338)

Focal adhesion kinase (FAK) (shown in blue, green cyan and cyan)

Dioxxygenase (shown in rainbow)

Contact surface residues 2 (shown in warm pink)

[FKPKEMVKEGKDKPY GLL] (291-301,176-179,193-195)

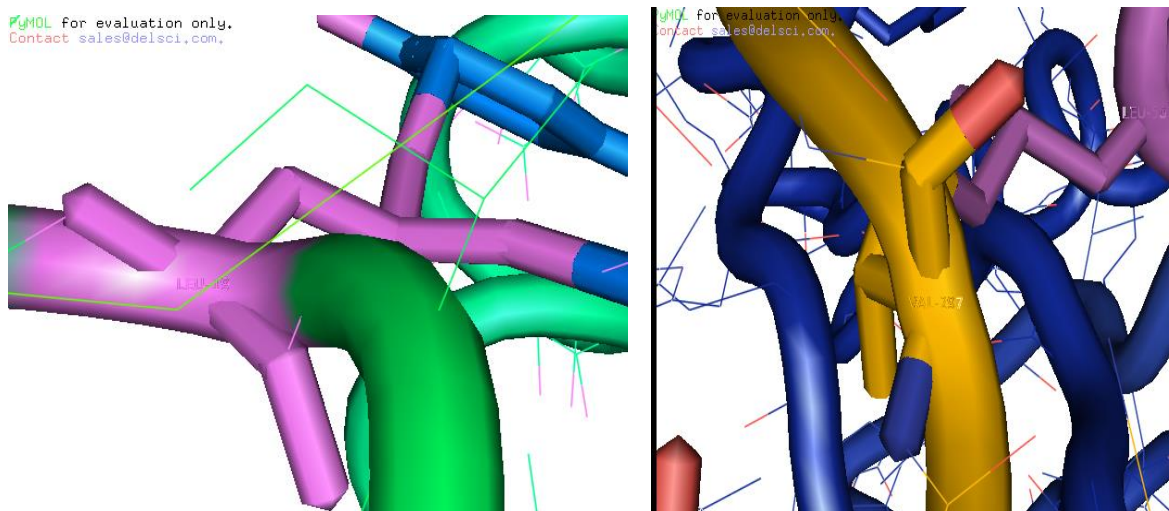


Fig. 3.8 Dioxxygenase-FAK binding residues; **a.** Leu19 and Tyr528, Leu525 **b.** Val297 and Leu501

3.4.3.6.1 Dioxxygenase and Focal Adhesion Kinase (FAK) interactions:

Interactions:

Diox	FAK
Leu19-	Tyr528, Leu525 (bound residues)
(CD1)	(CE2)
(CD2)	(CD1)
Pro293	Ile440, Glu423
Glu204	Ser443, Pro444
Val21	Asp558
Leu135	Asp558
Val297	Leu501 (bound residues)
(CG1)	(CD2)
His242	Ile440, Glu423, Met442, Ser443

Based on the substrate binding domains, present in the dioxxygenase structure, it is hypothesized that the dioxxygenase may be participating in a common pathway that includes

extracellular matrix and extracellular cues sensing receptor Integrin (ITG) in a way that the signal cascade involves dioxygenase in the complex process of tumorigenesis, cell proliferation, angiogenesis and DNA damage (De Mali *et al.*, 2003). Based on the present study an idea is being developed for making connections between the signal transductions of Integrin and G-coupled receptors in a way that involves dioxygenase in a common pathway, that underlies in the complex process of tumorigenesis, cell proliferation, angiogenesis and peripheral immune regulation mediated by TGF- β , TNF- α etc. Focal adhesion kinase (FAK), binds to the Integrin receptor in terms of transmitting signals from Integrin to PI3K, AKT, mTORC1, Erb1, MEK, ERK, Elk1, Bcl2, PAK, VEGF, COX2, PDK1, c-Jun, JNK, Wnt, β -catenine which involves other proteins and factors such as p53, p62, p21, cdc37, cdc42, cdc2, cdk2 regulating cell proliferation, cell cycle, tumorigenesis, angiogenesis and DNA damage. This study is first of its kind that an entirely new pathway has been studied on structural basis; as dioxygenase and focal adhesion kinase structure complex indicated bound amino acid residues Leu19 (CD1)(CD2) and Tyr 528(CE2), Leu525(CD1); and Val297 (CG1) and Leu501 (CD1) and other interacting residues dioxygenase Pro293, Glu204, Val21, Leu135 and His242 and FAK Ile440, Glu423, Ser443, Pro444, Asp558, Leu501 and Met442, which shows, how FAK or PTK2 focal adhesion associated-protein kinase are involved in cell adhesion including cell sticking and cell motility. The integrin engagement is associated with the phosphorylation of FAK, stimulation of growth factor and mitogenic neuropeptide the activity. Integrin receptor clustering upon extracellular cell matrix engagement leads to phosphorylation of FAK and recruitment of focal adhesion eliciting the down signaling. During early signals of apoptosis in the human endothelial cells caspase 3 cleaves the FAK, at Asp residue (Schaller *et al.*, 2010; Dasari *et al.*, 2010; Pandey *et al.*, 2012), however, dioxygenase interaction with FAK suggests that Glu residues in the dioxygenase may regulate the cell

differentiation, cell proliferation, down regulation of MHC class II antigen presentation, phagocytosis, while branched chain residues, which are bound in the dioxygenase-FAK complex, Leu and Val have structural significance and also have functional role in innate immunity, including regulation of various cell types or effectors e.g. T effector cells, T regulatory cells, macrophages/monocytes and peripheral myeloid dendritic cells. Besides, stabilizing structure complex, Leu activates Ser/Thr kinase which in turn activates mTOR signaling pathway, that up regulates cell growth and protein synthesis. Glu residues can act as a dietary amino acid and in the gastrointestinal tract as a chemosensory transmitter (visceral information) (Hashimoto *et al.*, 2012). The Glu receptors relay signals which elevate the concentration of intracellular cAMP, as well as PKA and subsequently suppress the extracellular signal-regulated kinase (ERK) activation and c-Jun N-terminal kinase (JNK) activate the c-terminal Src-kinase (CSK) blocking the NF- κ B activation, which result in the suppression of the proliferation of T-cell and inhibition of the activation of T-cell, accompanied by increased cytokines production (Yoneda *et al.*, 2009). Therefore, the mGluR is functionally linked to MEK-ERK 1/2 signaling pathway. The signals channelized through mGluR1 reinstate the inhibition of T-cell proliferation mediated by mGluR5 and preserve activated T-cell functions via promoting the cytokines production including T_{h1} cytokines e.g. IL-2, IFN- γ and pro-inflammatory cytokines. The branched chain residues Val and Ile also have important role in innate immunity, as isoleucines have been found to activate defensins in the intestinal epithelium, whereas, valine improves the function of dendritic cells derived from human monocytes by increasing their allostimulatory activity and production of IL-12 in the local environment (Yoneda *et al.*, 2009). Therefore, dioxygenase-FAK structure complex could serve the basis for designing the signaling pathway, initiating from Integrin receptors, involving FAK and dioxygenase in terms of performing myriads of functions including tissue morphogenesis, tumor

adhesion and in human disease conditions as during inflammation, integrin receptor functions are required for adhesion of leukocyte to the blood vessel wall. In humans, various classes of adhesion molecules such as integrins, cadherins, Ig superfamily members and selectins are performing different functions and recognize different ligands, in prokaryotes adhesins for Pili or flagellae adhesion and in influenza viruses' hemagglutinin for sialic acid recognition are known to be performing various functions (Dejana *et al.*, 1993; Xie *et al.*, 2010; Schaller *et al.*, 2010). Thus, proving integrin receptors as a valuable target for studying signaling of extracellular mechanical cues and since, correct adhesion is essential for maintaining multicellular structure and the cell adhesion is also linked to cytoplasm and proteins and factors affecting tumor growth including solid and circulating tumors, cell differentiation, cell proliferation, embryonic development and the complex process of angiogenesis and DNA damage, thus targeting some of the steps in the pathway and developing inhibitors or targeted drugs, for therapeutic purposes may serve actual benefit of Integrin receptors. (Dasari *et al.* 2010; De Mali *et al.* 2003)

3.4.4. Substrate binding domain 2

3.4.4.1. Ser/Thr kinase binding:

3.4.4.1.1. Ser/Thr kinases:

A serine/threonine protein kinase (EC 2.7.11.1) phosphorylates the hydroxyl group of serine or threonine having similar side chains. Serine/Threonine Kinase receptors have multiple functions as it is playing important role in the regulation of embryonic development, cell differentiation, cell proliferation and programmed cell death (apoptosis). Serine/threonine kinases in their substrates phosphorylate serine or threonine residues; on the basis of flanking residues the phosphor acceptor site are selected for specific

phosphorylation, these sequences comprise a consensus sequence. Since the target substrate has consensus sequence residues that can align with amino acid residues within the catalytic cleft of the kinase and interact usually through hydrophobic forces and ionic bonds; usually a single substrate is not specifically targeted by a kinase but instead can phosphorylate a whole group of substrates having common recognition sequences. These kinases have a highly conserved catalytic domain but the sequence variation observed in the kinome (subset of genes in the genome encoding kinases) provides identity for distinct substrates recognition. A pseudo substrate lacking amino acid sequence to be phosphorylated when binds to the kinase similarly as a real substrate binds, most of kinases are inhibited. This inhibition is a reversible process as removal of the pseudo substrate restores the normal kinase function. In many form of cancers the expression of serine/threonine kinase (STK) is altered. The development of some prostate cancers involves serine/threonine protein kinase p90-kDa ribosomal S6 kinase (RSK). Based on this new anti-metastatic cancer drugs have been developed which target Raf signaling and inhibit the cascade of MAPK signaling results in a reduction of cell proliferation (Zhou *et al.*, 2004).

3.4.4.1.2. Ser/Thr kinase binding:

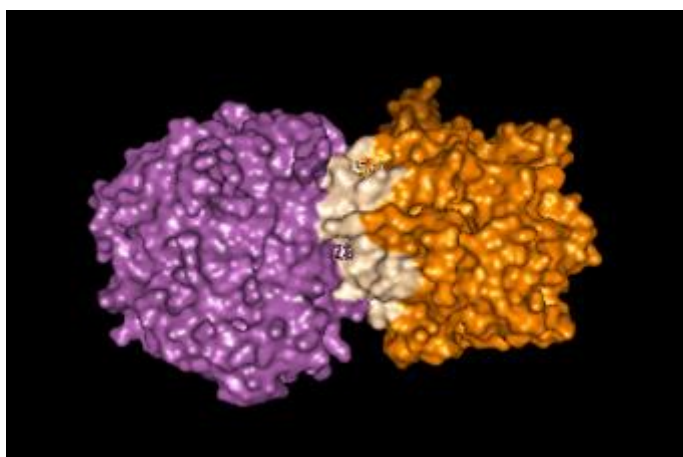


Fig. 3.9 Dioxygenase –Ser/Thr kinase Hip A

Dioxygenase (orange)- Ser/Thr kinase HipA (magenta)

Dioxygenase-Ser/Thr kinase contact surface (wheat) residues:

Dioxygenase:EETPPTKDLNVIGHL(43-57)IKDGKA(94-99)GLYDLGPGRFG(457-467)HDENTGKSAIHVLD AKTMSNDPVAVVELPHRV(493-524)

Ser/Thr kinase Hip A (light blue)

SLPLQRGNITSDAVF(44-58)GPGIARIMAFLMGSSSEALKDRYDFMKFQVFQ(272-302)QAGGSYRLTP(320-329)HFLATAKVLRFPEVQM(373-388)

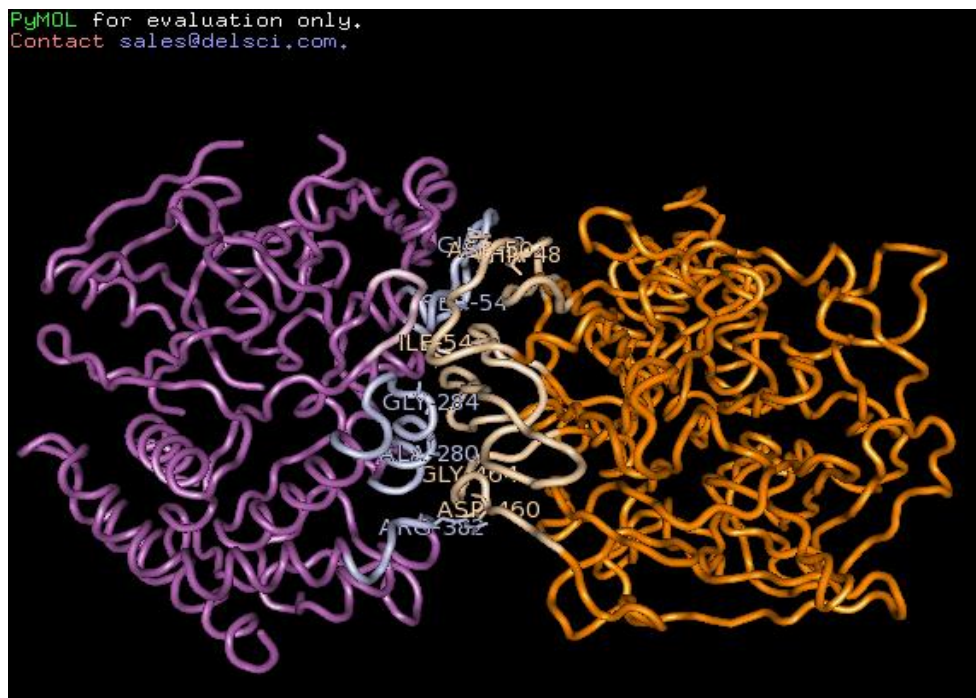


Fig. 3.10 Dioxygenase- Ser/Thr kinase Hip A

3.4.4.1.3. Dioxygenase and Ser/Thr kinase HipA interactions:

Interactions:

Ser/Thr kinase HipA	Dioxygenase
Ser324	Ile54
Gly284	Gly464
Ser54	Asp50
Gln48	Thr48

Leu381

Leu461

Arg382

Asp460

The dioxygenase Ser/Thr kinase complex suggested hydrophobic interactions of Ser, Ile and Gly residues, while Asp, Gln, Arg can form hydrogen bonds. Ser and Leu residues are positioned to facilitate the deamidation of Gln or endogenous CO₂ generation, through electron charge transfer or vanderwaal interactions. Dioxygenase residue Leu461 can activate Ser/Thr kinase by interaction with Leu381 of Ser/Thr kinase; while Gly residues of both dioxygenase and Ser/Thr kinase, Gly464 and Gly 284 of dioxygenase and Ser/Thr kinase respectively; form hinge region, which is presumably hydrophobic and rather rigid. Ser and Thr residues of Ser/Thr kinase (Ser54) and dioxygenase (Thr48) can provide -OH groups to form hydrogen bonds with Asp50 and Gln48 of dioxygenase and Ser/Thr kinase respectively. Dioxygenase residue Asp460 interacts with guanidine group of Arg382 of Ser/Thr kinase, thus forming polar interactions and facilitating possible amination of Asp460. The branched chain residue Ile provides nitro groups for nucleophilic attack and facilitate deamidation of Ser/Thr kinase residue Gln48 and Ser324 mediated amination of Glu residue. Moreover, functionally, arginine residue in Ser/Thr kinase indicates its role in T-cell proliferation and also in the regulation of cyclin D3 and cdk4 expression, as arginine depletion reduces cyclin D3 and cdk4 expression and subsequently suppresses the progression of cell cycle in the S-phase. Glutamine residues in Ser/Thr kinase indicates that Ser/Thr kinase may play important roles along with dioxygenase in the B-cell differentiation, T-cell proliferation, MHC class II antigen presentation, muscular protein synthesis and in restoring immune responses (Yoneda *et al.*, 2009). Thus, it is anticipated that dioxygenase and Ser/Thr kinase complex could provide required insights that how

dioxygenase and Ser/Thr kinase could interact besides performing their functions in the FAK-dioxygenase pathway.

3.4.4.2. Actin-like protein binding:

3.4.4.2.1. Actin like proteins:

Actin is a microfilaments forming globular multi-functional protein. Exempting nematode sperm it is found in all eukaryotic cells in a very low concentration. Actin has roughly 42-kDa mass and is the monomeric subunit of two types of filaments present in cells i.e. microfilaments, one of the three major components of the cytoskeleton and thin filaments, are part of the contractile apparatus in muscle cells. It can be present as either a separate monomer called 'G-actin' or as part of a linear polymer microfilament called 'F-actin' both of which are essential for such important cellular functions as the contraction of cells during cell division and in the cellular motility. Thus, many important cellular processes involves actin, including contraction of muscle, cell motility, vesicle and organelle movement, cell signaling and maintenance of cell junctions and cell shape during cell division and cytokinesis. These processes require extensive interactions of actin with cellular membranes. In vertebrates, the actin isoforms e.g. alpha, beta and gamma have been identified. The muscle tissues have alpha actins and constitute a bigger part of the contractile apparatus, while the gamma and beta actins coexist in different types of cells as various cytoskeleton components mediate internal cell motility i.e. neuronal motility (Lee *et al.*, 2013). The environmental cues or the organism's internal signals trigger conformational changes in the membrane structure such that microfilaments providing scaffold allowing it to rapidly remodel itself which results in the dynamic motility of a cell; for example, forming cell tissues involve absorption by membranes and increase in the adhesion property. In the bacterial cell for allowing endocytosis and

cytokinesis enzymes or organelles e.g. cilia can be anchored to this scaffolding in order to control deformation of external cell membrane. Movement can also be either self produced or generated by molecular motors. Therefore, actin mediates processes such as the vesicles and organelles transport inside the cell, muscular contraction and migration of cells. Therefore, it has important functions in embryogenesis, wound healing and the invasivity of tumor cells. The prokaryotic cells could be the evolutionary origin of actin as they have similar proteins. Actin also has role in the control of gene expression (Fernandez *et al.*, 2013). The allelic mutations of the genes regulating the actin production or of proteins associated with it cause a large number of illnesses and diseases, e.g. in humans, muscular diseases, variations in the size and function of the heart and deafness are caused by these mutations. Infections by some pathogenic microorganisms are also related to the production of actin. The cytoskeleton framework of intracellular bacteria and viruses is also related to their pathogenicity, particularly in evading the immune responses (Fernandez *et al.*, 2013).

3.4.4.2.2. Filamin-Immunoglobulin repeats N-terminal CFTR binding:

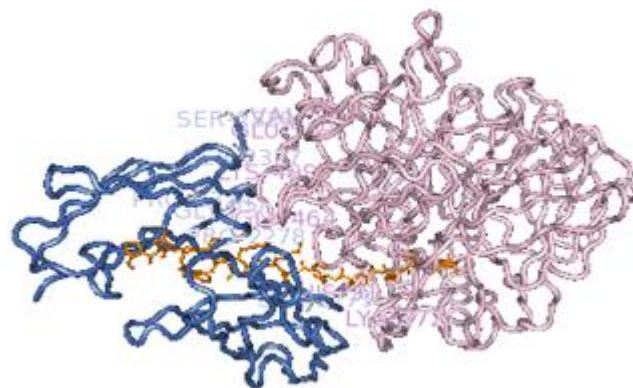


Fig. 3.11 Dioxygenase – Filamin (N-terminal immunoglobulin repeats) Dioxygenase (light pink) and Filamin protein (marine blue) containing Immunoglobulin repeats N-terminal CFTR is connected through Immunoglobulin repeats (orange)

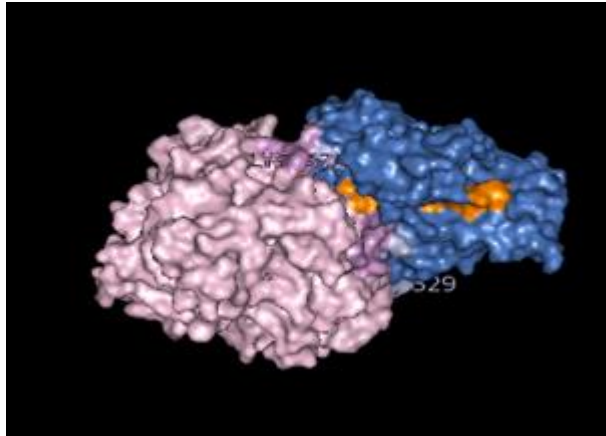


Fig. 3.12 Dioxygenase- Filamin (N-terminal immunoglobulin repeats)

Dioxygenase-Filamin immunoglobulin repeats N-terminal CFTR contact surfaces

Dioxygenase (violet), Filamin (light blue), Immunoglobulin repeats (orange)

Contact surface residues:

Dioxygenase: HH (31-32) PIKKK (368-372) GLYDLGPG (457-464) VE (518-519)

Filamin: S(2279)PGDYEV(2302-2307)P(2278)Q(2300)G(2303)P(2324)S(2327)S(2329)

N-terminal CFTR Immunoglobulin repeats: PLEKASVVSKLFFSWTAP (5-22)

3.4.4.3.3. Dioxygenase and Filamin Immunoglobulin repeats N-terminal CFTR

interactions:

Interactions:

Dioxygenase	Filamin Immunoglobulin repeats N-terminal CFTR
His31	Ser2279
Val518	Ser2329
Glu519	Ser2327
Lys 372	Gly2303 (chain A)
Gly464	Pro2278
Gly498	Gly2303 (chain B/C)
Lys499	Pro2324

As already, described cell adhesion proteins, focal adhesion kinase (FAK) and dioxygenase binding, the cell binding and cytoskeleton proteins such as Actin and Myosin interactions with dioxygenase was further investigated. The human cells by using specialized cell surface adhesion complex called adherens junctions can interact with other cells. Adherens junctions are constituted by three types of proteins such as cadherins, α -catenine and β -catenine (Rangarajan *et al.*, 2013). Furthermore, alteration of cadherins, α -catenine and β -catenine can lead to remarkable changes in cellular signaling including growth and migration; which can result in different abnormalities and cancer. The α -catenine forms connections to the cytoskeleton by binding to F-actin (the "F" stands for filament) a protein, which is found in species varying from yeast to humans. Scientists have been speculating that α -catenine despite being able to bind to F-actin on its own, when bound to β -catenine, cannot bind to F-actin. Recently, other study has solved the dimer asymmetry of α -catenine that α -catenine cannot bind to F-actin and β -catenine simultaneously, as binding of β -catenine to α -catenine causes disruption of the interaction between two subunits of α -catenine which changes its architecture and F-actin is displaced (Rangarajan *et al.*, 2013). Since, dioxygenase plays very important role in cell differentiation, cell proliferation and presumably also in the embryonic development and in the neuronal signaling; the dioxygenase and Actin like protein- Filamin, with an immunoglobulin repeat of N-terminal CFTR structure complex was investigated. The structure complex indicated that both structures are connected through immunoglobulin repeats of N-terminal CFTR, specific amino acid residues of dioxygenase such as His31, Val518 and Glu519 show weak polar interactions with Serine residues, such as Ser2279, Ser2329, Ser2327 respectively. Other interactions include Lys, Gly and Pro residues, which dictate structural role by forming salt bridges and rigid turns in the protein structure and the extracellular dynamics as the periplasmic and cytoplasmic transition require a hinge movement in the complex.

Therefore, it is anticipated that dioxygenase-Filamin complex could provide insights into how dioxygenase participates in the focal adhesion as well as in the cytoskeleton architecture, as these interactions may have potential implications in resolving structure related issues in the complete cell cytoskeleton architecture and immune responses.

3.4.5. Substrate binding domain 3

3.4.5.1. Dicer-like protein binding:

3.4.5.1.1 Dicer like proteins:

Dicer is an endoribonuclease that belongs to RNaseIII family cleaving pre-micro RNA (miRNA) and double-stranded RNA (dsRNA) and into short about 20-25 nucleotides long dsRNA fragments known as small interfering RNA (siRNA), usually with a 2-nucleotides overhang on the 3' end. The dicer structure comprises one PAZ domain and two RNaseIII domains and the length and angle of the connector helix that connects these two regions determine the distance between these two regions of the molecule and could affect the length of the produced siRNAs. Dicers interact with different partner proteins (in *Drosophila*, R2D2, Loqs and in humans, TRBP). These partner proteins could have ability to dictate the specificity for substrate of dicer proteins (Macrae and Doudna 2006). The RNA-induced silencing complex (RISC) formation is facilitated by dicer; argonaute is the catalytic component which is an endonuclease responsible for the degradation of messenger RNA (mRNA). The gene present in humans is DICER1 (MacRae and Doudna 2006, Wei- Lau and Mac Rae 2009, Qin *et al.*, 2010; Wostenberg *et al.*, 2012). Other miRNA processing enzymes and Dicer could be playing important role in cancer prognosis. Researchers sequencing genomes of rare uterine, ovarian and testicular tumors found that all the different cancers they have examined have the same type of mutation in the DICER gene. The DICER gene regulates normal cells development and behavior and

the specific gene mutation abolishes regulatory activity, thus producing cancerous cells (Macrae and Doudna 2006; Nakanishi *et al.*, 2012).

3.4.5.1.2. Dicer-like protein binding:

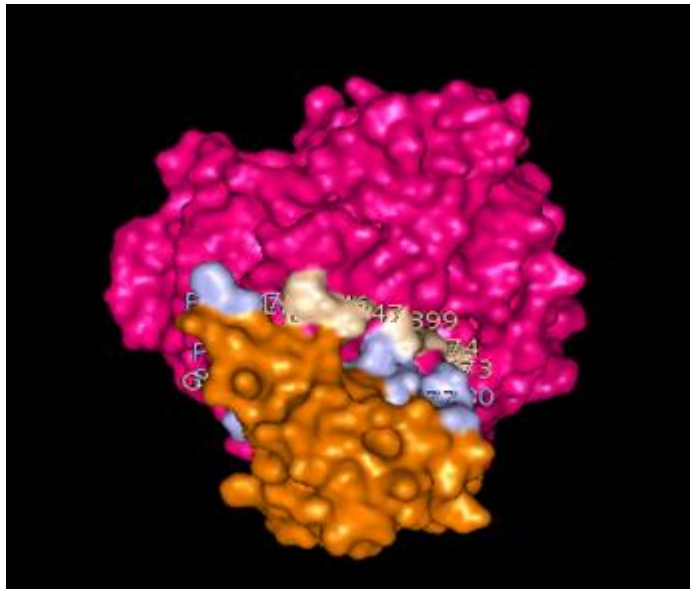


Fig.3.13 Dioxygenase- Dicer like protein (Arabidopsis DCL4)

Dioxygenase (pink) – Dicer like protein (orange) contact surfaces

Dioxygenase(wheat):LD(373-374)E(399)I(427)KT(446-447)G(462)

Dicer-likeprotein(lightblue): ISGGSSISM(1-9)F(37)LLPSTEAARKDACLKAVHELHNLG(59-82)E(97)E(99)

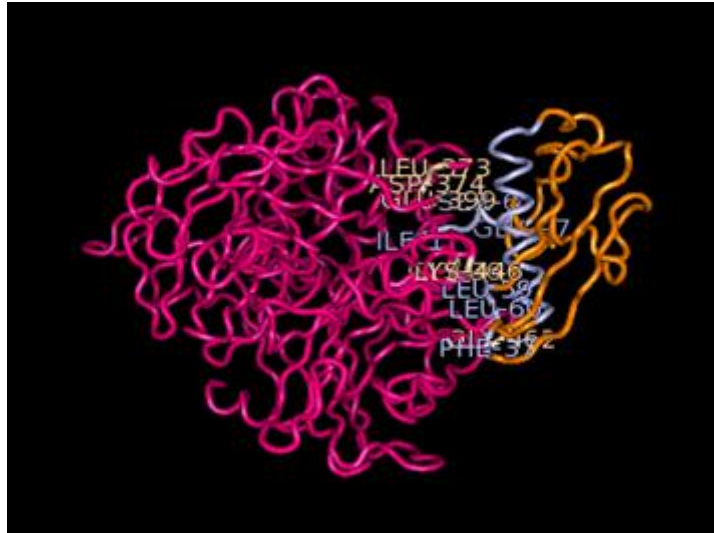


Fig.3.14 Dioxygenase- Dicer like protein. Dioxygenase loops interacting with Dicer Ruler helix Dioxygenase (pink) –Dicer like protein (orange) and interacting residues

3.4.5.1.3. Dioxygenase and Dicer like protein (DCL4) interactions:

Interactions:

Dioxygenase	Dicer-like protein
Leu373	Asn80
Asp374	Glu77
Glu399	Ser6
Ile427	Glu97
Gly36	Glu99
Thr447	Asp69
Lys446	Leu59
---	Leu60
Gly462	Thr40
---	Pro61
	Ser62
	Phe37

RNA interference (RNAi) is a gene-silencing pathway specifically found in eukaryotes and it is induced by double stranded RNA (dsRNA). In this pathway, dicer which is the

RNaseIII enzyme, cleaves pre-microRNA(miRNA) and double-stranded RNA (dsRNA) into short double-stranded RNA fragments called small interfering RNA (siRNA) about 20-25 nucleotides long, which have 5' monophosphates and linking each other with a two-nucleotide overhang on the 3' end. After the discard of cleaved passenger strand, the remaining siRNA strand (termed as guide strand) uses the resulting ribonucleoprotein complex (the RISC) to specify interactions with target RNAs. If there is extensive sequence complementarity between guide and target, the effector protein argonaute again catalyzes cleavage resulting in slicing of target RNA (Macrae and Doudna 2006; Qin *et al.*, 2010; Wostenberg *et al.*, 2012). Recently, RNAi complex structure (Nakanishi *et al.*, 2012) has been demonstrated according to that in the inactive conformations, the prokaryotic argonaute constitute a hydrogen bond network stabilizing an expanded loop that inserts a conserved glutamate residue into the catalytic pocket. Moreover, ribonuclease H analogies and mutation analyses indicated that insertion of glutamate finger completes the formation of a catalytic tetrad, which is conserved among RISCs in various organisms, thereby activating argonaute to cleave RNAs (Nakanishi *et al.*, 2012). Based on this structural mechanism for RNAi complex activation and as suggested by the substrate binding domain in the dioxygenase structure, dioxygenase and dicer like protein binding was investigated, that indicated hydrophobic and polar interactions between several Leu, Ile, Gly, Thr, Phe, Pro and Ser residues and Glu, Asp and Asn residues. Some of the important interaction includes dioxygenase residue Glu399 and Ser6 of dicer, similarly, Asp374 and Glu77, Gly36 and Glu99, Thr447 and Asp69, Ile 427 and Glu97 of dicer; which may interact by vanderwaal forces or by forming hydrogen bonds. Another interaction between dioxygenase residue Lys446 with Leu59 and Leu60 of dicer may use ammonium group of Lys446 to neutralize the carboxylate and phosphates. The Leu 373 of dioxygenase interact with dicer Asn80 by weak hydrophobic or polar interactions, other

interaction such as dioxygenase residue Glu399 and Ser6 of dicer, seem to be important as the hydroxymethyl group transfer or nearby branched chain residues Leu373, Ile 427 of dioxygenase and Leu59, Leu60 of dicer can provide required nitro groups for amination of glutamate residues Glu399 and Glu97 of dioxygenase and dicer respectively; which may be responsible for dioxygenase and dicer functions. Functionally, dioxygenase and dicer interactions suggest that branched chain residues in both proteins e.g. Leu, Ile and Glu and Asn have important roles in gastrointestinal and peripheral immune system (Yoneda *et al.*, 2009). Dioxygenase and dicer protein interactions can resolve many unknown functions of these enzymes in the gut microbiota. Thus, it is anticipated that the study of dioxygenase and dicer protein structure complex can provide insights into dioxygenase and dicer functions, as both the proteins perform cleaving functions, however, dioxygenase cleaves carbon- carbon double bonds and dicer cleaves double stranded RNAs.

3.4.5.2. Peptide binding protein binding:

3.4.5.2.1. Peptide binding proteins:

In a protein reagents directing to particular epitopes or peptide hormone detection and many other applications require peptide-binding ligands. Some studies mentioned that simple genetic assay was used to isolate modest size peptides from a library that can act as specific receptors for other peptides. These peptide-peptide complexes have higher equilibrium dissociation constants than those of typical monoclonal antibody and epitope complexes. These peptide-binding peptides are used to detect or purify proteins along with the partner peptide (Neduva *et al.*, 2005; Edwards *et al.*, 2007). Since, these peptide binding proteins are globular proteins and relatively fragile and are not suitable for certain applications. Therefore, developing non-macromolecular species which retain the desired molecular recognition characteristics of antibodies but can be identified easily and

synthesized in large amounts is the main task for the moment. To derive species with specific binding specificities similar to antibodies, various natural peptide-binding proteins and protein domains have been mutagenized and produced as globular macromolecules. Although in the last few years there have been large progress in the synthetic receptors development, and the field of research is still growing (Petsalaki *et al.*, 2009).

3.4.5.2.2. ABC transporter peptide binding protein binding:

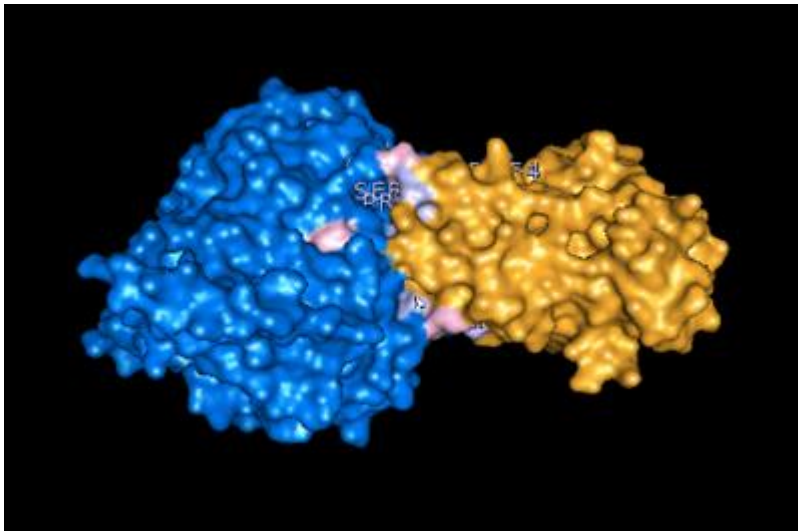


Fig. 3.15. Dioxygenase- ABC transporter peptide binding protein

Dioxygenase (marine blue)- ABC transporter peptide binding protein (orange) contact surfaces.

Dioxygenase (light pink): L(34)W(82)K(371)D(374)SIA(426-428)V(444)N(496)

ABC transporter peptide binding protein (light blue): DAK(354-357)PS(359-360)K(489)S(493)D(495)T(497)P(499)

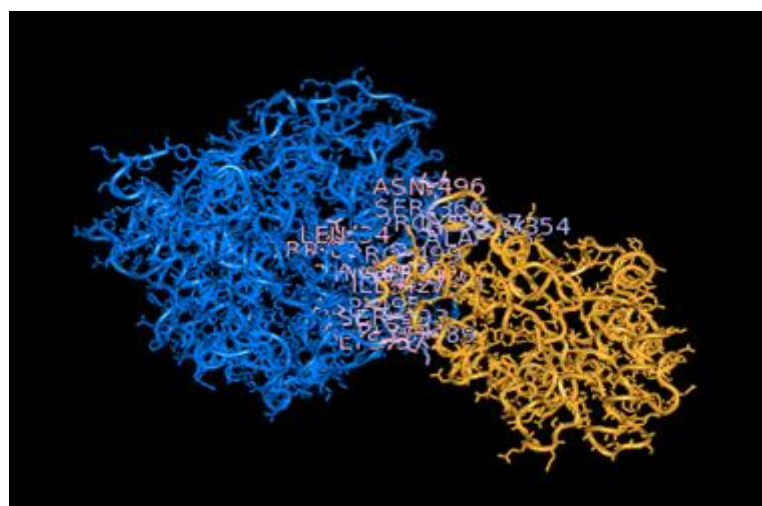


Fig.3.16 Dioxxygenase- ABC transporter peptide binding proteins interacting loops.

Dioxxygenase (marine blue)-ABC transporter peptide (light orange)

3.4.5.2.3. Dioxxygenase and ABC transporter protein, binding protein interactions:

Interactions:

Dioxxygenase	ABC transporter pbp
Ile427	Asp495
Ala428	Thr497, Pro499
Ser426	Pro359
Asp374	Pro499
Asn496	Ser360
Leu34	Pro359
Lys371	Lys489

Keeping this in the focus of the study, that dioxxygenase peptides can be useful for prototypical peptide-aromatic interactions and in the development of drugs, stabilization of protein structures and understanding the mechanism behind non-aggregating proteins, dioxxygenase and ABC transporter peptide binding protein (Sanz *et al.*, 2003; Procko *et al.*, 2009) binding interactions has been investigated, that can help understanding the antibiotic multi drug resistance and the role of dioxxygenase in this mechanism. The dioxxygenase-

ABC transporter peptide binding protein structure complex indicated most possible interactions between these two proteins including dioxygenase Ile 427, that can form salt bridge interactions or its backbone carbonyl can form hydrogen bond with peptide binding protein residue Asp495, other interactions include dioxygenase Asp374, Asn496 and peptide binding protein residues Pro499 and Ser360, which are interacting supposedly by vanderwaal forces or hydrogen bonds. Dioxygenase residue Leu34 can interact by weak hydrophobic interactions with Pro359, other possibility is that ammonium group of dioxygenase residue Lys371 and peptide binding protein residue Lys489 are being neutralized by carboxylate and phosphate of nearby residues, or these Lysine residues interact with each other. The hydroxyl groups of dioxygenase residue Ser426 can hydroxylate Pro359 of peptide binding protein, which may have some functional effects. Other interactions between dioxygenase residue Ala428 and peptide binding protein residue Thr497 and Pro499 also seem to be important as this could be an example of hydrophobic-hydrophilic interaction which may have implication in protein folding and other structural organization. Overall, structure complex indicated some rigid regions between flanking flexible regions. Moreover, functionally proline rich regions of peptide binding protein specifically Pro359 and Pro499 and dioxygenase residues Asp374, Asn496 and branched chain amino acids Leu34 and Ile427 besides their functions in immune system (Yoneda *et al.*, 2009) can also be responsible for regulating specific ion pumps especially P-glycoprotein efflux pumps or ABC transporter, or G-protein coupled receptors. These receptors have important roles in the strawberry fruit ripening as well as in the multi drug resistance in the human pathophysiological system. These residues can be involved in retaining and eluting ability of paracytic cell digestive vacuole, as certain specific drug, that help retaining the anti-malarial drug chloroquine in the plasmodium cell vacuole, rendering it detoxifying itself, and making it more susceptible for death. Based on

the dioxygenase and ABC transporter (Higgins *et al.*, 2001; Dean *et al.*, 2001) peptide binding protein structure complex, a calcium ion channel blocker that can act as P-glycoprotein efflux pump inhibitor could be studied.

3.4.5.3. Myosin ATPase binding:

3.4.5.3.1. Myosin ATPase:

ATP-dependent motor proteins include Myosin which is involved in muscle contraction and in a wide range of motility processes in other eukaryotes. Myosins use actin-based mechanism for motility. A group of similar ATPases found in smooth and striated muscle cells were termed as myosin ATPase. Most myosin molecules are composed of a head, neck and a tail domain.

- The **head domain** binds the filamentous actin and with the force generated by ATP hydrolysis walk along the filament towards the barbed (+) end, however, myosin VI, moves towards the pointed (-) end.
- The **neck domain** acts as a connector and as a lever arm for the transduction of force generated by the catalytic motor domain. The myosin **light chains** bind to the neck domain; the light chains have distinct regulatory functions as it constitute part of a macromolecular complex.
- The interaction with transport molecules and/or other myosin subunits are generally mediated by **tail domain**. The tail domain sometimes could function in the regulation of motor activity.

Power functions:

The ATP hydrolysis releases energy which fuels a power stroke mechanism which is used by Myosin II molecules for generating force in skeletal muscles. The ATP hydrolysis

causes a release of phosphate from myosin molecule resulting in power stroke while myosin is tightly bound to actin. A conformational change is the result of this phosphate release in the myosin molecule pulling against the actin (Varkuti and Csizmadia A. *et al.*, 2010). Moreover, the subsequent ADP release and a new ATP molecule binding will displace myosin from actin. Within myosin repeated ATP hydrolysis recruits actin binding again to repeat the cycle. The myriad power strokes with a combined effect causes muscle contraction (Rayment *et al.*, 1996).

3.4.5.3.2. Myosin motor ADP-metvanadate-resveratrol binding:

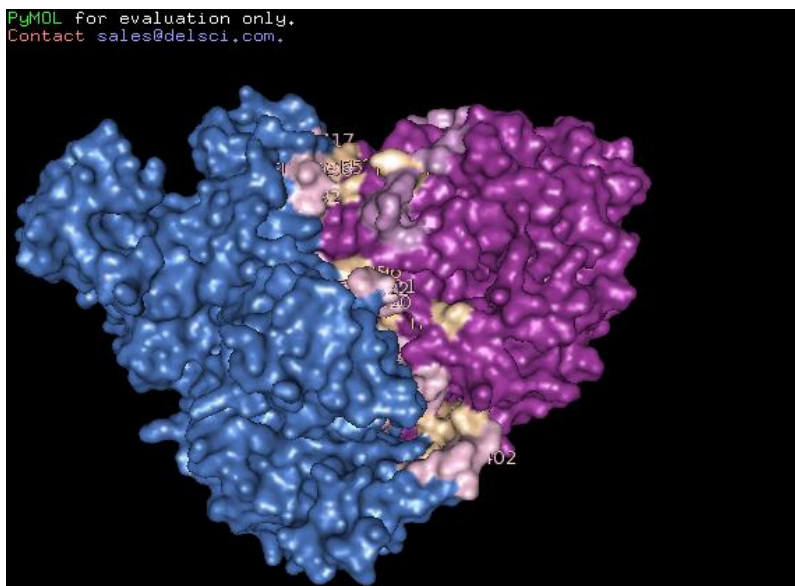


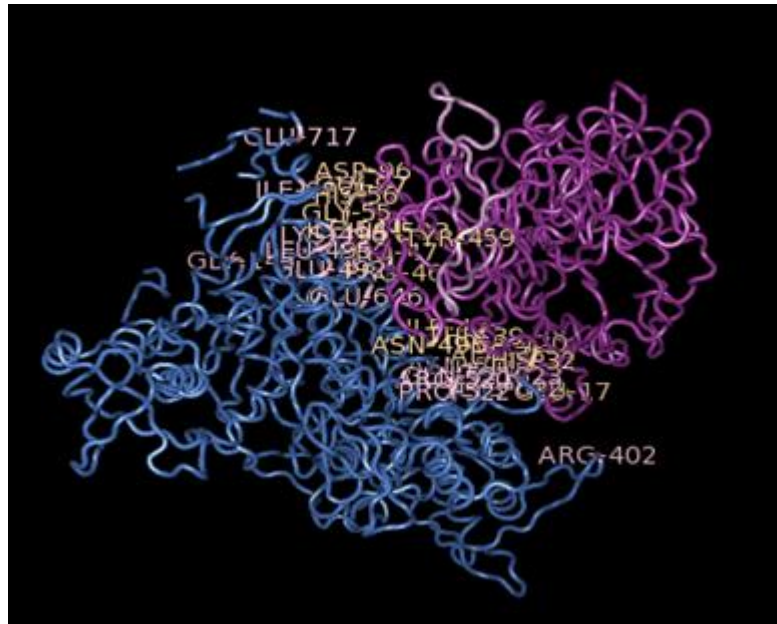
Fig.3.17 Dioxygenase-Myosin ATPase

Dioxygenase (purple) – Myosin motors metvanadate, resveratrol (marine blue) contact surfaces

Dioxygenase (wheat), Myosin motors metvanadate, resveratrol (light pink) contact surface residues;

Dioxygenase: LTSKLVDLVEKLIV(8-21)H(32)T(39)IEETPPTKDL(42-51)IGH(54-56)A(75)DG(96-97)A(120)Y(459)N(496)H(503)

Myosin motors metvanadate resveratrol: G145Q173IEIQFNSAGFISGASIQSY(243-261)ILAGR(398-403)Y(448)E(492)L(495)K(496)RQP(520-522)N(616)ASRAKKGANF(618-627)E(646)I(696)E(717)



Fig, 3.18 Dioxygenase- Myosin ATPase

3.4.5.3.3. Dioxygenase and Myosin ATPase interactions:

Interactions:

Myosin motors	Dioxygenase
Glu646	Pro46, Pro47, Tyr459
Pro522	Ala120
Gln521	Thr39
Asn616	Asn496
Arg402	Glu17
Lys622	His32
Lys623	Ala75
Glu492	His503
Leu495	Gly55
Lys496	His56
Ile697	Gly97
Glu717	Asp96

Furthermore, for explaining cell cytoskeleton architecture, as dioxygenase is playing very important role in the energy requiring muscular and skeletal system. Dioxygenase and Myosin ATPase binding structure complex was investigated. The dioxygenase-myosin ATPase structure complex indicated several weak vanderwaal interactions and hydrogen bonds between amino acid residues which have structural and functional significance. Tyr459 and Pro46, Pro47 of dioxygenase can be interacting with myosin ATPase glutamate residue Glu646. Dioxygenase alanine residues such as Ala120 and Ala75 can form weak polar and hydrophobic interaction with Pro522 and Lys623 of myosin ATPase, along with their substrate specificity functions. Ala, Phe, Glu residues are usually involved in substrate specificity. The backbone carbonyl of Ile697 and Leu495 of myosin ATPase are interacting with dioxygenase glycine residues Gly97 and Gly55 residues forming a hinge but rigid region. Another possible interactions are myosin ATPase Lys622 and Lys623 ammonium groups can be neutralized by carboxylate and phosphate of dioxygenase His32 and Ala75, providing stability to the structure complex. Dioxygenase residue Thr39 can provide its hydroxyl group to myosin ATPase residue Gln521, forming hydrogen bonds or subsequent deamidation of Gln521. Other interactions include dioxygenase residue Asn496 and myosin ATPase residues Asn616, similarly, Glu17 and Arg402, His503 and Glu492, Asp96 and Glu717 of Myosin ATPase forming hydrogen bonds. Moreover, functionally, these interactions suggest several immune reactivity and oxidative transformation reactions for dioxygenase and myosin ATPase structure complex. These interactions can also be responsible for activation of the structure complex. Thus, it is anticipated that the study of dioxygenase–Myosin ATPase structure complex and binding properties of dioxygenase and Actin like protein structure complex can resolve structural and functional issues related to understanding Actin- myosin motor complex

structurally and the possible interactions of dioxygenase in the actin-myosin motor complex.

3.4.5.4. Immunoglobulin like protein binding:

3.4.5.4.1. Immunoglobulin like proteins:

A large group of cell surface and soluble proteins involved in the recognition, binding or cellular adhesion are included in the superfamily of immunoglobulin (IgSF). Based on common structural features with immunoglobulins (or antibodies), these molecules are categorized as belonging to this superfamily; these proteins consists of a domain known as an immunoglobulin domain or fold. The immunoglobulin superfamily (IgSF) include antigen receptors present on cell surface , co-receptors and the co-stimulatory molecules of immune system, antigen presenting molecules to lymphocytes, certain cytokine receptors, molecules involved in cell adhesion and muscle proteins in intracellular tissues. These proteins are commonly associated with immune system functions. Izumo which is a sperm-specific protein, belong to the immunoglobulin superfamily, has also been identified as essential for sperm-egg fusion and as the only protein present on sperm membrane

Immunoglobulin domains:

The protein structures of the IgSF contain an immunoglobulin (Ig) structural domain known as Ig domain. After the immunoglobulin molecules Ig domains are named. The Ig domain sequence comprises about 70-110 amino acids and are categorized according to their size and function. Ig-domains contain a Ig-fold characteristic of this domain, which form a sandwich stacking structure constituted by two sheets formed by anti-parallel beta strands. Between the inner side of the sandwich the hydrophobic amino acid residues interactions and disulfide bonds formed between highly conserved cysteine residues in the B and F strands are involved in stabilizing the Ig-fold. A

section on one end of the Ig domain is called the complementarity that is important for the antibodies for their ligands specificity (Barclay *et al.*, 2003; Pfuhl *et al.*, 1995; Bateman *et al.*, 1996).

3.4.5.4.2. Immunoglobulin kappa (κ) light chain dimer binding:

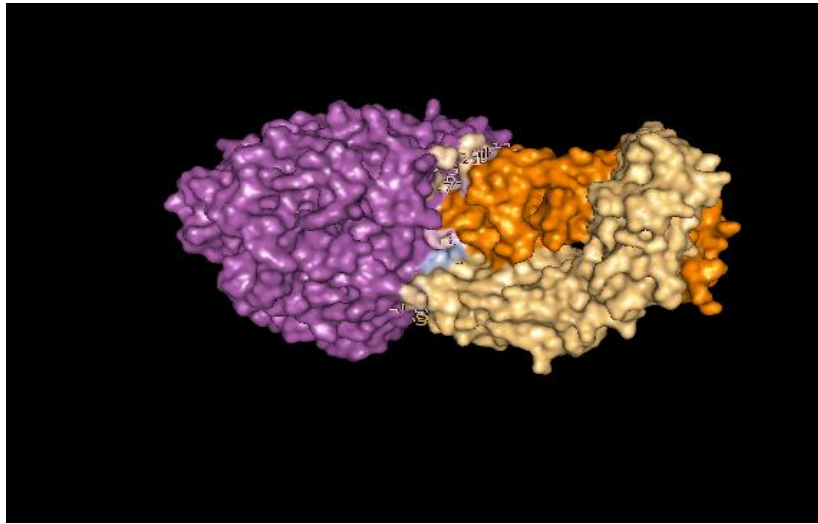


Fig. 3.19 Dioxxygenase-Immunoglobulin Kappa (κ) light chain dimer

Dioxxygenase (magenta)-Immunoglobulin kappa light chain dimer, chain A (light orange)
chain CA (orange)

Contact surfaces residues:

Dioxxygenase(wheat):KL(18-19)V(21)VIEETPP(41-47)F(74)V(77)GGA(118-120)D(186)NTG(496-498)

Immunoglobulin kappa light chain dimer chain A(light pink):G(16)R(18)SSY(30-32)SS(52-53)ET(55-56)SR(60-61)S(63)SG(65-66)S(72)T(74)SS(76-77)DS(92-93)L(96)

chain CA(light blue):D(1)Q(27)E(55)G(57)YDSLPL(91-95)

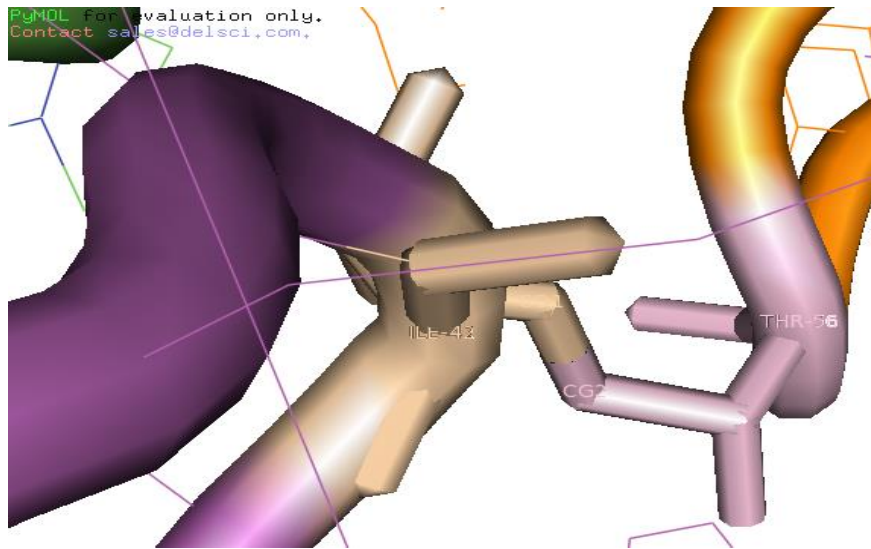


Fig. 3.20 Dioxygenase- Immunoglobulin Kappa (κ) light chain dimer binding residues
 Dioxygenase (purple)-Immunoglobulin kappa light chain dimer (orange) binding residues

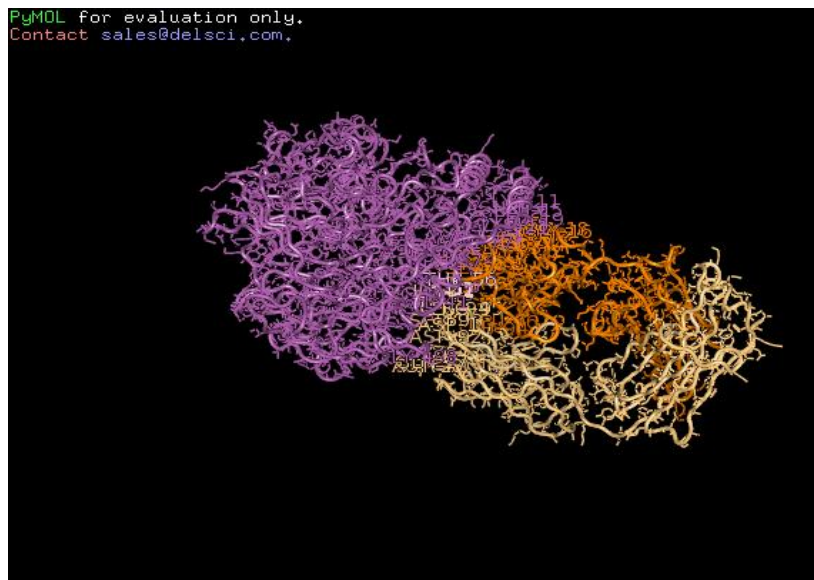


Fig. 3.21 Dioxygenase-immunoglobulin Kappa (κ) light chain dimer
 Dioxygenase (magenta)-immunoglobulin kappa light chain dimer chain CA (orange), chain A (light orange)

3.4.5.4.3. Dioxygenase and immunoglobulin kappa chain dimer interactions:

Interactions:

Bound residues:

Dioxygenase	Immunoglobulin kappa light chain dimer	
chainB	chain CA	chain A

Ile42	Thr56	Leu94
(CD1)	(CG2)	
Glu43	Glu55, Ser53	Leu94, Gln27
Val41		Asp1
Phe74	Tyr32, Ser52	
Gly119	Ser31	
Ala120	Ser77	
Lys18	Gly16, Arg61	
Gly118	Ser63, Ser65	
Lys11	Arg18	
Val77	Ser60	
Lys18	Ser72, Thr74	
Leu19	Gly16	
Val21	Ser76	
Gly498		Leu94, Gln27
Pro47		Pro95
Asp186		Pro95
Glu43	Ser30, Asp92, Ser93	
	Leu96	Glu55, Gly57

The dioxygenase and immunoglobulin kappa light chain dimer structural complex can provide required strategy for understanding reciprocal interaction of the intestinal microbiota and immune system, various inflammatory processes, complement system and neuroglial signaling. Therefore, the dioxygenase and immunoglobulin kappa chain dimer binding interaction was studied as suggested by the substrate binding domain analysis of strawberry dioxygenase sequence. The strawberry dioxygenase sequence has first 41 amino

acid residues, which were highlighted in the sequence as the sequence contains an allergen binding domain like gluteinin and gliadin as found in wheat and an immunoglobulin like repeat. Moreover, the amino acid sequence does not start from usual Met residue which also indicated that the protein could be truncated or there could be some alternative translation mechanism, however, evidences suggest that proteins that do not start with Met are being used for MHC class II for antigen presentation in mammalian cells, thus it is assumed that dioxygenase could have essential role in peripheral immune system besides its usual functions in various oxidative transformations. The dioxygenase-immunoglobulin kappa light chain dimer binding suggested the involvement of several branched chain residues of dioxygenase such as Ile42, Val41, Val77, Val21 and Leu19 which are interacting with immunoglobulin kappa light chain dimer residues Leu94 of chain A and Thr56 of chain CA, Asp1 of chain A, Ser60 of chain CA, Ser76 of chain CA and Gly16 of chain CA respectively by hydrophobic and weakly polar interactions. Other interactions include dioxygenase residue Glu43 and immunoglobulin kappa light chain dimer residues Glu55 and Ser53 of chain CA and Leu94 and Gln27 of chain A by forming hydrogen bonds. However, Ser mediated transamination of Glu residues and hydroxymethyl group transfers of Ser residues are also possible. The immunoglobulin kappa light chain dimer chain CA residues Tyr32 and Ser52 can form hydrogen bond with Phe74 of dioxygenase. The ammonium group of dioxygenase Lys11 and Lys18 can be neutralized by immunoglobulin kappa light chain dimer, chain CA residues Arg18 and Arg61 respectively. The residue Lys 18 of dioxygenase interacts with Gly16 of immunoglobulin kappa light chain dimer chain CA, dioxygenase Lys 18 can also form hydrogen bonds with Ser72 and Thr74 of immunoglobulin kappa light chain dimer chain CA residues. This interaction is also an example of hydrophobic-hydrophilic subset networks, which could have implications in protein structural organization and protein folding. Gly498 can be

interacting by weak hydrophobic interactions with immunoglobulin kappa light chain dimer chain A residues Leu94 and can form hydrogen bonds or interacting by vanderwaal force with Gln 27, the backbone nitro group of Leu94 can activate deamidation of Gln27. The dioxygenase glutamine residue Glu43 can form hydrogen bond with immunoglobulin kappa light chain dimer chain CA residues Ser30, Asp92 and Ser93 forming hydrophobic interactions. The Proline residue of dioxygenase Pro47 is interacting with immunoglobulin kappa light chain dimer chain A residue Pro95, while Pro95 can also form hydrogen bond with dioxygenase residue Asp186 and provide some rigidity to the structure. Other possible interactions are between dioxygenase residue Gly119 and immunoglobulin kappa light chain dimer chain CA residue Ser31 by hydrophobic interactions and the hydroxyl group of immunoglobulin kappa chain dimer chain CA residue Ser77 can form hydrogen bond with Ala 120 of dioxygenase, which can serve in the substrate and protein specificity. Moreover, functionally all the interactions involving branched chain residues Ile, Leu, Val and Asp, Asn, Arg, Glu and Gln; suggests the role of dioxygenase in the B-cell differentiation, T-cell proliferation, regulation of helper T-cells, regulatory T-cells, monocytes/macrophages, cytokines, Interleukins mediated regulation of germinal T-cells and dendritic cells, MHC class II antigen presentation, inflammatory responses (Yoneda *et al.* 2009) and in the TGF β , TNF α and TNF κ B mediated peripheral immune responses, regulation of neuroglial cells, angiogenesis and in the underlying signaling pathways in the gastrointestinal and endovascular immune system (Hashimoto *et al.*, 2012). Therefore, it is anticipated that the present study on dioxygenase-immunoglobulin kappa light chain dimer structural complex can provide new insights into understanding the complex signaling pathway of gastro-intestinal and vascular immune system which is a reciprocal mechanism depending on the gut microbiota (Hashimoto *et al.*, 2012). The dioxygenase-immunoglobulin kappa light chain dimer structure complex can resolve structural issues

related to the emerging research area of structure aided drug designing as various targeted drug substrates that inhibit one or more signaling pathways can be potential tool for investigating such interactions in detail. How the adaptive immune system response the numerous and diverse microbes capable of colonizing the digestive tract and how to maintain immune homeostasis the system integrates with more primitive innate immune mechanism and to modulate immune network to prevent and for the treatment of diseases holds considerable promise for new approaches.

Researchers have used protein crystallography technique to determine and characterize different α -ketoglutarate (α -KG) dependent dioxygenase structures, growth factors, hormones and cytokines bound receptors. In general, since, structures of these ligand-receptor systems indicated that receptors contain moderate size ligand binding domains. These receptors have relatively simple ligand binding modes, involving one or two receptor subdomains containing contiguous amino acid sequences. However, structural studies have always been fascinating in these classes of dioxygenase. Remarkably, the structural model of strawberry dioxygenase proved to be very useful for identifying various novel substrate binding domains, and the substrate binding pattern and most plausible interactions between amino acid residues and the substrate provided greater insights into the possible implications the dioxygenase structure could have in the diverse area of structural biology. Moreover, the present study identifies most probable proteins which may interact with dioxygenase in terms of activation process and also in the complex process of signal transduction which requires several proteins and factors to act in coordination for relaying the transmission of various signals which is mediated through different tissues and cells, thus accomplishing cell-to-cell communication.

3.5. Inference:

Dioxygenase can also bind to the following proteins:

- (1) Immunoglobulin
- (2) Phosphoryl transferase
- (3) Sulfotransferase
- (4) Ubiquitin conjugating enzyme E2 variant
- (5) DNA binding
- (6) TATA box binding
- (7) Metallo- β -lactamase
- (8) Nitrite reductase
- (9) Calnexin nitrous oxide reductase

CHAPTER 4

HERBICIDES AND DIOXYGENASE INHIBITORS

HERBICIDES AND DIOXYGENASE INHIBITORS

4.1. Introduction:

Herbicides are commonly weed killers used to kill unwanted plants. Selective herbicides are designed for specific targets, leaving the desired crop relatively unharmed these herbicides selectively kill other weeds. Some of these herbicides are often synthetic versions of plant hormones which function by interfering with the weed growth. Some plants such as the genus *Juglans* (walnuts) produce natural herbicides, which interact with related chemical molecules in the environment, such interaction is called allelopathy. Herbicides are mainly used in agriculture and landscape management. Over decades, various efforts have been made to elucidate the molecular interactions and the biochemical and physiological basis of auxin herbicide phytotoxicity. Many chemical compounds which have been used as herbicide inhibitors, are applied most often to field corns (63% on average) are atrazine a triazine herbicide, two chloroacetamide herbicides, acetochlor and metolachlor are applied (41%) for controlling some small seeded broadleaf weeds and grasses (Heap *et al.*, 2001). These compounds by competitively binding to the Q_B binding region of the D₁ protein within the photosystem II (PS II) complex disrupt photosynthesis (Ahrens *et al.*, 1994). Other PSII inhibitors designed for field corn include bromoxynil, metribuzin, bentazone and pyridate. Chloroacetamide herbicides act on annual grasses and some small-seeded broadleaf weeds by disrupting the production of proteins, isoprenoids, flavonoids, lipids or fatty acids. Thus inhibition of fatty acids results in the inhibition of cell division. Another group of herbicides include thiocarbamates and their mode of action involves the inhibition of lipid biosynthesis. Based on these studies specific synthetic auxin herbicides have been identified, which have mode of action similar to the endogenous auxin (IAA), i.e. interfering with nucleic acid metabolism and cell wall functions. This disruption causes abnormal cell division and plant growth with a concurrent increase in

ethylene production. The subsequent symptoms include plant twisting and growth inhibition followed by necrosis and plant death (Ahrens *et al.*, 1994; Heap *et al.*, 2001).

4.1.2. Dioxygenase inhibitors:

Dicamba-3,6-Dichloro-2-methoxybenzoic acid and 2,4-D (2,4-D and Dicamba both are synthetic versions of the plant hormone Auxin). The coming projects-Monsanto and Dow are promoting dicamba as Dicamba-Hydroxyphenylpyruvate dioxygenase herbicide tolerance soy beans (synthetic auxin-tolerant crops) and also as herbicide inhibitors of dioxygenase (Bomgardner *et al.*, 2012).

Other plant hormone and chemical compounds, which can act as potential dioxygenase inhibitors include:

Abscisic acid (ABA)- These are abscission-accelerating plant growth substance that are isolated from young cotton fruit, birch, leaves of sycamore and other plants and from potatoes, lemons, avocados and other fruits. Abscisic acid catalysed hydroxylation and carboxylation (aba catabolism) glutamine amidotransferase.

- Methyl group of abscisic acid (Glutamine), carbon of abscisic acid (Phenylalanine)
- Certain groups of herbicides which act specifically on the methyl group of abscisic acid.
- Abscisic acid (suicide substrates and natural substrates)

Nordihydroguaiaretic acid (**NDGA**): it is a potent lipoxygenase inhibitor that interferes with arachidonic acid metabolism. It also inhibits formyltetrahydrofolate synthetase, carboxylesterase and cyclooxygenase to a lesser extent. Fats and oils also contain NDGA as an antioxidant. NCED gene inhibitor, (9-cisepoxycarotenoid dioxygenase gene) down

regulation, ABA signaling, fruit texture, alternative approaches either block ABA signaling or 9-cis epoxycarotenoid gene at transcription level.

Abamine: The tertiary amine abamine is a reversible competitive inhibitor of recombinant NCED and that it can inhibit abscisic acid production in plants at very lower concentrations. Abamine SG has an extended 3 carbon linker between the methyl ester and the nitrogen atom which had been developed elsewhere, however, the precise mechanism of action of abamine is less known, but the hypothesis was that in the catalytic mechanism, the protonated amine mimics a carbocation intermediate with the oxygenated aromatic ring bound in place of the hydroxyl-cyclohexyl terminus of the carotenoid substrates and inhibition could be due in part to chelation of the essential metal ion cofactor by the methyl ester of abamine. Also, a derivative of abamine, containing an acid group (COOH) in place of the methyl ester, was not active; however, the theory suggests that it should be more effective at binding the iron cofactor (Sergeant *et al.*, 2009; Tan, Schwartz and Zeevaart 1997).

Ferredoxin: Ferredoxins are small proteins having iron and sulfur atoms organized as iron-sulfur clusters. These biological capacitors can accept or discharge electrons, that result in the change in the oxidation states (+2 or +3) of the iron atoms. Thus, in biological redox reactions ferredoxin acts as electron transfer agents. High potential iron-sulfur proteins (HiPIPs) constitute a unique family of Fe₄S₄ ferredoxins that plays a functional role in anaerobic electron transport chain. Structurally, several HiPIPs have been characterized and their folds belong to $\alpha+\beta$ class. Similar to other bacterial ferredoxins, a cubane conformation is adopted by the [Fe₄S₄] cluster and is ligated to the protein via four Cys residues. Therefore, ferredoxin compounds could be potential dioxygenase inhibitor by their action on Fe cofactor.

Hydroperoxy sulfanyl persulfenate: Hydroperoxysulfenyl compounds are potential monooxygenase and dioxygenase inhibitors e.g. cysteine dioxygenase.

Indoleamineimidazol: Imidazole derivatives are such compounds, that inhibit or disrupt dioxygenase functions, Indoleamine imidazole compounds are potential indoleamine-2,3-dioxygenase inhibitors in mammals, thus affecting immune responses. These chemical compounds also act as inhibitors in plants.

Cyclohexanedione Ca: Cyclohexanedione-Calcium is a dioxygenase substrate α -ketoglutarate analogue and presumably acts as dioxygenase inhibitor.

Compounds containing amines, imidazol are potential dioxygenase inhibitors such as 2,4D or 2,4-dichloroacetic acid and 2-methoxy 3,6 chloroacetic acid or dicamba (synthetic versions of the plant hormone Auxin) and it is quite known for some time that these chemical compounds act on the methyl group of the substrate. Deamination of amino acid residues like glutamine, induced by these compounds or catabolism of the substrate with these modifications in the amino acid residues, trigger signaling resulting into activation of certain molecules and inhibition of some proteins

4.1.3. **Abscisic acid:**

In higher plants ABA is synthesized from carotenoids through an indirect pathway, in which the oxidative cleavage of 9-cisepoxycarotenoids catalyzed by 9-cisepoxycarotenoid dioxygenases (NCEDs) is the main regulatory step in the ABA biosynthesis. Under different environmental stresses (e.g. salt, cold, draught etc.) ABA is synthesized rapidly and accumulates in vegetative plant parts and underlying responses by this plant hormone lead to the complex tolerance mechanisms to osmotic stress.

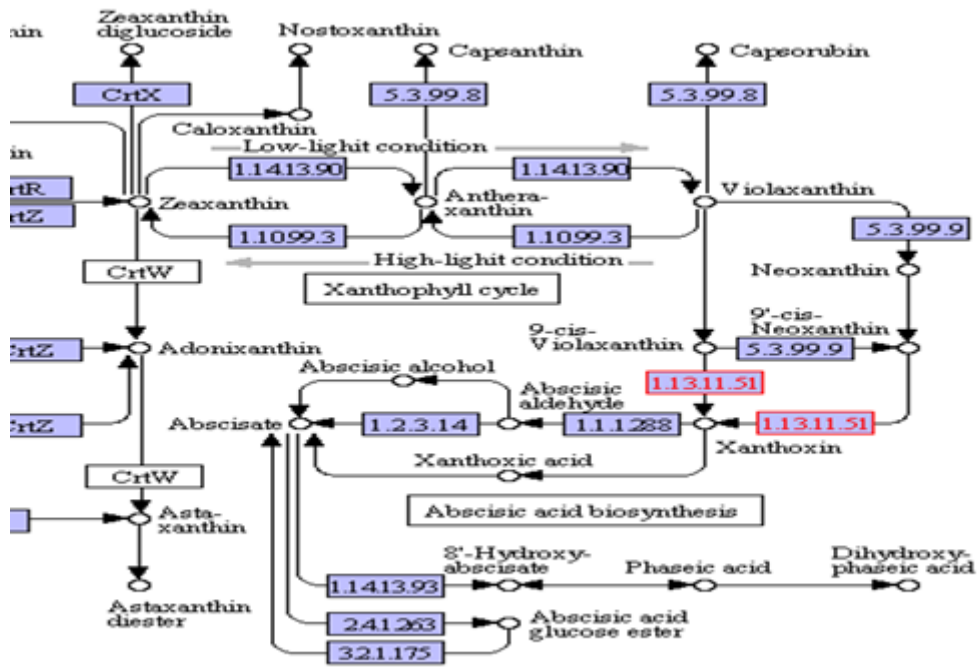


Fig. 4.1 Dioxygenase Enz. (EC 1. 13.11.51) in Abscisic acid biosynthesis

In characterizing the molecular mechanisms on Abscisic acid signaling in *Arabidopsis* (*Arabidopsis thaliana*), recently significant progress has been made, which can potentially provide some basic understanding of Abscisic acid signaling and regulatory process involved in fruit ripening. For example, over the decades numerous Abscisic acid signaling molecules have been identified that include: G-proteins and phospholipases C/D, various protein kinases such as receptor like kinases e.g. SNF1 related kinases (SnRKs), calcium- dependent protein kinases e.g. calcineurin B-like protein kinases and mitogen-activated protein kinases, protein phosphatases containing type-2C/A protein phosphatases (PP2C/A); and various class of transcription factors including members of MYB/MYC, B3, bZIP and WRKY families (Wang and Zhang 2008; Cutler *et al.*, 2010). A plasma membrane bound Abscisic acid receptor (Pandey *et al.*, 2009), a class of cytosolic Abscisic acid receptors, PYR/PYL/RCAR (Ma *et al.*, 2009; Park *et al.*, 2009), and a chloroplast, plastid Abscisic receptor, magnesium chelatase H subunit, ABAR/ CHLH (Wu *et al.*, 2009;

Shang *et al.*, 2010) have also been identified resulting in the suggestion that there are certain kinds of Abscisic acid signaling pathway in Arabidopsis. ABA may act as a signaling molecule at the transcriptional and also at the post transcriptional level. Recent integrated transcriptomic and metabolomic analyses have shown that under stress conditions in various plant families, some ABA regulated molecular mechanisms of metabolic networks are present (Chunli li *et al.*, 2011). Many metabolic pathways, such as carbohydrate glycolysis, photosynthesis and nitrogen metabolism are modified under stress conditions. As part of tolerance to abiotic stresses the expression of several genes involved in draught-induced metabolite accumulation, e.g. sugars (trehalose, sorbitol and mannitol), amino acids (asparagines and proline) and amines (polyamines and glycine betaine) are increased. Among these, amino acid accumulation is of a major importance in the plant protection but how during water deficit or after exogenous ABA application, the regulation of amino acid accumulation occurs; it is still a good question to solve.

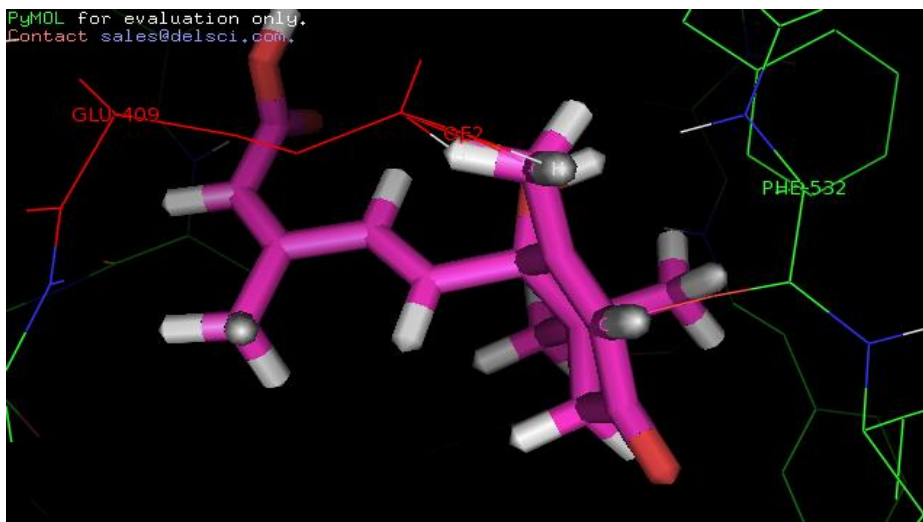
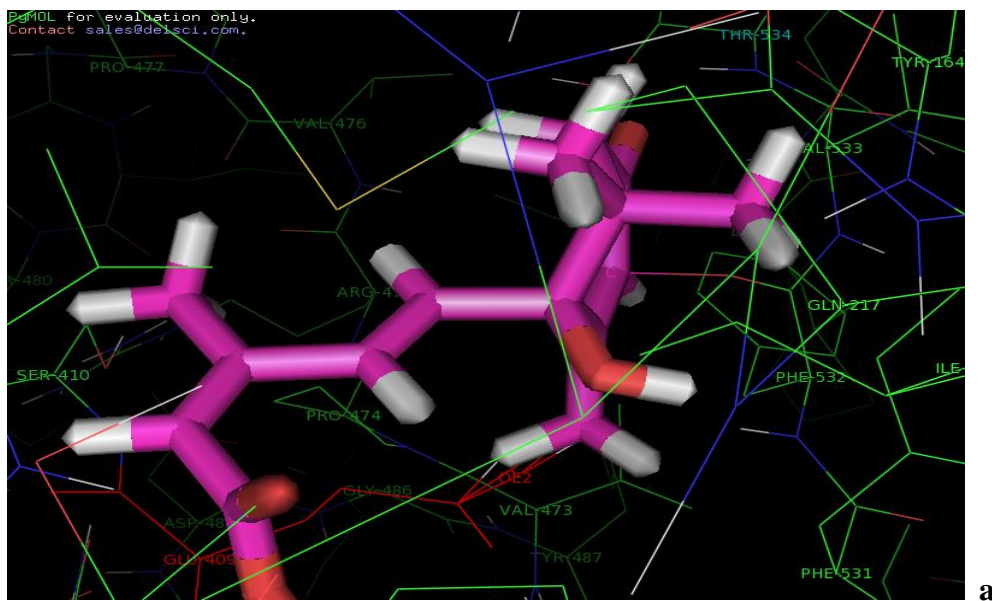


Fig 4.2. Dioxygenase- ABA interactions (Glu409 and Phe532, attached with ABA molecule)

The Abscisic acid is bound to Glu 409 and Phe532, which indicates that Glu and Phe would have essential functional and structural significance. We know that glutamate is

essential in the carbon and nitrogen metabolism in higher plants and plays a central role in amino acid metabolism (Young and Ajami 2000). Glutamate, which is (1) a product of the assimilation of ammonia through the reaction catalyzed by glutamine synthase which exists as cytosolic and chloroplastic isoforms and co-operate with glutamate synthase (2) the substrate and product for aminating and deaminating glutamate dehydrogenase (3) the preferred precursor for proline synthesis through pyrroline carboxylate synthetase (P5CS) activity (4) and both the acceptor and donor of many different reversible aminotransferase reactions, such as alanine aminotransferase and aspartate aminotransferase. The homeostatic control of glutamate concentration requires or involves combined actions of these different glutamate pathways within plants. Evidences indicate that glutamate is a key signaling molecule in higher plants. Glutamate is produced by ammonium assimilation pathway which is catalyzed by glutamate synthase/glutamate synthetase cycle. In the asparagines synthesis process glutamate is also regenerated as a by-product by asparagines synthetase (AS). The glutamate dehydrogenase (GDH) deamination activity catalyzes the degradation of glutamate. The oxidative glutamate deamination facilitates carbohydrate metabolism by playing a significant role in the regeneration of respiratory substrates as α -ketoglutarate, ammonium and NADH act as reducing power. Since, in plants during stress the glutamate pathway is the predominant route for proline biosynthesis, the replenishment of glutamate pool is essential for the production of protective metabolites (Young and Ajami 2000). The integrated transcriptomic and metabolomic analyses of *A. thaliana* suggested that during stress condition, stress induces the ABA accumulation that regulates global metabolic networks and changes in amino acid concentrations. The proline and asparagines accumulation is correlated with induced expression of P5CS and AS genes. As P5CS expression can be induced by exogenous ABA and this hormone does not affect AS expression, under water deficit changes in proline content could be controlled at

transcriptional level via an ABA-dependent pathway, but under such conditions the changes in asparagines content could also be controlled through an ABA-independent pathway at the transcriptional level. Although, glutamate metabolism genes does not require ABA for regulation of its expression other control mechanisms which involve interaction with elements involved in ABA regulation must exist (Raghavan *et al.*, 2006; Szabados *et al.*, 2009; Sun *et al.*, 2012). These control mechanism for the regulation of ABA encoding genes may also present in strawberry plants, such as the existence of regulatory processes that control metabolic content during stress and in the water deficit tolerance or fruit ripening, the putative central role played by glutamate and the ABA signaling could extend information on how changes occur in the nitrogen metabolism under adverse environmental conditions.



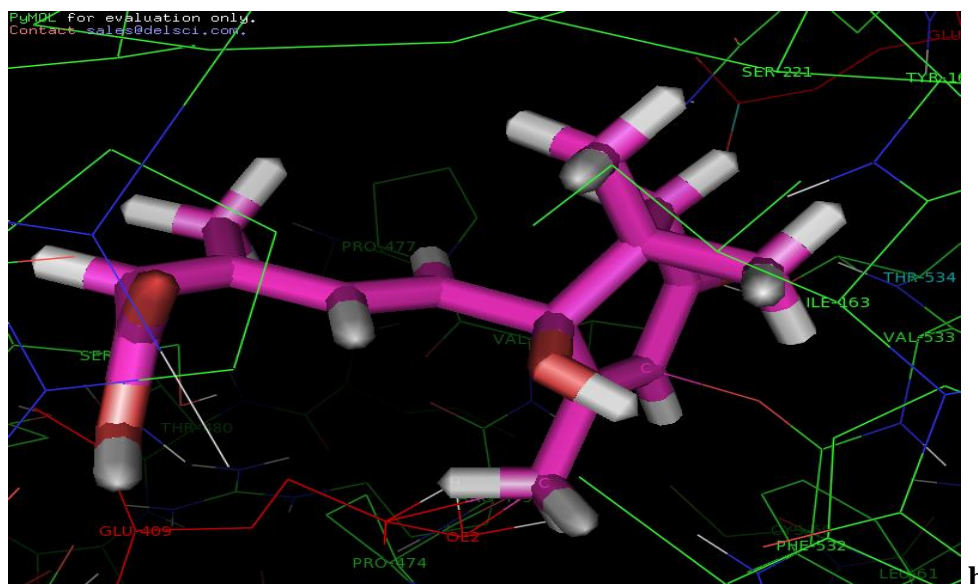


Fig. 4.3. Dioxygenase-ABA binding interaction

1.(Glu409 and Phe532; and other residues Phe531, Ile163, Gln217, Val533, Tyr164, Thr534, Val473, Val476,Pro474,Pro477, Ser410, Arg475, Gly486, Tyr487,Asp484)

2(Glu409 and Phe532; and other residues Pro474, Pro477, Ser410, Ser221, Tyr164, Cys60, Leu61, Val533, Thr534)

The intermolecular dioxygenases family enzymes which catalyze hydroxylation of aromatic amino acids include phenylalanine hydroxylase (PAH), tyrosine hydroxylase (TH) and tryptophan hydroxylase (TPH) (Schofield and Zhang 1999). Phenylalanine hydroxylase catalyzes the rate determining step in the phenylalanine catabolism i.e. hydroxylation of phenylalanine to tyrosine. Tyrosine hydroxylase catalyzes the initial step in the biosynthesis of catecholamine neurotransmitters i.e. the oxidation of tyrosine to 3, 4-dihydroxyphenylalanine (L-DOPA). Tryptophan hydroxylase catalyzes the formation of serotonin precursor 5-hydroxy tryptophan. These amino acid hydroxylases contain histidine and glutamine residues with an iron centre in the catalytic domain. The normal phenylalanine pathway, which include formation of tyrosine by PAH activity; some deleterious mutations in the gene encoding PAH could lead to phenylketonuria (PKU) causing severe progressive mental retardation, in which phenylalanine is accumulated in

the blood and serves as a substrate for tyrosine amine transfer (TAT) facilitating the neurotoxin phenylpyruvate, whereas TAT activity leads to the conversion of tyrosine 4-hydroxyphenylpyruvate (Abu-Omar *et al.*, 2004). The global alignment of bacterial and eukaryotic PAH, TH, TPH indicated that there is an absolute conservation of a few amino acid residues such as Ala, Trp, Cys, Phe, Gly, Pro, Arg, Glu, Tyr and Asp. Since, these amino acid residues have been highly conserved through prokaryotic and eukaryotic evolution, the importance of these residues cannot be ignored. Many of these residues are absolutely required for activity, indeed. For example, the first coordination sphere of PAH, contains iron and the residues histidine and glutamine which is required but phenylalanine can act as cofactor and substrate binding at the active site (Abu-Omar *et al.*, 2004). While in the type of dioxygenase currently studied, the residues His225, His273, His339 and His529 are present in the first coordination sphere of the iron, while Glu409 and Phe532 can act as a cofactor and substrate (ABA) binding at the active site (Messing *et al.*, 2010). This strict residue conservation clearly indicates that phenylalanine, tyrosine, glutamine and tryptophan hydroxylases and 4-hydroxyphenylpyruvate dioxygenases have some functional correlation (Zhang *et al.*, 2007). The PAH ternary complex molecular structure with substrate analogues e.g. L-norleucine and 3-(2-thienyl)-L-alanine show that in the full length enzyme the binding of L-Phe shifts the tetramer/dimer equilibrium toward the tetrameric form (Abu-Omar *et al.* 2004), which suggests that the structural changes are observed in the overall crystal structure and not just limited to the catalytic domain. The His binding at the active site changes the glutamine and iron atom coordination from monodentate (before addition of His) to bidentate. The most significant change has been observed in Tyr, which from a surface position moves to a buried position near the active site causing maximum displacement. Similarly, on addition of different substrate and amino acid residues other structural displacements in the structure are possible. The

structure indicated direct contact between substrate and amino acid residues (Glu404 and Phe532), and also some conformational changes associated with substrate binding. Moreover, using X-ray absorption spectroscopy substrate/cofactor induced structural changes or changes in the active site coordination sphere can be observed. Furthermore, molecular docking analysis combined with crystallographic results can provide insights into the molecular basis of dioxygenase inhibition.

4.1.4. Abscisic acid signaling:

The L-Phe to L-Tyr conversion by phenylalanine hydroxylase is regulated by cofactor inhibition, substrate incubation and specific Ser residues phosphorylation (Abu-Omar *et al.* 2004), which are phosphorylated by specific kinases such as, cAMP dependent kinases or calmodulin dependent protein kinases, which decreases K_{phe} at the regulatory site and results in catalytic enhancement. Phosphorylation causes an increase in total protein volume, which facilitates L-Phe532 binding at the regulatory domain. At the catalytic domain there are some additional regulatory sites. Cys60 located adjacent and Arg475, which is in the vicinity of the regulatory domain, could affect basal activity and L-Phe532 affinity. Ser410 and Ala224 are interacting by polar and non-polar interactions with residues at the regulatory domain. These residues are playing key roles in dioxygenase substrate or ligand binding and stabilizing the structure. In addition, mitogen activated protein (MAP) kinases phosphorylate Ser221. Other kinases, including Protein kinase C e.g. mitogen and stress activated kinases (MSK) and PRAK (p38 regulated/activated) kinases can also be involved in Ser410 phosphorylation. This phosphorylation by increasing the rate of inhibitor dissociation enhances enzymatic activity. Recent spectroscopic and crystallographic studies indicate that binding of both substrate and phenylalanine causes change in the iron geometry. This observation indicates that the active site can accommodate molecular oxygen provided under the catalytic condition the

order of substrate binding involves O₂ as the last substrate to bind (Abu-Omar *et al.*, 2004; Bugg *et al.*, 2003). Furthermore, this oxygen binding is reversible, which would support metal site coordination. However, away from metal at a hydrophobic site the binding of O₂ is also possible. The substrate hydroxylation requires iron, thus the metal is possibly involved in hydroxylating species, phosphorylation by different kind of kinases, seems to be an agreed upon pathway, thus far, is consistent with cleavage of oxygen-oxygen bonds taking place in a different step from substrate hydroxylation. The substrate (L-Phe532) dependent oxidase activity independent from metal is also possible. This enzyme has chemical mechanism which is distinctly different from other known mechanism of cytochrome P450 enzymes because of the catalytic active site residues and changes occurred in the catalytic coordination due to substrate binding residues and cysteine bridges. The substrate oxidation step is also distinct from that proposed for Cytochrome P450, the “oxygen rebound mechanism”. In this case, Phenylalanine substrate with para (p) substituents, such as methyl, chloride or bromide, (e.g. Abscisic acid) is catalytically converted to products in which the para (p) substituent migrated or shifted to the meta (m) positions. This is also known as hydrogen shift mechanism. To rationalize this observation, which argues against C-H abstraction, “aromatic epoxidation” has been suggested. Another possibility is the formation of a cationic intermediate that is consistent with 1,2-hydrogen shift. The considered interpretation is a hybridization change on the iron terminal, which is also consistent with cationic electrophilic substitution mechanisms (Bugg *et al.*, 2003).

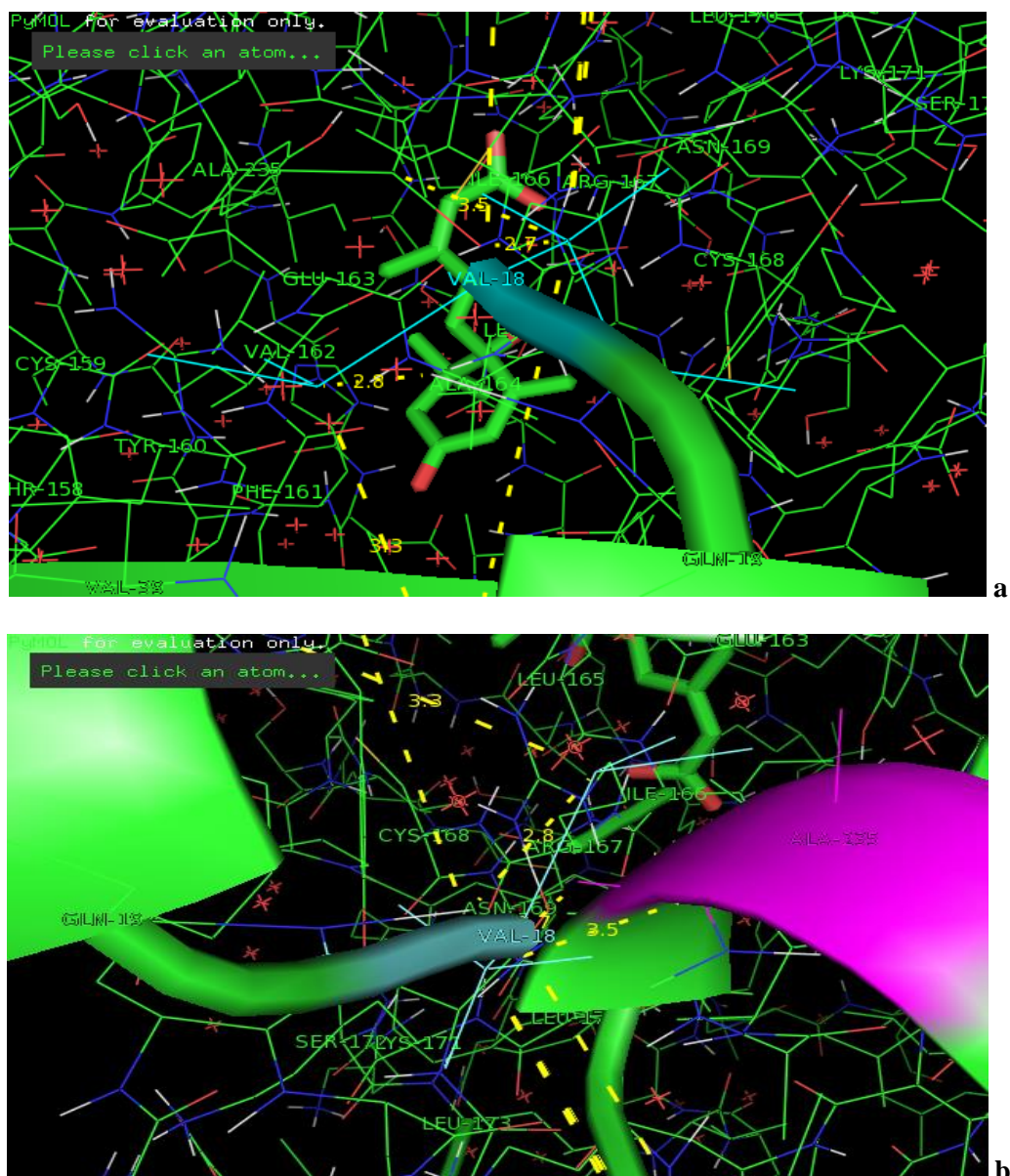


Fig. 4.4 Dioxygenase and ABA receptor (PYL9) tagged with ABA complex, and residues (Val18 of PYL9 interacting with Ala235, Tyr160, Thr156, Phe161, Arg167, Asn169, Cys159, Cys168, Ser172, Glu19, Phe21 of PYL9 and Leu165, Leu170, Glu163, Ile166, Val162 and Tyr164)

However, structural data provided subtle differences in active pocket residues (second coordination sphere or beyond) and their potential role in catalysis and substrate specificity. Of course, structural data provides the basis for hypotheses and speculations of structure-function relationships to be investigated experimentally via biochemical techniques. The structure model provides evidences to solve mechanistic issues. There are

many active site residues identified by crystallography or other techniques of importance to one extent to another, the study focuses on ones that have been characterized and provided insights on what role any residue could have in defining the enzymatic functions.

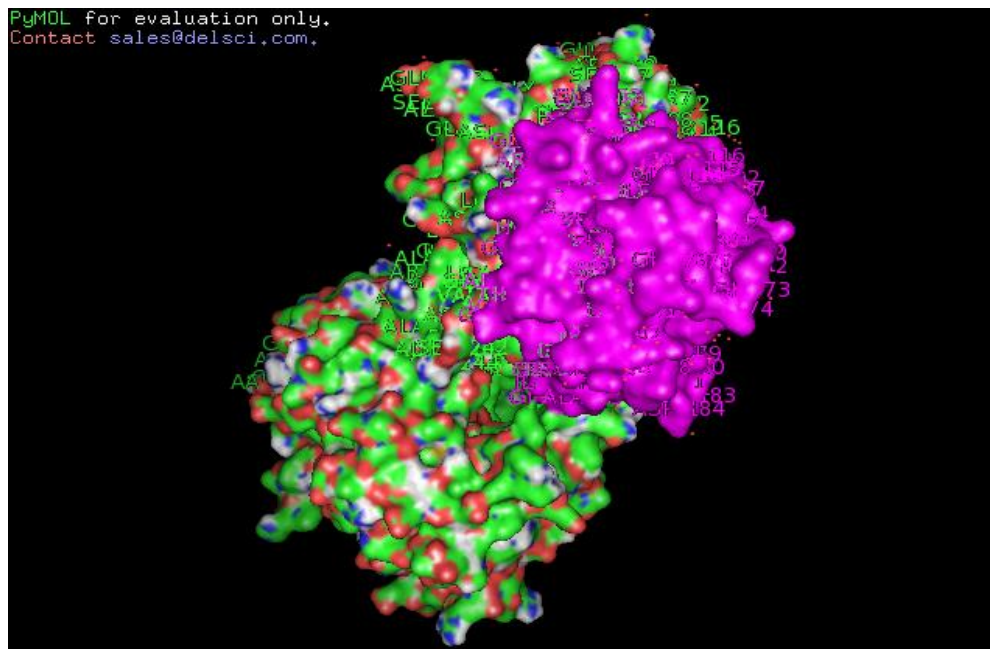


Fig.4.5 Dioxxygenase and ABA receptor (PYL9) tagged with ABA complex, surface view, showing groove, and surrounding residues, Val18 of PYL9 interacting with Ala235, Glu163 and Ser172)

An elegant study is required to understand the structural basis of certain enzymes which have evolved to discriminate between two very similar substrates. Of the 9-10 residues which have variable side chains in different dioxxygenases or hydroxylases, some most important residues are essential for tyrosine hydroxylation or phenylalanine carboxylation are identified such as His165, Tyr164, Arg167, Asn169 and Val476. Val533 could afford the kind of enzyme with an enhanced affinity for Phe532 hydroxylation. However, the involvement of a loop in the second coordination sphere of iron is a discriminating factor between the binding order of substrate and residues (60-248). Comparison of dioxxygenase structures and its binary or ternary adduct with various substrates or ligands, shows a

change in the conformation of this loop from closed to open, as Tyr164 loses its hydrogen bond to a water molecule (Messing et al., 2010, Zhang et al., 2007; Fritze *et al.*, 2004; Kiser *et al.*, 2009). The ternary structure superpositions of dioxygenases and different substrates indicated that this conformational change could be responsible for opening the binding site for phenylalanine, Phe532. This mechanism is another example of induced-fit, where binding of the first substrate, creates the binding site for the second substrate. The gaseous plant hormone, ethylene regulates many plant responses and development steps. In the presence of carbon dioxide, it is involved in fruit abscission, seed germination, tissue differentiation and root elongation; in this process a five-coordinate distorted square pyramidal geometry for iron is adapted only when both ACC and ascorbate are bound. Carbon dioxide also has a complimentary stabilizing role for the ACCO/Fe(II) /ACC distorted octahedral geometry. In the absence of carbon dioxide, because of the ACC alone 30% of the octahedral complex converts to the square pyramidal form primarily preparing the complex for dioxygen binding, which leads to deactivation of enzyme presumably via oxidative damage (Abu-Omar *et al.*, 2004). α -keto acid-dependent enzymes are one of the most important mononuclear non-heme iron hydroxylases. They have been reported to be involved in plethora of biological processes ranging from DNA and RNA repair, antibiotic synthesis, to even herbicides. The α -keto acid-dependent enzymes include 4-hydroxyphenylpyruvate (HPP) dioxygenase, taurine/2 oxoglutarate dioxygenase (tauD), deacetoxycephalosporin C synthase (DAOCS), asparagine hydroxylase, lysyl hydroxylase and prolyl-4-hydroxylase etc. (Abu-Omar *et al.*, 2004). The reaction mechanisms of the catalytic core of various protein structures can be explained by the iron intermediates involved in substrate binding or coupling with oxidized/non oxidized states of iron, before and after the substrate binding. For example, the alkylsulfatase AtsK, which is responsible for the conversion of alkyl sulfate esters to aldehydes and sulfate; the

molecular structure of alkylsulfatase AtsK was solved for Apo enzyme as well as in combination with α -ketoglutarate, α -ketoglutarate and iron, α -ketoglutarate, iron and alkylsulfate ester for catalytic turn over. The substrate binding effect on the coordination environment of iron and the correct positioning of the substrate as to acquire function of the oxidation state of 2Fe-2S was probed by High resolution Q-band pulse ^2H -ENDOR spectroscopy, nitric oxide (NO) was found to bind with the water ligand causing water retention and the positioning of naphthalene was largely dependent on the oxidation state of the centre (Abu-Omar *et al.*, 2004). This study suggests that the oxidation state of the 2Fe-2S centre is dependent on the substrate binding site. While benzoate dioxygenases, show that the substrate binding influences the rate of electron transport between Fe-S cluster and the catalytic iron. Substitution studies, of the equivalent bridging aspartate in anthranilate 1,2-dioxygenases with asparagines or glutamate and alanine consistently deny the role of aspartate in facile inter-site electron transfer, or substrate gating, suggesting a new mechanistic role for the substrate in oxygen activation (Abu-Omar *et al.*, 2004; Wolfe *et al.*, 2001). Under steady conditions dissociation of products is the rate determining step that occurs upon reduction of the iron due to electron transfer from the enzyme's reductase component involving the centre. The role of aspartate in substrate dihydroxylation and reduction of the iron through electron transfer could be responsible for the mechanistic importance of the "peroxide shunt", which is the use of H_2O_2 , as an oxidant replacing oxygen, which is an important characteristic of this class of enzymes (Abu-Omar *et al.*, 2004). Intradiol dioxygenases, like catecholate dioxygenases, in which the catechol substrates donate its protons from hydroxyl radicals to the hydroxide ligands and the axial tyrosine to facilitate the formation of a bidentate iron (III) catechol complex that retains the essential residues such as endogenous two histidine and tyrosine ligands at equatorial positions. The different trans influence of histidine and tyrosine plays a key role in enzyme

catalysis, being a thermodynamic parameter, not a kinetic effect. Here, this study has covered the role of enzymes affecting C-H oxidations, aromatic hydroxylations, allylic peroxidation, desaturation and C-C oxidative bond cleavage (Abu-Omar *et al.*, 2004; Bugg *et al.*, 2003).

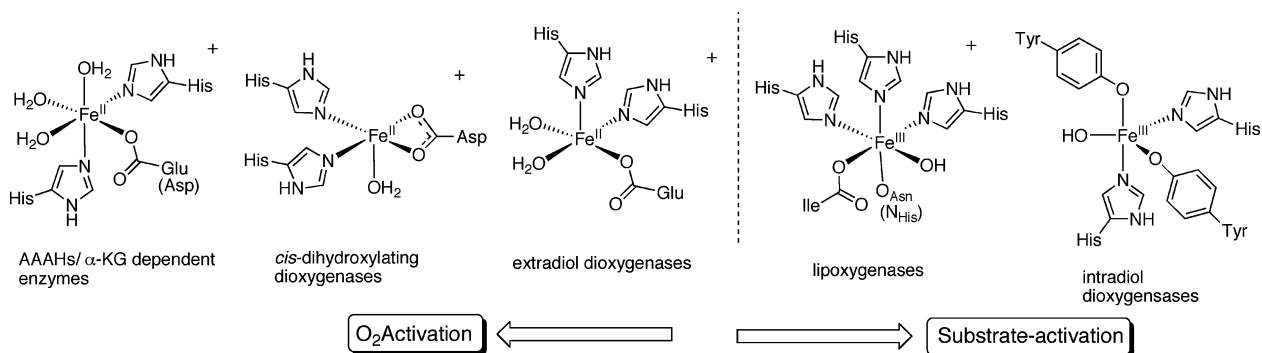


Fig. 4.6 Active sites of non-heme iron hydroxylases and ligands to favor O_2 versus substrate activation.

As the advancement in crystallography and cloning techniques combined with mechanistic investigations has progressed the knowledge of non-heme iron clusters has expanded tremendously. The suitable ligands for the enzyme's targeted functions are selected. Histidine carboxylate (for O_2 oxidation), histidine (in the case of lipoxygenases) and tyrosinate (in the case of intradiol cleaving dioxygenases) are used to favor substrate activation. The variation in ligands to favor the activation of O_2 versus activation of substrate has been shown in the figure 4.6. In addition to the ligand differences, more common differences are monodentate versus bidentate carboxylate coordination of Glu or Asp (i.e. hPAH vs cPAH and α -KG- dependent enzymes vs cis-dihydroxylating dioxygenases) (Abu-Omar *et al.* 2004). The study aims at understanding the enzymatic reaction well enough to be able to design an environmentally significant molecular catalyst or even inhibitor for the enzyme. Therefore, the mechanistic chemistry defines structure and functional roles of model systems and enzymes. Although, the molecular mechanisms,

for ripening of non-climacteric fruits such as papaya, berries have been extensively studied and many ABA signaling components have been identified, including hexose transporter or ASR proteins, various protein kinases such as receptor like kinases, SNF1 related kinases (SnRKs), G-proteins and phospholipases C/D, sugar inducible protein kinases VvSK1, calcium dependent protein kinases, calcineurin B-type protein kinases and mitogen activated protein kinases, protein phosphatases containing type 2C/A protein phosphatases (PP2C/A); and various class of transcription factors such as members of MYB/MYC, bZIP, B3 and sugar related WRKY families (Wang and Zhang 2008; Cutler *et al.*, 2010). ABRE binding factors (ABF) and AP2 transcription factors. A plasma membrane Abscisic acid receptor (Pandey *et al.*, 2009), a group of cytosolic Abscisic acid receptors, PYR/PYL/RCAR (Ma *et al.*, 2009; Park *et al.*, 2009) and a chloroplast, plastid Abscisic acid receptor, ABAR/ CHLH and magnesium chelatase H subunit (Wu *et al.*, 2009; Shang *et al.*, 2010) have also been identified. This provides strong evidence that a crucial role has been played by ABA in the regulation of ripening related gene-expression through ABA perception and strawberry fruit signaling transduction (Chunli Li *et al.*, 2011). There are some reported evidences for the activation of an aquaporin complex at the time of strawberry fruit ripening, which provides several lines of evidences that in the regulation of the ripening of non-climacteric fruits, such as strawberries, the interaction between sugar and ABA could be a core mechanism (Chunli Li *et al.*, 2011). Abscisic acid as a plant hormone, not only plays a significant role in adaptation to the adverse environment but also it is involved in regulating various aspects of plant growth and development. In accordance with that, it has been mentioned that exogenous ABA can significantly accelerate strawberry fruit ripening and due to the down regulation of the transcript of a 9-cisepoxycarotenoid dioxygenase gene (FaNCED1), a remarkable decrease in ABA content and significantly retard in the ripening process could be observed. FaNCED1 expression

is an important factor in the ABA accumulation in the ripening fruit, its functional characterization in strawberry has been intensively studied. Another study shows that down-regulation of the putative ABA receptor *FaPYR1* gene not only significantly delay strawberry fruit ripening, but also markedly alter ABA content (Chunli Li *et al.*, 2011). Collectively, these results demonstrate that in the strawberry fruit ripening process putative ABA receptor *FaPYR1* (Ma *et al.*, 2009; Park *et al.*, 2009) acts as a positive regulator. Subsequently, determination of the ability of *FaPYR1* and *FaCHLH/ABAR* proteins to bind to ABA will provide more information of the strawberry fruit ripening regulatory mechanism (Wu *et al.*, 2009; Shang *et al.*, 2010). In response to environmental and developmental signal a series of ABA signaling cascade in plant cells should be transmitted from the perception of hormone to gene expression and regulation (Jia *et al.*, 2011; Chai *et al.*, 2011). At the advent of strawberry fruit ripening the sugar accumulation, low acids and enlarged cells derived physiological changes, such as osmotic stress, sugar and pH, turgor, might act as early signals to promote ABA accumulation. The two putative ABA receptors, *FaABAR/CHLH* and *FaPYR1* receive ABA signal or may be strawberry homolog of G-protein coupled receptors (GTGs and GCR2 (the upstream of GPA1, named *FaGTG/GCR*) are also involved. These receptors could relay its signal respectively as follows, *FaABAR/CHLH* through *SigE* (a sugar and pigment stimulated gene) or sugar related WRKY transcription factor (Osanai *et al.*, 2009), *FaPYR1* through PP2C (*ABI1*, an ABA responsive gene) and *FaGTG/GCR*, presumably through GPA1, another ABA responsive modulator SnRKs are also participating. In the ABA signaling pathway, second messengers e.g. Inositides (IP_3 , IP_6), protein kinases, protein phosphatase2Cs and transcription factors form a signaling pathway integrated to the Abscisic acid signaling. Finally, the *FaABAR/CHLH* and *FaPYR1* signaling pathway could be involved in regulating fruit sugar and through ABF and *SigE* transcription factors in the pigment

metabolism, where as the FaGTG/GCR2 signaling pathway could be involved in the regulation of fruit softening through Ca-signals or ion channels (Chunli Li *et al.*, 2011; Gambetta *et al.*, 2010; Shang *et al.*, 2010).



Fig. 4.7 A model for strawberry fruit ripening via ABA signaling

The information, on how ABA perception is transmitted to their downstream events and its regulation with the complexity of cross-talks between ABA signaling pathway and other hormone signaling pathways (e.g. ethylene or IAA) and the signaling of different chemical inhibitors, which could have potential to affect the ABA signaling pathway at any point is the subject of the study.

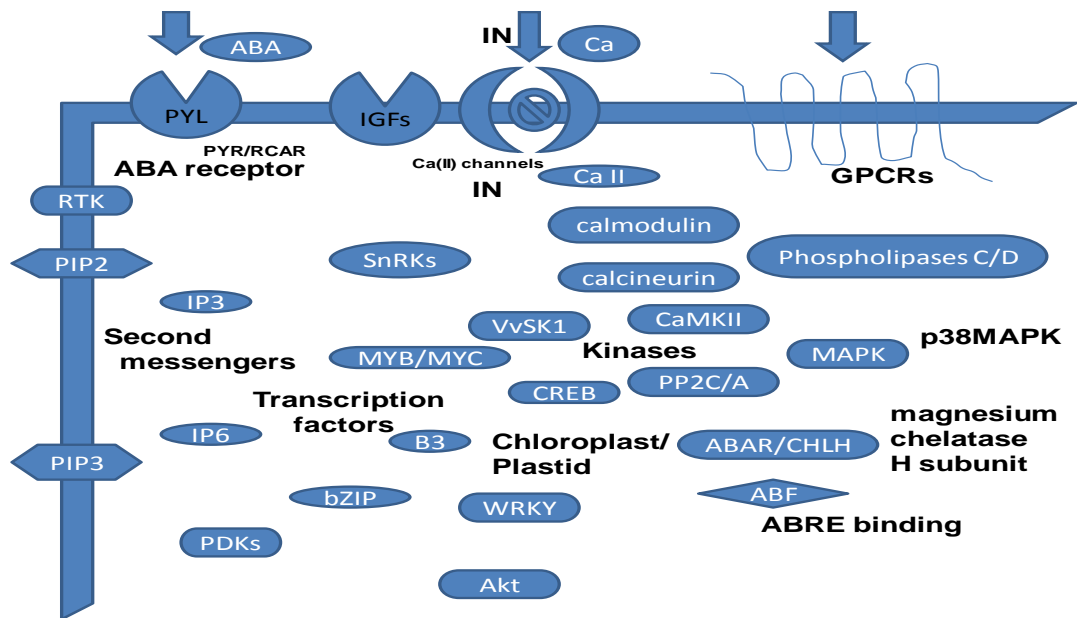


Fig. 4.8 G-proteins and phospholipases C/D, various protein kinases such as receptor like kinases, SNF1 related kinases (SnRKs), calcium dependent protein kinases, sugar inducible protein kinases (VvSK1), calcineurin, B-type protein kinases and mitogen activated protein kinases, protein phosphatases containing type 2C/A protein phosphatases (PP2C/A); and various class of transcription factors such as members of MYB/MYC, bZIP, B3 and sugar related WRKY families. Second messengers e.g. Inositides (IP₃, IP₆), ABRE binding factors (ABF) and AP2 transcription factors. ca-signals or ion channels. A plasma membrane Abscisic acid receptor, a class of cytosolic Abscisic acid receptors, PYR/PYL/RCAR and a chloroplast, plastid Abscisic receptor, ABAR/ CHLH and magnesium chelatase H subunit have also been identified.

Several lines of indirect evidences suggest that there are multiple types of receptors: different ABA responses have different stereospecificity of ABA analog activities and there seem to be both intracellular and extracellular sites of ABA action. An intracellular site of ABA action is evident from stomatal regulation. These studies indicate that the recognition of ABA receptor on the cytoplasmic side of the plasma membrane and a close interaction between the presumed ABA receptor and the Ca channel (Chunli Li *et al.*, 2011; Shang *et al.*, 2010).

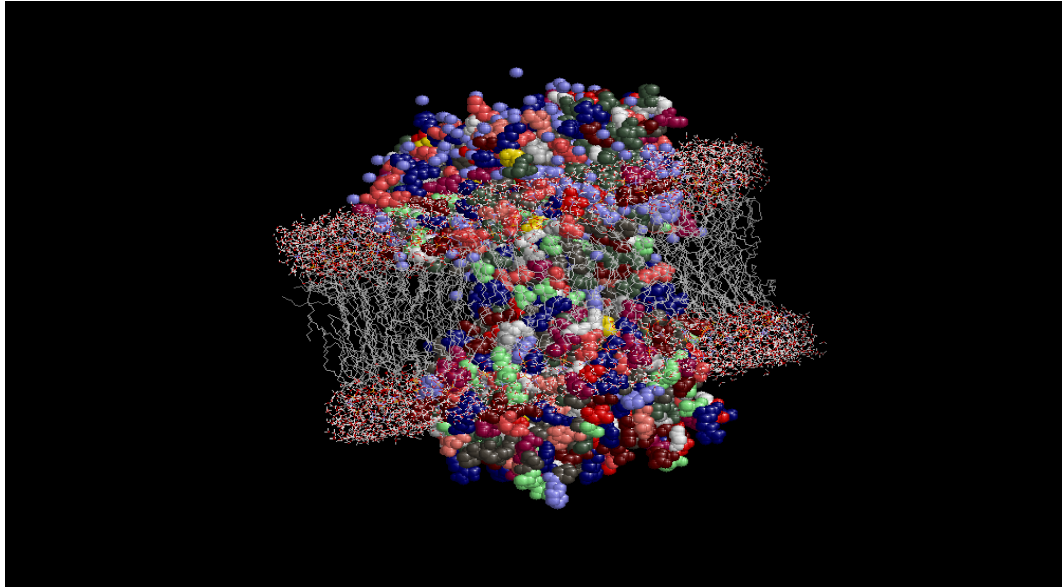


Fig. 4.9 A model for dioxygenase, ABA receptor (PYL9) and ABA binding based on ABA signaling mechanism (extracellular perception)

In two different studies using ABA protein conjugates that cannot enter the cell but are biologically active to induce gene expression and ion channel activity, the extracellular ABA perception was also observed. The structure model is consistent with the existence of both extracellular and intracellular ABA receptors. [However, other interpretations are also possible, for example, direct action of Abscisic acid on plasma and tonoplast membranes (or ion channels) from the cytoplasmic side, higher affinity of an ABA receptor for the protonated form or pH-dependent pathways] (Taylor *et al.* 2005).

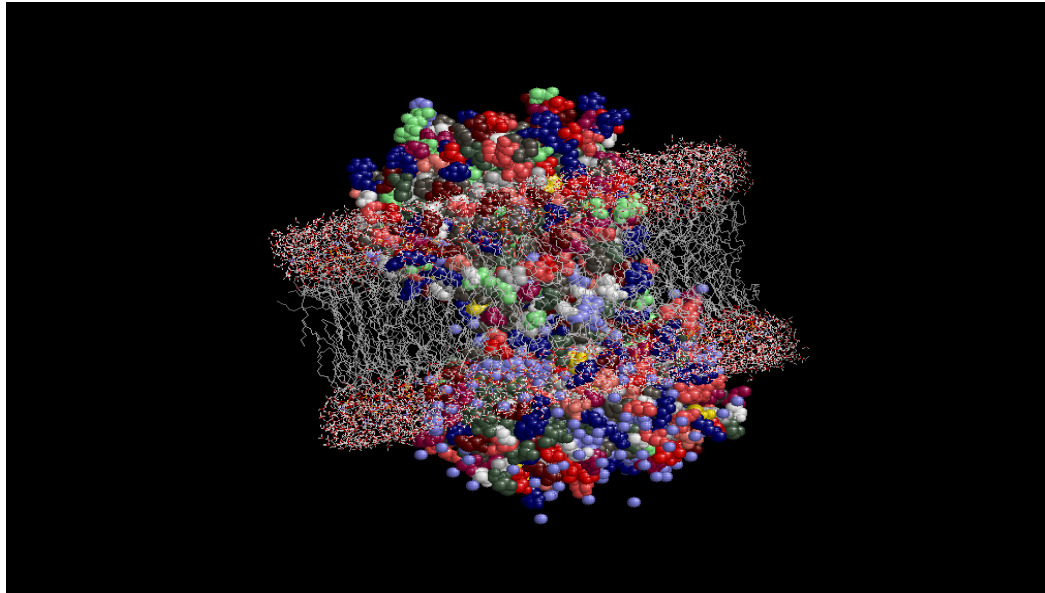


Fig. 4.10 A model for Dioxxygenase, ABA receptor (PYL9) and ABA binding based on ABA signaling mechanism (intracellular perception).

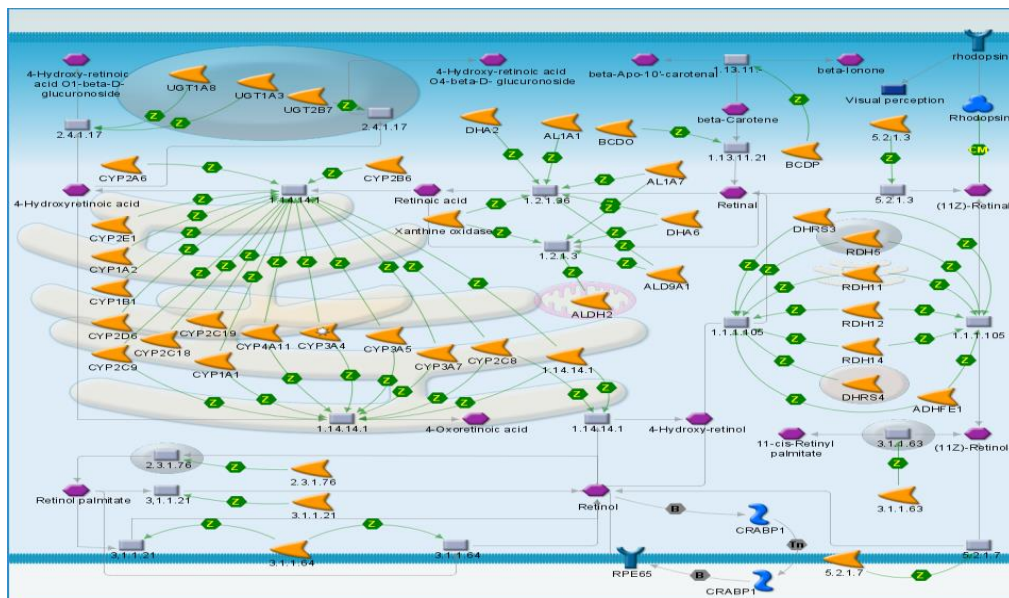


Fig. 4.11 In animals RPE65 (conserved domain in the strawberry sequence studied) and Retinol. Retinoic acid biosynthesis that is supposed to be similar to ABA in plants.

Although, proteins binding to ABA and ABA uptake mediated by carriers have been reported, but there is lacking evidence of linking these proteins to the physiological effects of ABA. ABA has direct effects on membrane fluidity and thermal behavior implicated in

ABA activity that it does not require receptor interaction. It is entirely plausible, indeed, that ABA could be analogous to lipophilic vitamins such as tocopherol (vitamin E) or vitamin K, which are the low-molecular weight compounds required in animals for fertility and blood clotting, respectively. On the other hand, there are similarities between ABA in plants and retinoic acid in animals. Both are synthesized ultimately from-carotene (also known as pro-vitamin A) by an evolutionarily conserved enzyme, dioxygenases by an oxidative cleavage catalyzed reaction. If perception mechanisms are conserved, ABA could bind to an intracellular receptor and would be acting as a transcription factor. In theory, the polyclonal antibodies could have epitopes (anti-idiotypic) mimicking the structure of ABA and therefore bind to proteins capable of ABA binding, including an ABA receptor. Homologs of (ABA) are membrane-localized proteins which are involved in the organization of actin, myosin, dyenin, microtubule cytoskeleton. The Abscisic acid is involved in the activation of endogenous Ca^{2+} dependent Cl^- current. The significance of ABA signaling in plants was demonstrated by the correlated disruption of ABA-regulated ion fluxes or (using inhibitors of vesicle trafficking). The ABA signaling response is indirectly involved in the regulatory functions of membrane vesicle trafficking and changes in cell surface and ion channel composition of membranes in the cell, direct binding and regulation of ion-channel activities (e.g. as found in neurons) and a receptor complex scaffolding. The ABA signaling network core comprises a subfamily of type 2C phosphatases (PP2Cs) and three snf1 related kinases e.g. snRK 2.2, 2.3 and 2.6. whose activities are actively controlled by ABA. In the absence of ABA, PP2Cs, inactivate SnRK2 kinases including ABI1, ABI2 and HAB1, which interact with SnRK2s and cause dephosphorylation of a serine residue in the kinase activation loop (S175 in SnRK2.6) which on phosphorylation facilitates kinase activity. The binding of ABA to the PYR/PYL/RCAR group of ABA receptors facilitates receptor binding to the catalytic site

of PP2Cs and inhibit their enzymatic activity. In turn, ABA-induced inhibition of PP2Cs leads to SnRK2 activation by autophosphorylation of activation loop thus allowing the SnRK2s to transmit ABA signal to downstream effectors. In plants, phosphorylation of SnRK2 activation loop may also involve upstream kinases which are yet to be identified. It seems that phosphatase in the inactivation of kinase as observed in the SnRK2.6-HAB1 complex could represent a widespread mechanism for phosphatase-kinase regulation. Alternately, ABA-bound receptor and kinase bind to a phosphatase respectively turn on and off this stress-responsive pathway. ABA is terpenoid which prevents germination, promotes seed maturation and during vegetative growth curtails excessive water loss. In the climate warming and water shortage conditions, modifying ABA synthesis or perception as possible solutions to sustain crop yield for food and fuel has attracted the interests of scientists. ABA is anchored deep inside the receptor and is enclosed by two substructures which on conformational changes take the forms of “gate” and “latch”. The “gate” form creates a new surface on the receptor which can tether the PP2C, leaving the SnRK2 to autophosphorylate (Ser175 in the activation loop of SnRK2.6) and subsequently phosphorylate downstream targets. In the absence of the hormone-receptor complex, the PP2C by dephosphorylating the SnRK2 turns off the signaling pathway. Although, ABI1, ABI2 and HAB1 interact interchangeably with SnRK2.2, SnRK2.3 and SnRK2.6. In the activation loop of SnRK2.6 the phosphate from the phospho-serine is removed and binding further suppresses any remaining kinase activity. Between this kinase and phosphatase the major contact surface runs along their catalytic sites. The contact surface between the α G helix of SnRK2.6 and a loop structure in HAB1 is adjacent to the HAB1 Trp*385-PYL (ABA-receptor protein) interaction loop. Thus, when ABA binds to the receptor, the receptor adopts a surface structure remarkably similar to that of the kinase. This is because, a second acidic motif, in addition to the kinase catalytic domain, called the ABA box in the

C terminus of SnRK2.2, SnRK2.3 and SnRK2.6 can also interact with HAB1 (but not with the receptors). But in the absence of ABA, receptor bound to ABA can dislodge the kinase from the phosphatase without fully dissociating these two proteins; the kinase will remain tethered by its ABA box to the phosphatase. This could also explain that by using increasing concentrations of the isolated kinase domain of SnRK2.6, the PYL2-HAB1 complex can be disrupted, whereas the ABA-bound receptors cannot dissociate SnRK2-PP2C, even if the ABA bound receptors have much stronger affinity for the phosphatases relative to the kinases (Leung *et al.*, 2012; Soon *et al.*, 2012). This could be a common mechanism in many biological processes involving kinase-phosphatase complexes for kinase-phosphatase co-regulation and co-evolution.

4.1.5. Signal Transduction:

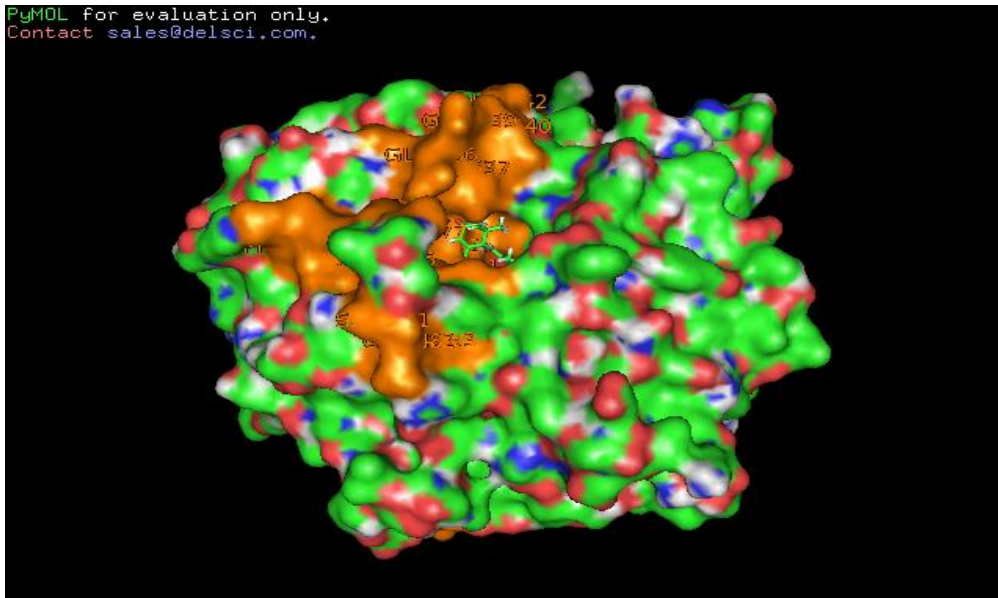
Biochemical and pharmacological studies indicated that in ABA signaling, G-protein, protein kinases, phospholipases and protein phosphatases participate in early events. The phospholipase C (PLC) activity produces Inositol triphosphate (IP₃), that acts as secondary messenger in ABA signaling (Chunli Li *et al.*, 2011; Shang *et al.*, 2010). Inositides which are more highly phosphorylated also can act as signals in animal cells, i.e. IP₆ seems to function in the inhibition of ABA. The phosphoinositide metabolism defects result in hypersensitivity to ABA and abiotic stresses further emphasize the role of phosphoinositides as secondary messengers. In the combinatorial control of gene expression, transcription factors are participating possibly by forming regulatory complexes and mediating specific ABA-inducible expression in various plants (monocot or dicot) or their human homologs are supposed to be involved in cell cycle control. Two separate signaling pathways mediate ABA induction and repression. Various steps in this signaling pathway involve G protein-mediated activation of PLD and as secondary messengers signaling by PA and Ca²⁺. Similarly, ABA could delay the ethylene-induced

programmed cell death which occurs at the end of the grain-filling period in endosperm development in cereals. Possibly interactions among some of the regulatory elements and hormonal and environmental signals regulating germination are involved in programmed cell death. The ABA is considered as a growth inhibitor that result from a combination of limited cell extensibility and inhibited cell division due to an arrest at the G1 phase of the cell cycle. Depending on the response, this G protein acts as either a promoter or an inhibitor of ABA response. The potential molecular mechanisms of ABA effects on cell cycle or extensibility include an inhibitor of ABA response that encodes a subunit of farnesyl transferase. There are so many specific targets of dependent farnesylation have been identified that include potential targets containing a C-terminal sequence conserved among yeast and animals. These include transcription factors, cell cycle regulators, cell wall modifiers, ROPs. As these proteins are farnesylated, it could affect their functions by regulating their membrane localization or potential for protein–protein interactions. The ABA signaling is a response to growth and environmental stress: regulatory effects of seed maturation/germination inhibition and stomatal aperture reduction, respectively. In contrast, most of the identified genes playing major roles in vegetative growth encode proteins which affect processes such as protein phosphorylation or farnesylation, phosphoinositide metabolism and RNA processing. These could also display combinatorial interactions among many family members of Rops, kinases, phosphatases and their potential targets and activators. However, events such as stomatal regulation and ABA-regulated gene expression define many similarities in terms of the relevant secondary messengers and a requirement for S-type anion channel activity (Zhang *et al.*, 2008; Daszkowska-Golec. *et al.*, 2013).

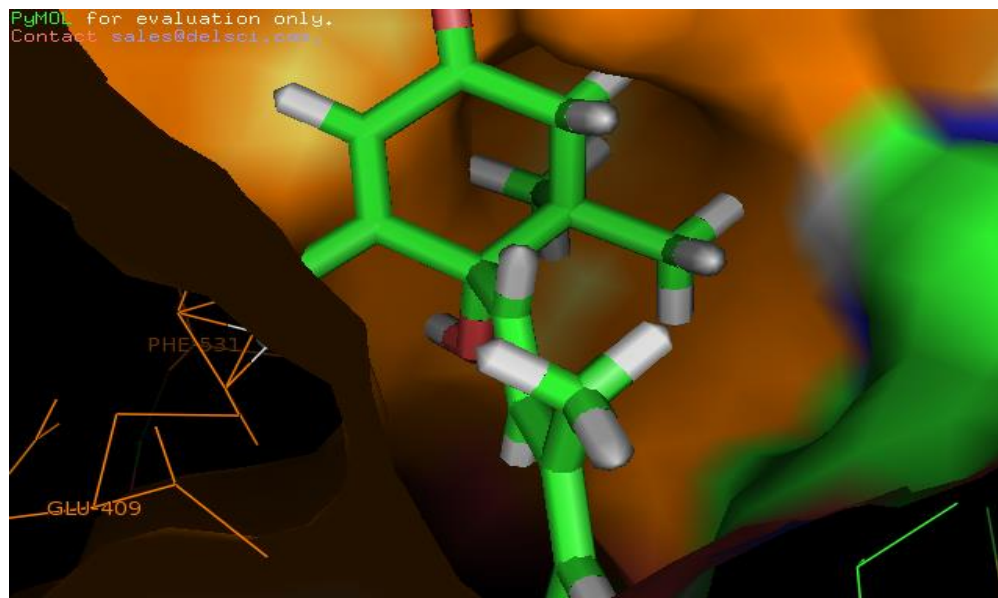
4.1.6. Inhibition induced by ABA response or selective herbicides:

Although, many evidences show that ABA plays an important role in strawberry fruit ripening, however, it is clear from ABA signal transduction that relies on various kinds of ABA receptors, G-protein like signaling. Ca-ion channels, second messengers like Inositides, and insulin like growth factor receptors; itself can act as promoter or inhibitor of ABA response, even ABA can act as a suicidal substrate. As we know, that ABA belongs to sesquiterpenoids ($C_{15}H_{20}O_4$), with one asymmetric and optically active carbon at C-1'. The form of ABA that occurs in nature is S-(+)-ABA, containing a side chain, with 2-cis and 4-trans conformations. Again, we know that trans conformation of ABA is biologically inactive, but the R-(-)-ABA, which is the possible product of racemization of ABA-trans-diol does have biological activity. Highly, saturated or unsaturated carbon chain Absicic acid such as allyl-ABA, farnesyl-ABA could act as suicidal substrates for dioxygenases, since for biological activity the enzyme containing, L-amino acids are acting more specifically, and natural form of ABA is the favored substrate for dioxygenases. Also, there are some inhibitors of ABA response, encoding a subunit of farnesyl transferases, which act in a targeted manner on a C-terminal sequence, this is a kind of dependent farnesylation and the potential targets which include transcription factors, cell cycle regulators, cell wall modifiers and ROPs (Zhang *et al.*, 2008; Sun *et al.*, 2012). Farnesylation of these proteins could affect their membrane localization and their interactions with other substrates or their potential for protein-protein interactions. So, these protein-protein interactions do not have natural small molecule partners always. Thus, targeting these protein-protein interfaces for potential therapeutic drugs or inhibitors and discovering small molecules as substrates or drugs, which could disrupt these protein-protein interfaces or could affect the stability of binding anyway, is challenging. The present study finds small chemical compounds or molecules which could act as drugs or herbicide inhibitors, with potencies to bind on

contact surface`s, so called “hot spots” involved in protein-protein interactions (Buckingham *et al.*, 2004). Remarkably, such small molecules bind much deeper in the contact surfaces of the dioxygenases with much higher efficiencies depending on their force fields. As we know, the electrostatic force fields could determine the binding stability of small molecule with contact surfaces of targated proteins. In protein-protein interfaces the contact surfaces involve amino acid residues, although, these contact surfaces are found to be large, a small subset of residues involved constitute most of the free-energy of binding; such small subset of residues are called hot spots, these hot spots constitutes much part of the contact surfaces and are usually located in the centre of contact surfaces. Proteins involved in protein-protein interactions can be promiscuous or could bind many targets using the same hot spot regions, structural studies show that these contact surfaces are adaptable, or it could engage variety of structurally diverse partners in a protein-protein pair. It has been shown that peptides or small molecule often show competitive binding to the hotspot with the natural protein-partner. As mentioned above, not only electrostatic interactions determining the stability of binding (Buckingham *et al.*, 2004), but also, for tight binding, there are many chemical solutions; also large contact surfaces are engaged in compact structures. Here, in this study, structure models of known crystal structures have been used, which are more commonly found in the protein data bank (PDB), which very nicely allow comparisons of the protein-protein complexes and protein and a small-molecule e.g. substrate complexes and are particularly instructive in many sense. This provides the opportunity to structurally analyze that how a small molecule or a peptide directly binds with the natural protein partners.



a



b

Fig. 4.12 Small molecules like Abscisic acid bind much deeper in the contact surfaces of the dioxygenases

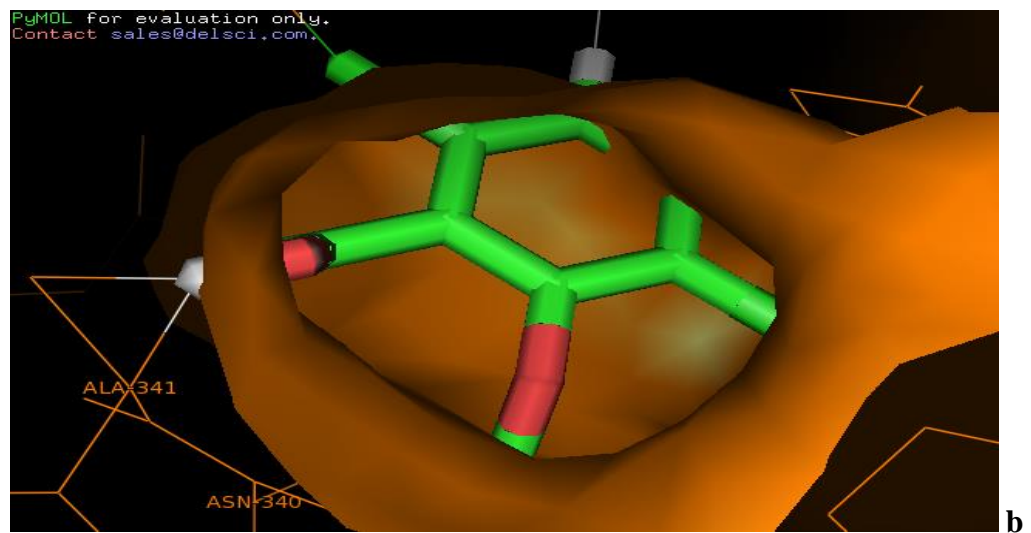
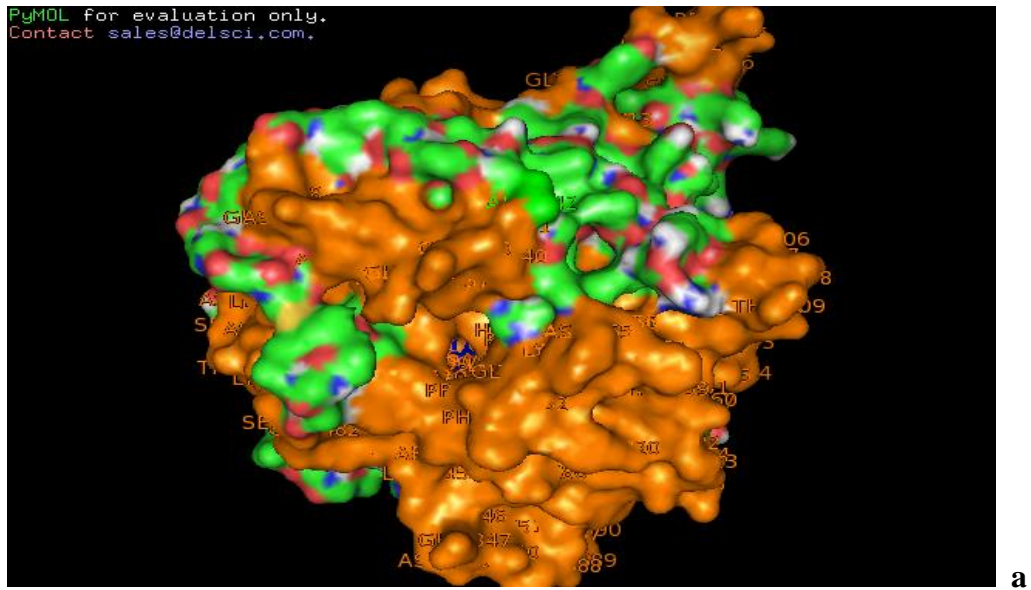


Fig. 4.13 Dioxxygenase-Dicamba (3,6-dichloro-2-methoxybenzoic acid), dicamba binding much deeper in the contact surfaces of dioxxygenases in comparison to ABA.

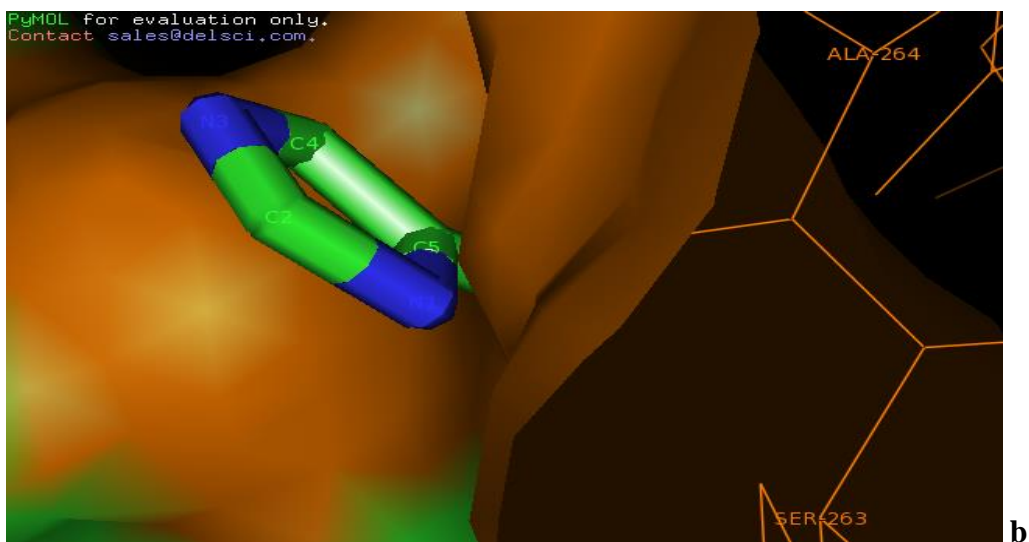
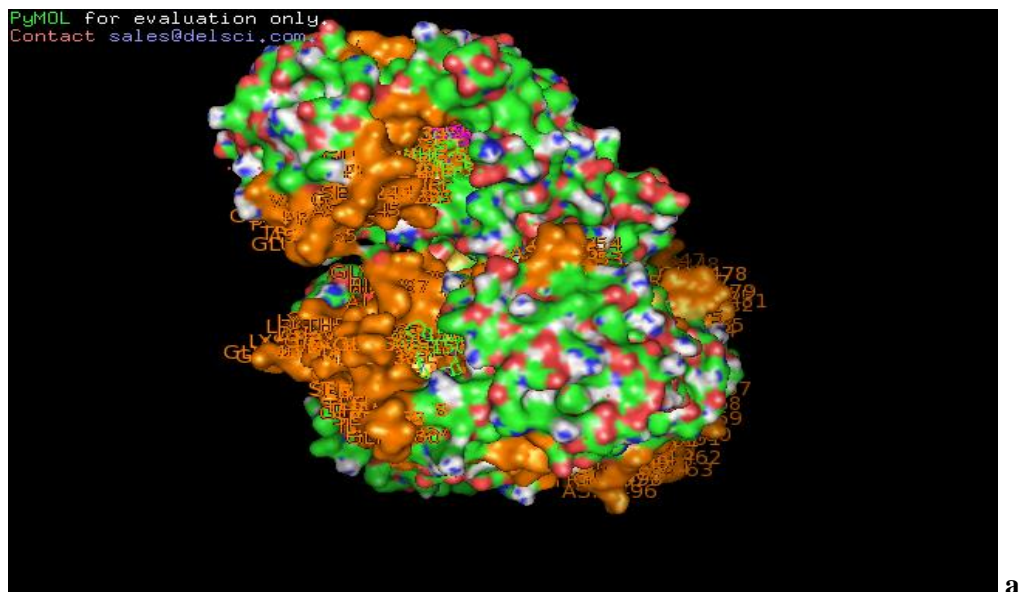


Fig. 4.14 Dioxygenase-Hydroxyphenyl-indoleamineimidazole, hydroxyphenyl (NHE, cyan) and indoleamineimidazole (PIM, magenta), contact surfaces, indoleamineimidazole binding deep in the contact surface, binding residues.

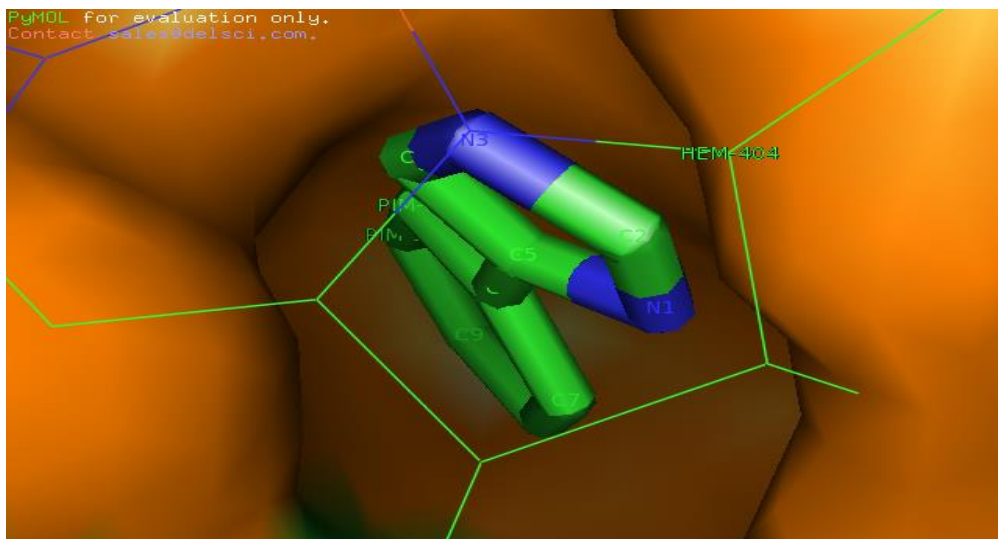


Fig. 4.15 Dioxygenases-binding with Hemoglobin I and indoleaimeimidazole

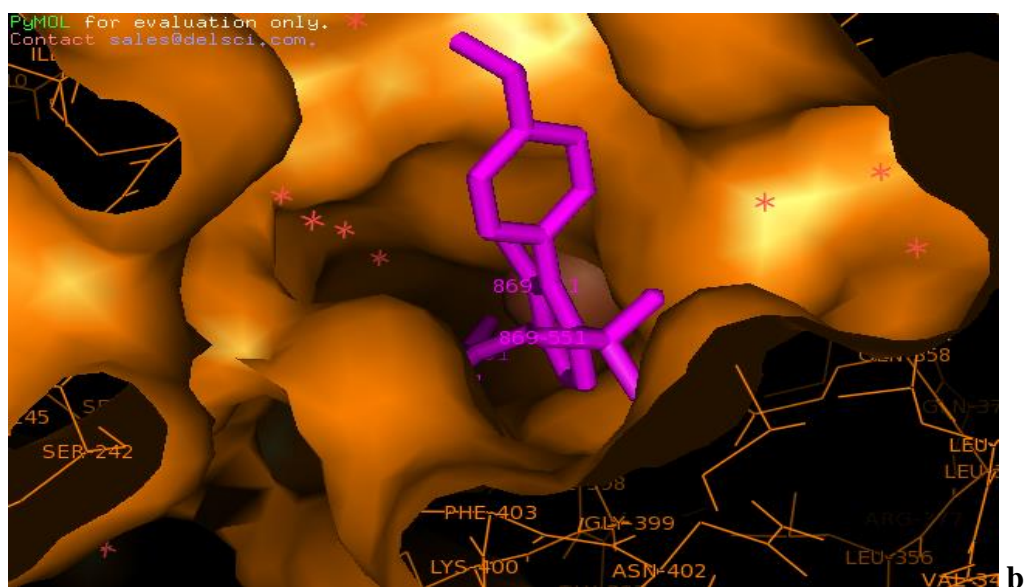
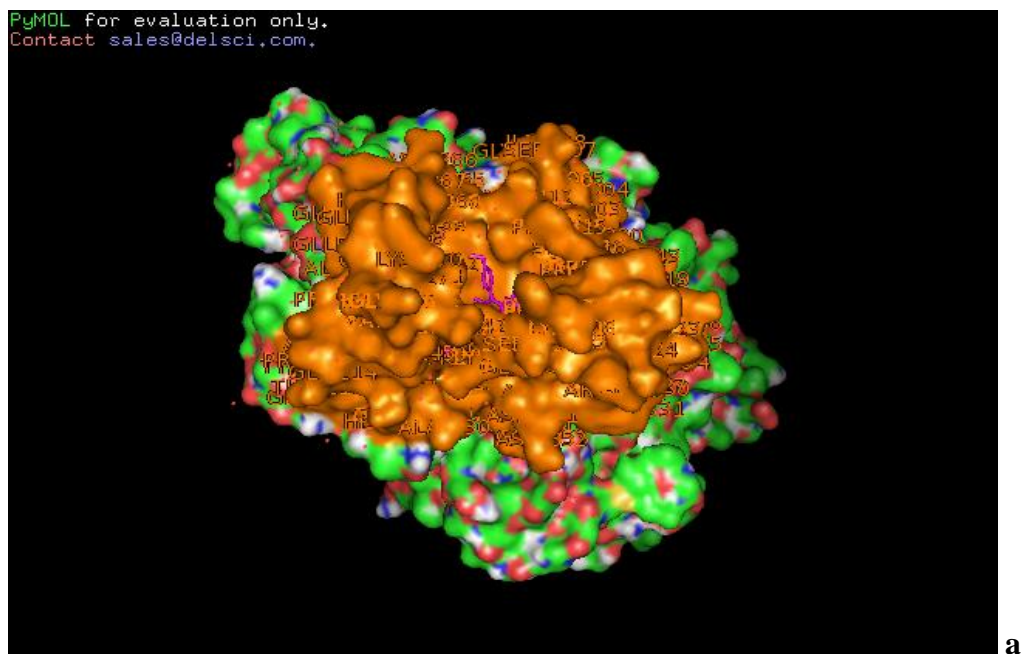


Fig.4.16 Dioxygenase-Hydroxyphenylpropionic acid contact surfaces, Hydroxyphenyl binding deep in the contact surface binding residues.

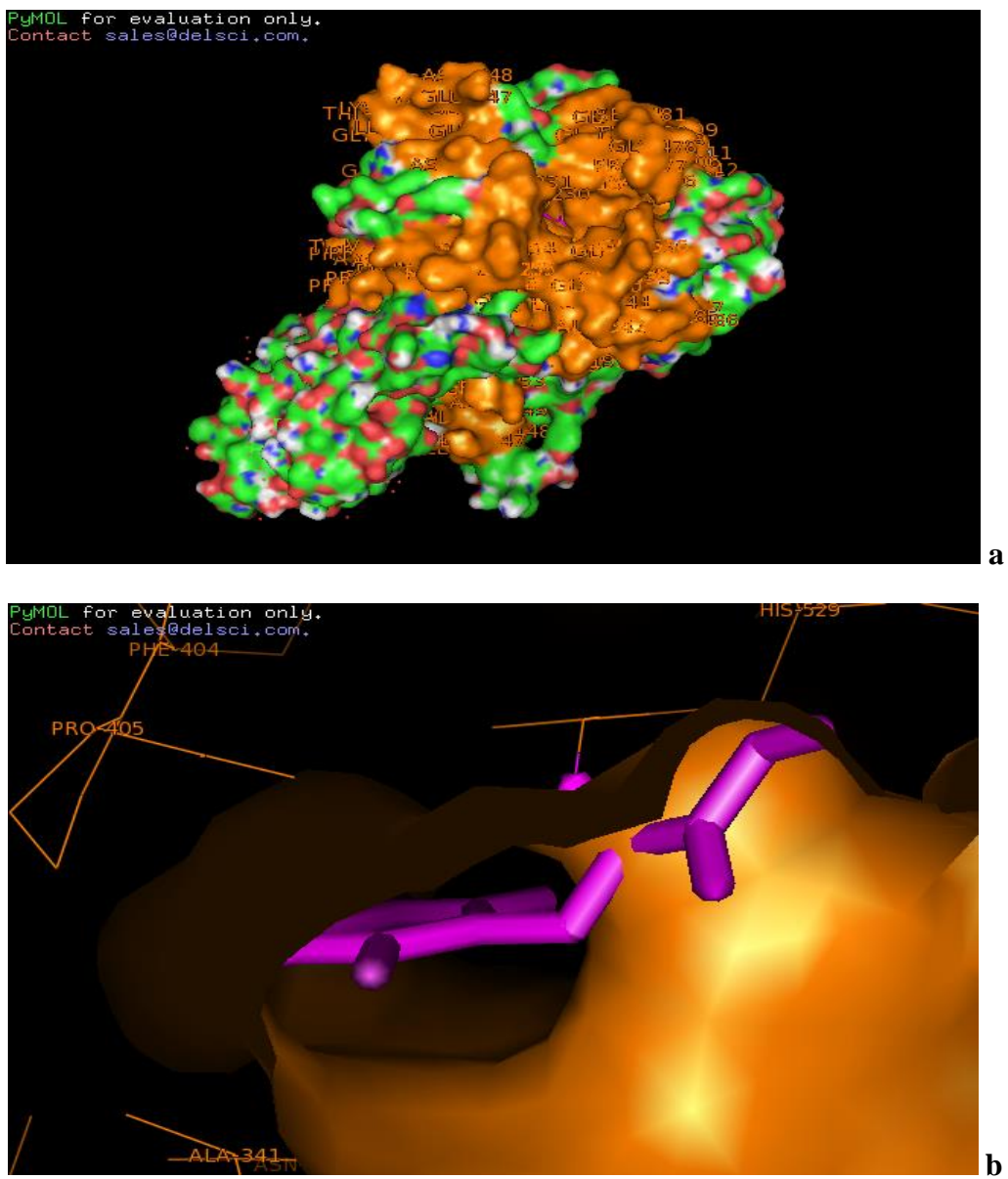


Fig. 4.17. Dioxxygenase-2, 4-D contact surfaces, 2, 4-D binding much deeper in the contact surface binding residues.

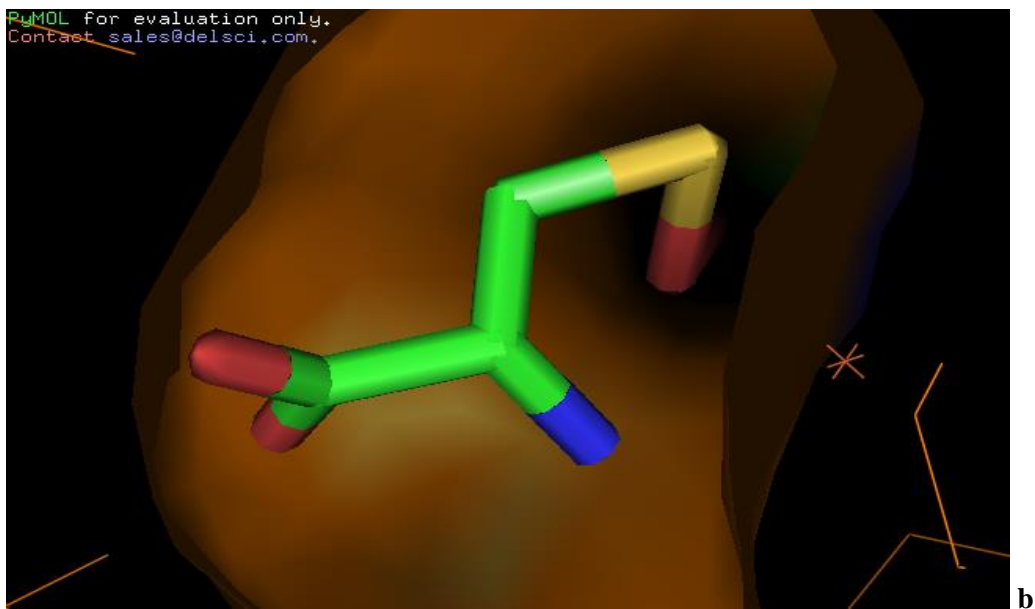
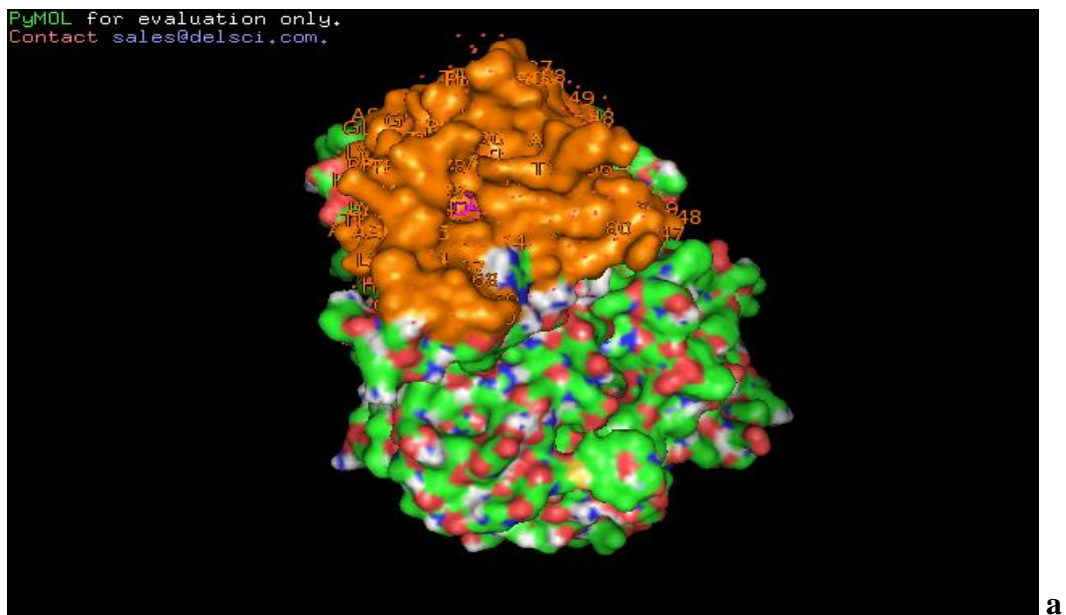
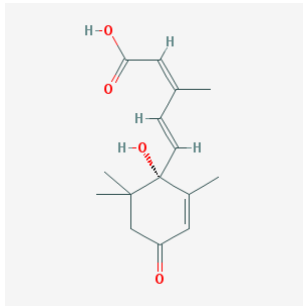
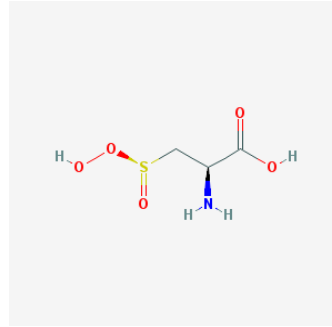


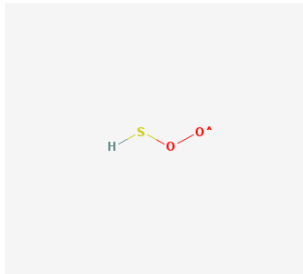
Fig. 4.18 Dioxygenase-Hydroperoxysulfenyl contact surfaces, Hydroperoxysulfenyl binding much deeper in the contact surface residues.



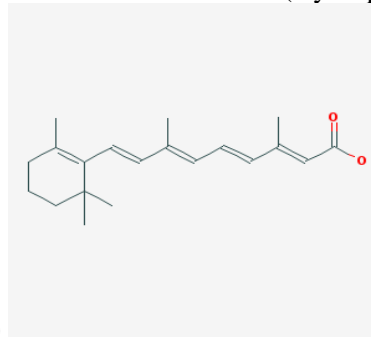
(Abscisic acid)



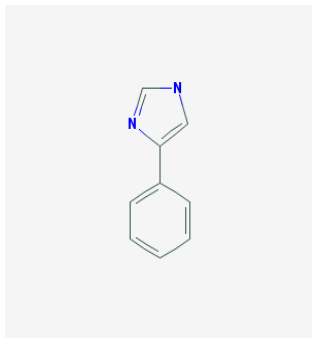
(Hydroperoxysulfenyl)



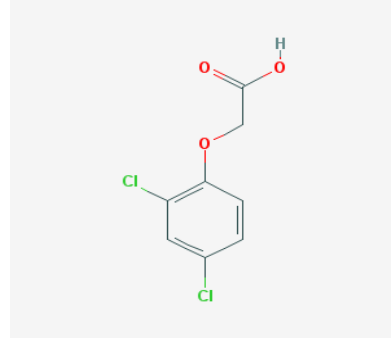
(Hydridisulfido dioxygen)



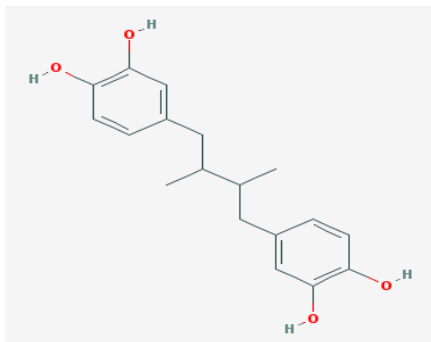
(Retinoic acid)



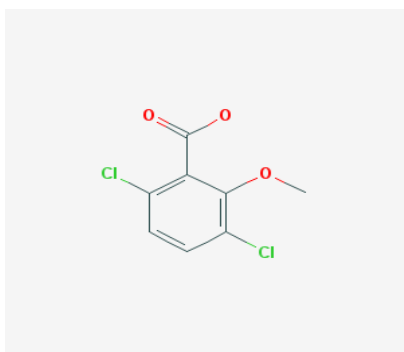
(Indoleamineimidazole)



(2,4-D, or 2,4-Dichlorophenoxyacetic acid)



(Nordihydroguaiaretic acid or NDGA)



(3,6-dichloro, 2-methoxybenzoic acid, or dicamba)

Fig.4.19 Potential herbicides

These are some small chemical compounds or molecules, which could inhibit or affect protein-small molecule complexes and protein-protein complexes. Some of these compounds i.e. 2,4-D and Dicamba are synthetic versions of the plant hormone Auxin (Hagood *et al.*, 2001; Heap *et al.*, 2002). These compounds have mode of actions similar to endogenous Auxin (IAA), acting on the cell wall and metabolism. This results in unusual or defective cell division and plant growth with an increase in ethylene production. Indoleamineimidazole is also such compound which could inhibit or disrupt functions of dioxygenases. Hydroperoxysufenyl compounds are also potential inhibitors of monooxygenases, as well as dioxygenases. Norhydroguaiaretic acid (NDGA) is a potential inhibitor of the expression of NCED (9-cisepoxycarotenoid gene) gene expression, affecting dioxygenases enzymatic function, by down regulation of its function. Besides this, herbicides are available in the market, which have various modes of action. However, weed resistance, application timings, constrained crop rotation, environmental and non-target plant safety, these factors limit the use of herbicide, which leads to further applications of new chemistries, to control weeds and recognize additional herbicides with different modes of action. In addition to this, genetically modified crops, with tolerance to broad spectrum of weeds, permit the use of new herbicides effectively. Starting from Auxin signaling and perception and a plant hormone interaction in signaling between ethylene, auxin and the up-regulation of 9-cisepoxycarotenoid dioxygenases (NCEDs) in Abscisic acid (ABA) biosynthesis leading to the accumulation of Abscisic acid are the initial step towards knowing the herbicide inhibitor action for dioxygenases. As we know, auxin herbicides belong to a chemical class, which includes benzoic acids, phenoxy-carboxylic acids, aromatic carboxymethyl derivatives, pyridine-carboxylic acids and quinolinecarboxylic acids. In higher plants, auxin herbicides mimic the action of the main auxin indole-3-acetic acid (IAA). However, they are long lasting, particularly due to

their higher stability in the plants, therefore, more effective than IAA. Auxin herbicides when present at low concentrations at the cellular sites of actions stimulate numbers of growth and development processes, but when the concentration of IAA and auxin activity in the tissue increases, growth is disturbed and plants are lethally damaged.

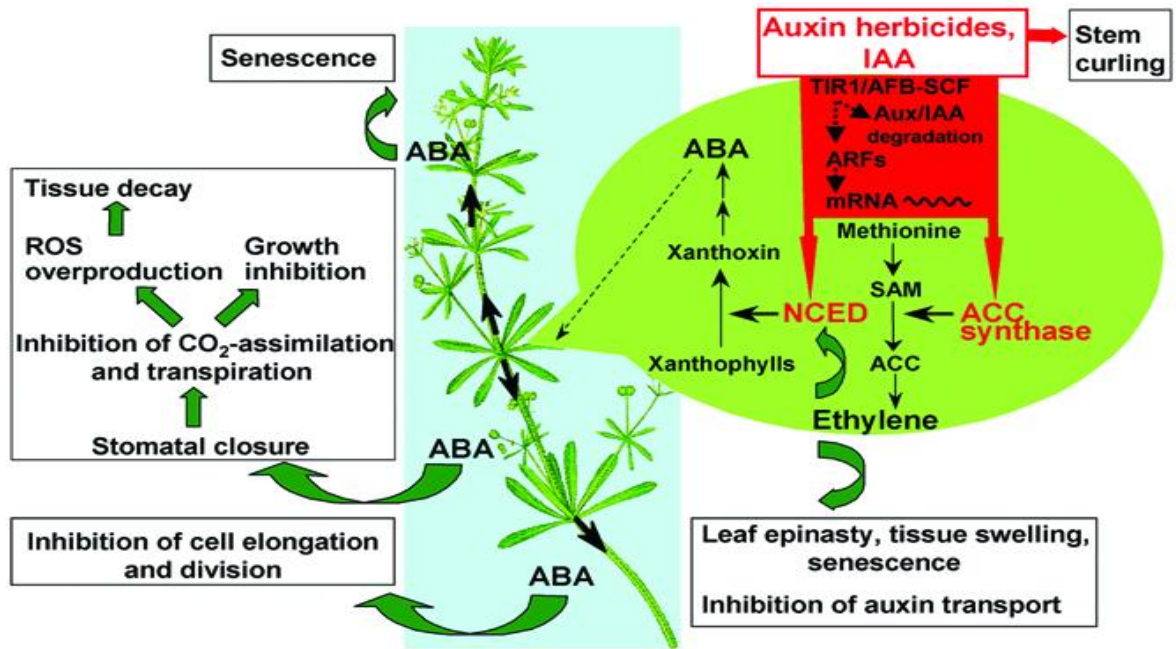


Fig. 4.20 ABA signaling and Auxin herbicides

The primary action of these compounds is to affect the cell wall architecture and nucleic acid metabolism. These compounds by stimulating the activity of a membrane bound ATPase proton pump acidify the cell wall. The reduction in apoplasmic pH induces an increase in the activity of an enzyme responsible for cell wall loosening thus causing cell elongation. Low concentration of auxin-mimicking herbicides stimulates RNA polymerase; result in subsequent increase in DNA, RNA and protein biosynthesis. Unusual increase in these processes presumably leads to uncontrolled cell division and growth, resulting in vascular tissue destruction. In contrast these herbicides in high concentration inhibit cell division and growth, usually in meristematic regions where photosynthate assimilates are accumulated and herbicide is transported through phloem. Herbicides which mimic auxin, stimulate evolution of ethylene which can produce characteristic epinastic symptoms

associated with the exposure of these herbicides in some cases (Grossmann *et al.*, 2003; Dharmasiri *et al.*, 2005; Kepinski *et al.*, 2005; Hagen *et al.*, 2002; Tan *et al.*, 2007).

We know that phthalamates and semicarbazone are chemical compounds that inhibit auxin transport. These compounds functions by inhibiting polar transport of naturally occurring auxin, indole-acetic acid and synthetic auxin-mimicking herbicides in sensitive plants. The auxin transport inhibition causes an unusual accumulation of IAA and synthetic auxin agonists in meristematic shoot and root regions, disrupting the delicate auxin balance needed for plant growth. After dicamba application, dicamba is translocated in the meristematic sinks, where at reduced dicamba rates, it, across a wider range of weed species delivers effective weed control (Gleason *et al.*, 2011). In sensitive grasses after the application of dicamba symptoms are characterized by stunted growth, rapid and severe plant hormonal effects (e.g. Epinasty). After rapid metabolism of dicamba in plants tolerance could occur (Gleason *et al.*, 2011).

Table 4.1: Selected crops in development tolerant to two or more herbicides (Glyphosate resistance threatens roundup hegemony, Emily Waltz, Nature Biotechnology, 2010)

Company (location)	Crop	Herbicides tolerated
Bayer CropScience (Monheim am Rhein, Germany)	Soybean	HPPD inhibitors, glufosinate, glyphosate
	Cotton	Glufosinate, glyphosate
	Corn	Phenoxy auxins (e.g., 2,4-D), aryloxyphenoxypropionate ACCase inhibitors (e.g., quizalofop-p-ethyl), glyphosate
Dow Agrosiences	Cotton, soybean	2,4-D, glyphosate
	Corn, cotton	Dicamba, glufosinate, glyphosate
Monsanto	Soybean	Dicamba, glyphosate
	Corn	Dicamba, glufosinate, glyphosate
Pioneer Hi-Bred (Johnston, Iowa)	Corn, soybean	ALS inhibitors, glyphosate
Syngenta (Basel)	Soybean	HPPD inhibitors (e.g., mesotrione), glufosinate, glyphosate

HPPD, hydroxyphenylpyruvate dioxygenase; 2, 4-D, 2, 4-dichlorophenoxyacetic acid; ACCase, acetyl coenzyme A carboxylase; ALS, acetolactate synthase

As this is known that 2,4-D (2,4-dichlorophenoxyacetic acid) and Dicamba (3,6-dichloro-2-methoxybenzoic acid) are the synthetic versions of plant hormone auxin (Hagood *et al.*, 2001; Heap *et al.*, 2002) and despite the wide use of dicamba, responses by plants to dicamba have not been extensively studied yet, thus to understand the dicamba mode of actions (Gleason *et al.*, 2011) plants sensitivity to auxinic herbicides 2,4-dichlorophenoxyacetic acid and dicamba can be compared and very recently some of the transgenic crops companies have proposed dicamba tolerant transgenic plants and dicamba as a potential p-hydroxyphenylpyruvate dioxygenases inhibitor as they are supposed to introduce dicamba tolerant transgenic soybean plants and dicamba as p-hydroxyphenylpyruvate dioxygenase inhibitor by 2014. After that farmers would be able to first grow dicamba tolerant crops (Bomgardner *et al.*, 2012). The dicamba mode of action is quite known for some time, different molecular studies with mutants insensitive to auxin explains about their sensitivity to dicamba and its mode of actions was compared with auxinic herbicide 2,4-D. Also, the whole genome analysis of dicamba induced gene expression shows that many stress responsive and signaling genes are stimulated by dicamba, including those involved in auxin biosynthesis and signaling, ethylene and abscisic acid (ABA) (Gleason *et al.*, 2011). The selectivity of auxin receptors and dicamba regulated gene expression and has provided molecular insights into dicamba regulated signaling that supported the development of herbicide resistance in crop plants (Gleason *et al.*, 2011). There are many recent patent files also, which describes the composition of the herbicide dicamba as a potential inhibitor of p-hydroxyphenylpyruvate dioxygenases. It has also been described that after dicamba exposure the plastid enzyme 9-cis-epoxycarotenoid dioxygenase genes were up-regulated. Abscisic aldehyde oxidase 3, the enzyme which

catalyzes the final step of Abscisic acid biosynthesis was also found to be up-regulated after dicamba treatment.

The dicamba treatment also induced the enzyme zeaxanthin epoxidase gene (ABA1) which catalyzes the first step in ABA biosynthesis. Dicamba specifically suppresses significant number of cell wall genes, suggesting that dicamba could be involved in inhibiting cell wall biosynthesis. Transport associated genes including, lipid transport protein as well as water transport protein and peroxidases were also specifically repressed by dicamba (Gleason *et al.*, 2011). Repression of peroxidases seems to be having significant implications in deciphering the dicamba mode of actions. Although, in the plant both 2,4-D and dicamba induces similar phytohormone responses. Overall, dicamba treatment helped *de novo* ABA biosynthesis in plants. Such auxinic herbicides functions through manipulating the plant hormone responses specifically causing large increase in ethylene and ABA contents. Phytohormones released in response to these herbicides can be manipulated which could provide additional means of auxinic herbicide resistance. Substrates for another group of dioxygenases such as Indoleamine 2, 3-dioxygenases, include indoleamines e.g. serotonin, melatonin and tryptamine. IDOs are more indiscriminate than catalytically related hepatic enzyme tryptophan 2, 3- dioxygenases (Zhang *et al.*, 2007). By locally depleting tryptophan, IDOs block the proliferation of T-lymphocytes, which are sensitive to Trp shortages. It indicates that Trp residues could play an important role in anti-inflammatory allergic response of dioxygenases (Hashimoto *et al.*, 2012). It is also the case in Glycines (Glyphosate) inhibitors which inhibit 5-enolpyruvateshikimate-3-phosphate (EPSP) synthase, which is involved in the synthesis of EPSP from shikimate-3-phosphate and phosphoenolpyruvate in shikimate acid pathway. Glyphosate action includes depletion of certain amino acid residues such as tyrosine,

phenylalanine and tryptophan, which are required for protein synthesis or for biosynthesis pathways leading to growth.

In higher plants the exogenous addition of these amino acid residues is not a successful approach to completely overcome the glyphosate toxicity; it indicates that it involves protein factors other than responsible for the inhibition of protein synthesis. Although, plant death can be due to other events occurring in response to EPSP synthase inhibition, which could be related to fatty acid metabolism and can be better explained by benzofuran like inhibitor actions, 2,4-D and Dicamba are such inhibitors which are being used as a glyphosate substitute for quite some time. The strawberry dioxygenase has allergen binding site, as strawberries also have allergic property such as anaphylactoid reactions to the consumption of strawberries have been reported. Strawberries also contain some chemical compounds, which can potentially fight metabolism associated diseases called metabolic syndrome. The strawberry dioxygenase also has cholesterol and lipid binding sites, therefore, it could be assumed that, at the same time, all these genes, proteins, which get modulated by these chemical compounds in strawberries, working simultaneously in different fronts against the components of the disease, including obesity, inflammation, diabetes, and cardiovascular diseases. The study explains the role of each compound, substrate, inhibitors, binding amino acid residue on the molecular mechanisms (Thornalley *et al.*, 2012).

This suggests that IDOs are potential target, highlighting the efficacy of chemotherapy or therapeutic vaccination by concomitant administration of an IDO inhibitor. Until recently, the most used IDO inhibitors have displayed affinities in the micromolar range and comprised mainly tryptophan derivatives such as 1-methyltryptophan (Zhang *et al.*, 2007). Other indole based IDO inhibitors includes brassinin and methylthiohydantoin-tryptophan. In the study, phenylimidazole (PIM) was used as a potential IDO inhibitor. With an

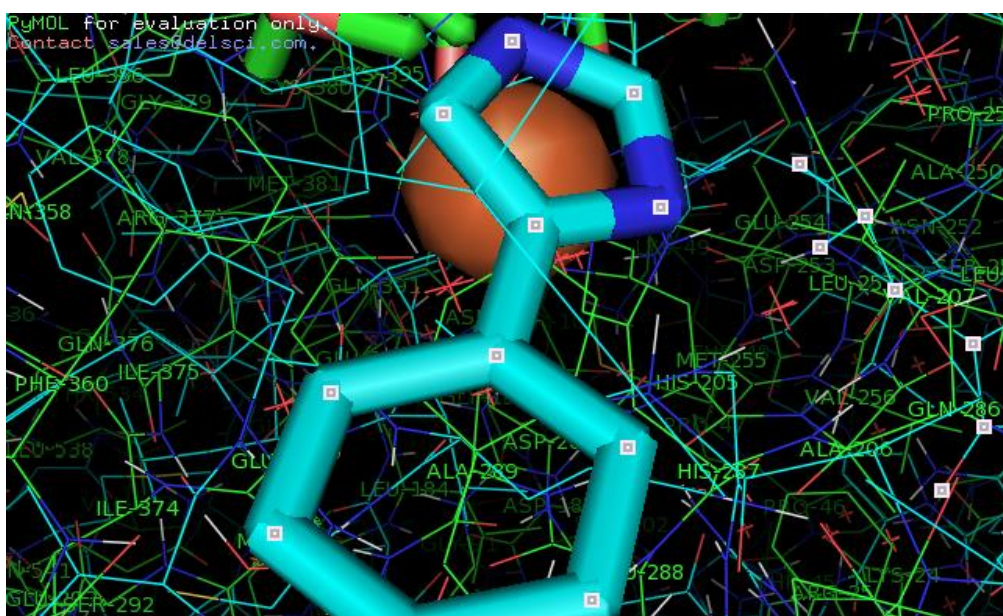
uncompetitive inhibition mode, PIM incorporates into the active site of IDO. Moreover, the preferred binding of PIM into the active site of IDO is confirmed by other crystal structure analysis. The essential features of PIM binding into the IDO binding cleft are particularly that (1) the oxygen atom of ketone functions, is positioned $\sim 2\text{\AA}$ above the plane of the heme coordinates to the heme iron (2) the indole ring is stabilized in the lipophilic A pocket of IDOs. The 3-pyridyl group is projected towards the active site entrance and is stabilized in the aromatic pocket B interacting with Phe and Tyr residues (Zhang *et al.*, 2007). According to the binding orientation, the interior side of the active sites (pocket A) contains amino acid residues and their interactions can be probed by introducing selected substituents on the indole ring, the lipophilic character of this pocket include chlorine, fluorine, bromine, methyl etc. Interactions in other pocket can be probed by substitution of 3-pyridyl group with other aromatic rings compounds containing hetero atoms. The heme iron binding interactions can also be explored by substitution of the ethanone moiety with other indole ring and the aromatic side chain linkers (Zhang *et al.*, 2007). Indole ring compounds can inhibit not only IDOs but also other kind of dioxygenases. Indoleamineimidazole compounds are also a good inhibitor of p-Hydroxyphenylpyruvate dioxygenases (HPPD) that converts p-hydroxymethyl pyruvate to homogentisate (Fritze *et al.*, 2004; Cheng *et al.*, 2004). The plastoquinone biosynthesis involves a reaction and the inhibition of this key step gives rise to bleaching symptoms of new growth. An indirect inhibition of carotenoid synthesis gives rise to symptoms resulting from plastoquinone the cofactor of phytoene desaturase. A group of HPPD inhibitors such as amitrole, inhibits chlorophyll and carotenoid accumulation in the light and precursors of carotenoid synthesis including phytoene, phytofluene, carotene and lycopene accumulate in amitrole -treated plants, suggesting that phytoene desaturase, imidazoleglycerol phosphate dehydratase, nitrate reductase or catalase, lycopene cyclase may be inhibited. However, other research

reports show that histidine, carotenoid and chlorophyll biosynthetic pathways probably are not the primary sites of amitrole action; instead, amitrole could have greater effects on cell division and elongation than on pigment biosynthesis. Above mentioned dicamba is also such kind of inhibitors, which has large effects on cell wall synthesis. Understanding proper dicamba mode of actions (Gleason *et al.*, 2011) could show that, how dioxygenase functioning is similar to shikimate acid biosynthesis enzymes. Despite economically significant applications in both pharmaceutical and agriculture sectors, detailed study is needed to identify the most catalytically relevant substrates binding model for dioxygenases. The role of the conserved glutamine, asparagines, tyrosine phenylalanine, tryptophan, valine, isoleucine, leucine, alanine and serine residues particularly interacting with substrates in the active site is highlighted in this study, which could be confirmed by quantum mechanical/ molecular mechanical calculations of the enzyme substrate complex and key reaction intermediates.

4.2. Important role of conserved residues and Inhibitor interactions:

The result highlights the following:

4.2.1. Indoleamineimidazole binding:



a

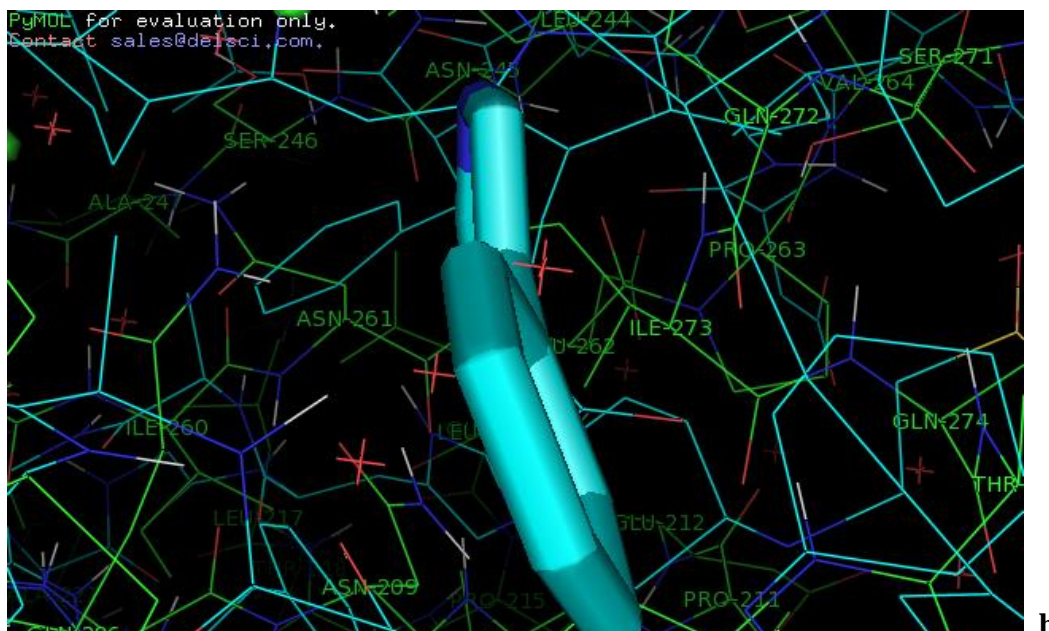


Fig.4.21. Dioxygenase-indoleamineimidazole binding residues (Glu119, Gln376, Gln286, and Gln358, Asn261and Ser246)

- (1) Central role of Glu119, Gln376, Gln286 and Gln358 in Hydroxyphenylpyruvate dioxygenase binding with indoleamineimidazole and the first nucleophilic attack.
- (2) The important movement of aromatic ring during the reaction,
- (3) Key role played by Asn261and Ser246 in C1 hydroxylation and the final ortho-rearrangement steps.

Furthermore, this study also shows that the last step of the catalytic reaction, the 1, 2-shift of the acetate side chain which is believed to be unique to the HPPD activity is also catalyzed by a structurally unrelated enzyme (Bugg *et al.*, 2003). HPPD (Hydroxyphenylpyruvate Dioxygenase) is a Fe (II)-dependent non-heme oxygenase catalyzing the conversion of Hydroxyphenylpyruvate (HPP) to Homogentisate (HGA). HPPD catalyzes the second step of tyrosine catabolism in most aerobic life forms, In most photosynthetic organisms; HPPD plays an anabolic role, as HGA is essential for the formation of isoprenoid redox molecules such as plastoquinone and tocochromanols. Plant HPPD is the molecular target of many natural compounds and of a range of very effective

herbicides that are currently used commercially. Interestingly, HPPD inhibitor/herbicide molecules also act as therapeutic agents for the debilitating and lethal inborn defects associated with type 1 tyrosinemia. Inhibition of HPPD prevents the accumulation of toxic metabolites in this disease. HPPD inhibitors are also currently being used in treatments of Parkinson disease, based on assumption that the inhibition of tyrosine catabolism will increase tyrosine availability for conversion to 3, 4-dihydroxyphenylalanine in the brain (Cheng *et al.*, 2004).

4.2.2. Substrate binding and reaction mechanism of HPPD:

Here, on the basis of enzyme-substrate complex models, and binding interactions which are hypothetical, a reaction-mechanism for HPPD has been proposed, which is supposed to be complex. It first involves a nucleophilic attack of the α -keto acid side chain by activated dioxygen leading to its decarboxylation by heterolytic cleavage. C1 hydroxylation of the aromatic ring then occurs via electrophilic attack by an Fe (IV)-oxene (Bugg *et al.*, 2003; Abu Omar *et al.*, 2004). An 1,2 shift of the acetate side chain then leads to the formation of HGA. It has been shown that the substrate binding mechanism of HPPD involves an ordered addition of substrates and an ordered release of products. HPP is the first substrate to bind and CO₂ is the first product to dissociate. Most of the previous biochemical studies have stressed the importance of 4-hydroxyl group. Particularly, this group involves in supplying electrons required for hydroxylation at position C1 on the aromatic ring (Bugg *et al.*, 2003). Some of the biochemical studies have also mentioned the role of an arene oxide intermediate, which forms during the aromatic C1 hydroxylation prior to the ortho-migration, which remain controversial, as this arene oxide intermediate could also be a side product released by mutant forms of enzyme and not actually directly involved in the main reaction mechanism. All of these structure models seem to agree on α -keto-acid moiety binding through bidentate

coordination with the metal ion and hydrogen bonding with a conserved glutamine residue, but to explain the binding of 4-hydroxyl group of the substrate, there could be various approaches (Bugg *et al.*, 2003; Cheng *et al.*, 2004). One of the potential explanations could be the conserved Gln272, Asn209 and Ser246, residues interacting with the 4-hydroxyl group. Inhibitors containing 4-hydroxyl group are taken into consideration for this reason, that 4-hydroxyl group forms hydrogen bonds with possibly Gln272 and Gln286 or with Asn261 and Ser246. Furthermore, Asn209, Asn261, Ser246 and Gln286 are involved in a hydrogen- bonding network stabilizing these residues in catalytic positions. To understand better, how these conserved residues affect substrate binding and the reaction mechanism, mutational studies of these conserved residues could be done, which could lead to further deep insights into the functioning of such dioxygenases. Informations on target sites of agrochemicals or herbicides now allow rational target site based herbicide design and approaches to improve both selectivity and potency towards herbicide and targeted or non-targeted organisms/plants. HPPDs are non-heme Fe²⁺-dependent dioxygenases that are mechanically related to α -keto-glutarate dependent dioxygenases, such as proline hydroxylase etc. Many potent HPPD inhibitors show that they exhibit the characteristic of slow, tight binding inhibitors. The majority of HPPD inhibitors effectively inhibit both plant and mammalian HPPDs. The present study describes the structural basis for potent inhibition of the plant enzyme by herbicidal benzopyrazoles and the possible modes of selectivity for the plant and mammalian enzyme for herbicide design (Cheng *et al.*, 2004).

The structural model for dioxygenases preceding the C-terminal helix has similar long and flexible loop (Val378-Gly397) with a disulfide bond between residues Cys380 and Cys395, as in *Arabidopsis thaliana* HPPD (Cheng *et al.*, 2004). Based on sequence alignment comparing the currently available plant HPPD sequences and many sequences from other

dioxygenases, it is found that these residues seem to be unique to plant HPPD. The section Ala230 to Gly243 may form two strands of β -sheet that loop over the top of the binding pocket and are a little closer to the C-terminal helix, thus indicating that this binding site is more accommodating. The active site is located within an open winding barrel like β -sheet. The metal ion is coordinated by His225, His273, His339 and His529 forming an octahedral complex, as other members of α -keto glutarate-dependent dioxygenase family (Messing *et al.*, 2010). The co-crystallization structures of enzyme-inhibitor complexes are not the focus of the present study, as found in structural models of enzyme-inhibitor complex; the dioxygenase structure can potentially define topography of the binding pocket of the enzyme to a greater extent. Gross differences in the main chain atoms between the ligand-bound or non-bound forms cannot be properly described by superimposed structure, for this, molecular replacements and refinements are being done at certain resolutions. In the enzyme-inhibitor complex, in the active site, the amino acid residues coordinating to the metal ion remain the same but the water molecules is displaced by 1, 3-diketone moiety of the inhibitor (Pallett *et al.*, 2001; Lee *et al.*, 1998; Meazza *et al.*, 2002). The distance of the metal ion and the oxygen molecule can be refined to better resolution, maintaining the octahedral geometry and provide a binding force and strong ligand orientation. All potent HPPD inhibitors in the structural classes possess an acidic 1, 3-diketone moiety and for tight binding the coordination of both oxygen molecule and the metal ion in the active site is essential (Cheng *et al.*, 2004; Pallett *et al.*, 2001; Lee *et al.*, 1998; Meazza *et al.*, 2002). In the inhibitor binding site the metal coordination involves the side chains of several residues, i.e. the phenyl groups of Phe360 and Phe403, are forming a π -stacking interaction with the benzoyl moiety of the inhibitor (869) as shown in figure 4.22. A T-shaped structure for Phe-Phe (Phe360 and Phe403) is no more stable than the preferred parallel displaced orientation. The majority of the aromatic-aromatic interactions are due to T-

shaped configurations. Moreover, clustering minimizes the effect of π -stacking interactions. The focus of the present study remains, however, to understand the role of π -stacking interactions in determining the tertiary structural change on binding with inhibitor compounds and its possible role in the structure-based drug designing.

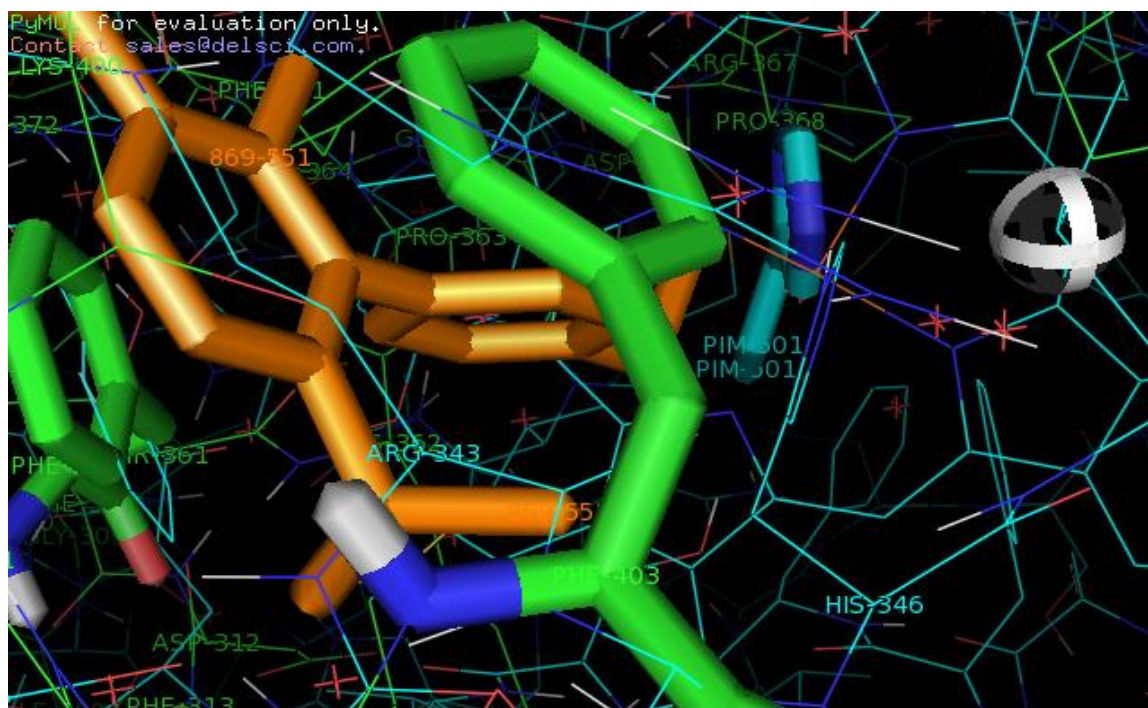


Fig. 4.22 Phe403 and Phe360 are forming π -stacking interaction with the benzoyl moiety of the inhibitor (869), shown in orange, in Dioxxygenase-Hydroxyphenyl-indoleamineimidazole inhibitor complex.

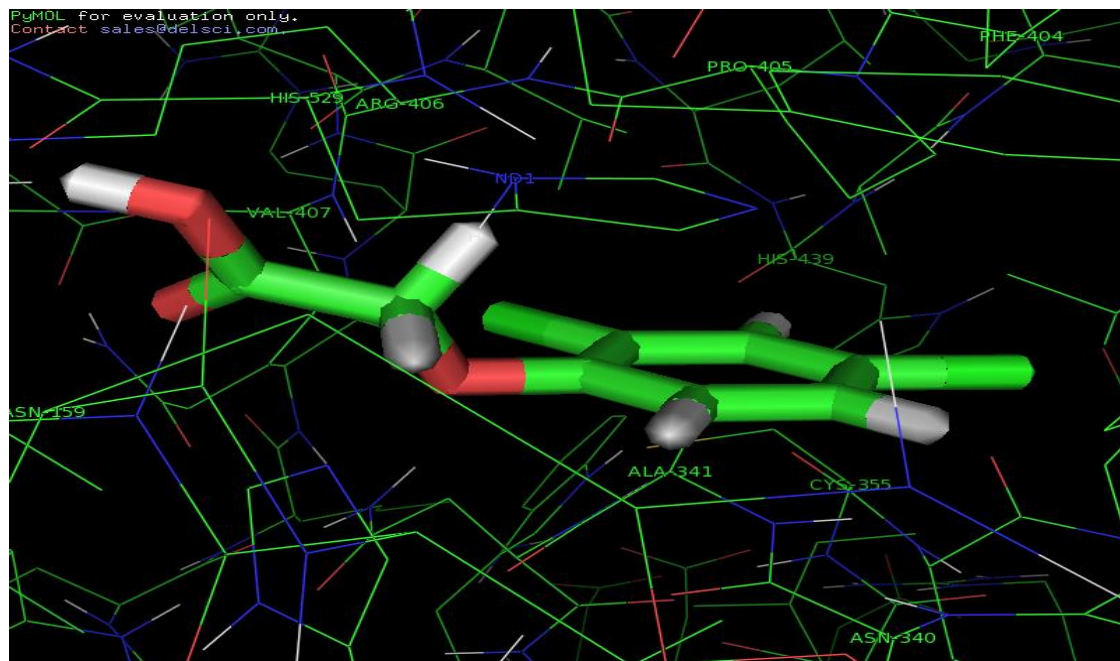
Pro259 causes a shift in the position of uncomplexed enzyme after binding with the imidazole inhibitor. Other shift in the complex is possible, that is Phe403 can rotate away from the inhibitor after substitution on pyrazole ring such as 3- phenyl-pyrazole. A significant difference is found among plant and animal HPPDs, that Phe403 is distally placed from the active site in the plant enzyme compared to the mammalian enzyme, allowing, larger inhibitor molecules to bind at the plant active sites (Cheng *et al.*, 2004). The structure indicates that preceding the C-terminal, the model sequence has Gly-Gly pairs before and after Cys395, which indicates that, it can act like a hinge region and that

the C-terminal helix seems to be slightly hinged (Cheng *et al.*, 2004). These shifting and some flexible residues in the complex and helix rearrangements are considerably, to accommodate the inhibitor compound or substitutions on the inhibitor compounds. The binding of inhibitors containing butylpyrazole groups (869) shift Pro259 ring atoms more distally, but in other cases, for example, in animal enzymes, like in rats the equivalent Pro259, cannot be readily repositioned, leaving the region, slightly constrained and the ligand binding site smaller (Cheng *et al.*, 2004). This also indicates that steric barriers prevent the shifting of same amino acid residue Pro259 in rats and the other possibility is that the shifting of Pro259 would cause the displacement of inhibitor towards the centre of the ligand binding site, instead. This could be a possible explanation, why some ligands bind much deeper into the active site, whereas, some ligands bind superficially. The size, molecular structure and the substitution pattern of the inhibitor are the key factors in generating their selectivity and these steric barriers can dramatically increase or decrease their binding affinities. It is anticipated that selected plant dioxygenase inhibitors and the subsequent insights into the structural basis of plant selectivity will offer unique tools for designing agrochemicals or herbicides with improved toxicological profiles.

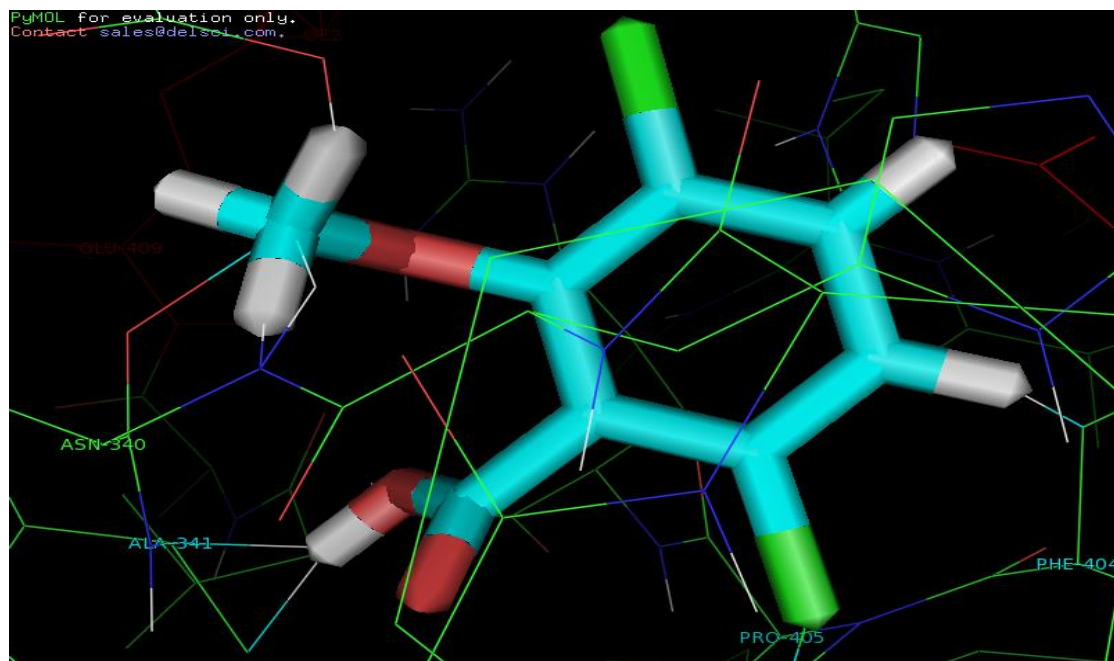
4.2.3. 2, 4-D and Dicamba binding:

On the basis of structure models and interaction of inhibitor compounds with the substrate binding residues, the physiological and biochemical effects of 2,4-D and dicamba like inhibitors on plants were investigated. Binding of other substrates, shows that there are certain conserved residues in the complexes, which are assumed to be playing very important role in 2, 4-D and dicamba. These residues include, specifically, glutamine, asparagines, tyrosine, phenylalanine, tryptophan, valine, isoleucine, leucine, alanine, and serine residues particularly interacting with substrates in the active site.

4.2.3.1. Dicamba binding:



a



b

Fig.4.23. Dioxygenase-Dicamba binding residues

[(His439, Asn155, Asn159, Asn340, Ala341, Phe404, Pro405, Arg406, Cys355, Val407 (surrounding)] (Diox-Fe-Hb-dicamba: His529, Asn340, Ala341, Phe404, Pro405, Arg406)

4.2.3.2. 2,4-D binding:

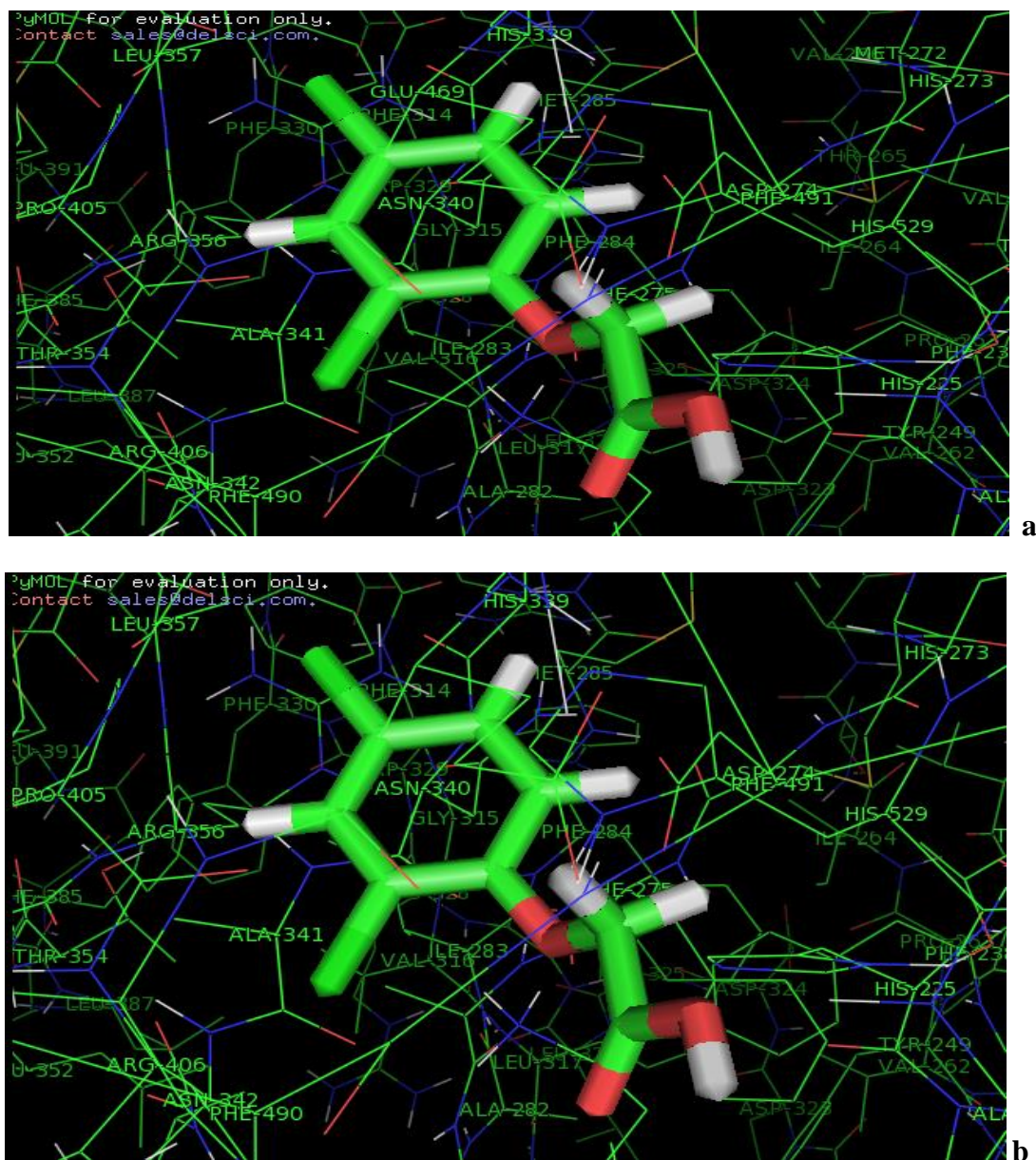


Fig. 4.24. Dioxygenase-2, 4-D binding residues

(Lys177,Leu219,Leu172,Asp176,Glu469,Pro405,Arg406,Ala224,His225,Ser1,Ser152, Tyr153,Glu174,His529,Asn159,Phe284,Thr223,Asn340,Ala341,Asp274,His273,Phe275, Ile283, Arg420)

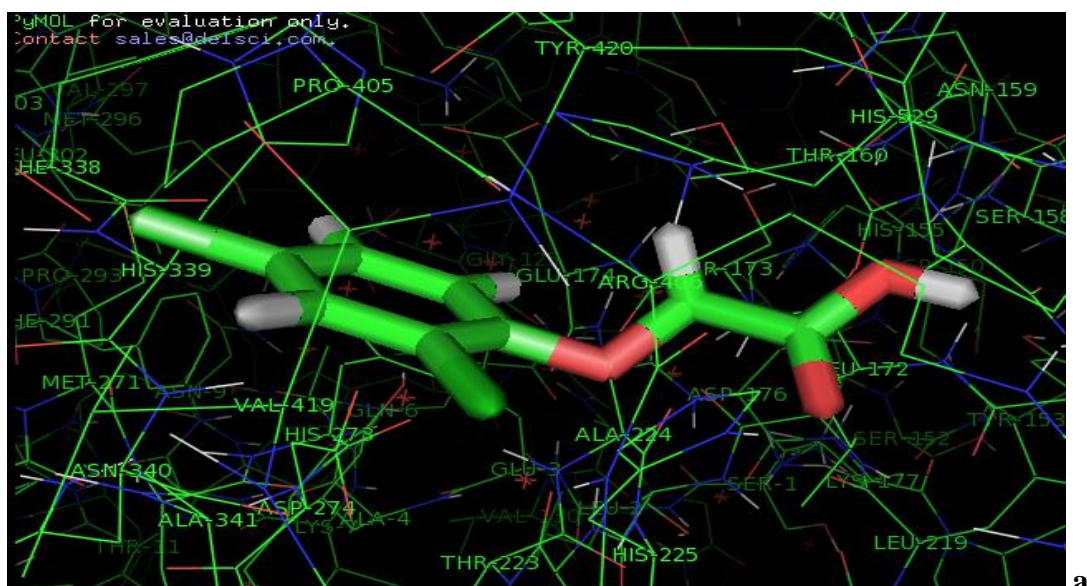


Fig. 4.25. 2,4-D binding residues (**His339**, Val419, **Pro293**, Phe291, **Tyr420**, Thr160, Ser158, **Leu172**)

Structure models for α -ketoglutarate dependent dioxygenases, like the one studied, including examples e.g. taurine/ α -KG dioxygenases (TauD), alkyl sulfatases (AtsK), 4-Prolyl hydroxylases, the factor inhibiting hypoxia-inducible factor (FIH), phytanoyl hydroxylases (PAHX), show a common β -helix fold containing ion-binding motif (His¹-X-Asp/Glu-X_n-His²), (Bugg *et al.*, 2003; Hegg *et al.*, 1997; (where n varies from 43-152 residues). In the inactivated enzyme water molecules occupying metal-ligand position get displaced upon substrate binding with the C-2 keto group of α -KG coordinating opposite the carboxylate side chain and the C-1 carboxyl group of α -KG binding opposite either His¹ or opposite His², with nearby Arg residue (located 16-23 residue beyond His²) providing more stabilization to the C-1 carboxylate. Another arginine residue located about 18 residues in the sequence beyond His² is forming an ion-pair with C-5 carboxylate of α -KG in other structures except FIH, whereas a lysine residue located in the sequence provide stabilization. This binding with His¹ or residues nearby His² serves the basis of grouping these enzymes. α -KG bound metallo centre exhibits a characteristic charge

transfer transition from metal-to-ligand conferring a definite color to the enzyme at this state, aforementioned crystallographic studies and additional spectroscopic evidences could provide clues and that the removal of final water molecule on substrate binding, which creates site for oxygen binding. The motif ($\text{His}^1\text{-X-Asp/Glu-X}_n\text{-His}^2$) are comprised of His339, Asp176, and His529; His225, Asp27 and His273; His493, Asp494, Glu495 and His503. Residues beyond His^2 , which are presumably interact with C-1 carboxylate of the α -KG are Phe531, Glu536, Gln537; Arg406, Arg420 and Phe284; Asp506, Asn512 and Asp513 respectively. Furthermore, Arg420 and additional attraction involve Thr223, Thr160 form salt bridges with 5-carboxylate of α -KG. Besides, these conserved residues, other residues are also involved in the substrate binding or surrounding, especially with respect to the structural basis of enantio-specificity (Koivunen *et al.*, 2012; Chalifoux *et al.*, 2012). This selective allylation of β -diketone moiety of the substrate would greatly enhance its utility in chemical synthesis of molecules and could have important implications in the binding of other dicarbonyl substrates. The phenoxy ring of the substrate is interacting with the amide nitrogen of Ser152, Ser158 and -OH group of Tyr153, Tyr420 and guanidine nitrogen of Arg406. The Tyr153 is lying near the oxygen atom and could be playing an important role in enantioselectivity. Residues lining the hydrophobic substrate- binding pocket include Val407, Val419, Leu172, Leu219, Ile283, Phe284, Phe275, Phe291 and Phe406. In the active site near polar carboxylate include residues His225 and His273. Also, of interest, the substrate oxygen atom predicted to lie within the interaction limit (\AA) of two carboxyl oxygen of Glu174 and Glu469. It is assumed that the Glu174 carboxyl group, if protonated or bridged by the proton of bound water molecule, could confer S-enantiomer specificity by making a hydrogen bond with the oxygen atom of the substrate (Chalifoux *et al.*, 2012). Different types of substrate enantiomers can be predicted to bind with geometries that are remarkably similar with the

carboxyl groups and aromatic rings nearly overlapping at distinct angles. Corresponding residues include Leu172, Leu219, Ile283, Glu174, Phe284 and Arg406; other residues include Ala224, Ala341, Val419, Val407, His439, His339 and Phe404. These residues are in contact with the hydrophobic ring of the substrate and cation π -interaction of the positive side chains of highly flexible Arg406 and the aromatic side groups as in the substrates facilitate the ligand binding. The residues Glu174, Ser152, Asn159, Asn340, His439, His339 and Arg420 could restrict binding of the substrate. Site directed variants (substitutions) could be used to check the subset of the residues for their possible role in controlling enantioselectivity. Glu174 prevents the binding of incorrect enantiomer and inhibition of the enzymatic functions. Tyr153 and Arg406 contribute to enantiospecificity. Phe284 could hinder the binding of correct (R) enantiomer, while helping the (S) enantiomer to exclude from the active site, so that the Phe284 variant has expanded active site and more readily binds to its substrate. This shows that Glu174 interacting with substrate oxygen with steric constraints on substrate binding imposed by Glu174, Arg406 and Asn340. This Glu174 binding prevent binding of the opposite enantiomer while slightly hinder binding of correct enantiomer (Chalifaux *et al.*, 2012). Several amino acid residues e.g. His273, His225, His339, His439, His529, Glu174, Glu469 have been identified which are likely to bind or interact with (R)- and (S)- enantiomers of the substrate, including several that possibly contribute to the enantioselectivity or enantiospecificities of the enzyme. However, many enzymes, including currently studied, the representative, of Fe (II)/ α -KG dioxygenases upon ligand-binding undergo large scale conformational changes. Thus, the task is to predict the conformational changes within structure, as well as, ligand binding orientation, the relative degrees of flexibility within protein structure and how this flexibility gets redistributed upon the formation of the complex, large scale conformational changes between two protein states, in the mean time,

calculating conformational ensembles of the ligand and protein structures that must be consistent with binding and ligand binding by docking into known enzyme conformations that approximate large-scale protein flexibility. The predicting value of this homology modeling approach is identifying residues which are capable of recognizing the (R) and (S) enantiomers of the substrate and can be assessed by the ability to comply with the binding effects of mutations or substitutions of these residues. The present study, which describes the residues involved in substrate or cosubstrate binding which could provide essential information about the structural basis of enantiomeric specificity of the enzyme. In case of dicamba binding, residues like His439, His529 and Tyr153 indicates primary catalytic roles for each of these residues. Moreover, residues like His439 and His529 play important roles in determining the unusual pH dependence of the enzyme. It should be remembered that when two distinct ionizable groups are involved in catalysis, a bell-shaped curve is usually measured for the dependence of the activity from pH (Cheng *et al.*, 2004). There are several proposed mechanisms for 2-3 ionizable groups generating two active forms of the enzyme, each catalyzed by a different catalytic efficiency. Similarly, the curve describing the dependence of the activity on substrate from pH can be explained by complex models involving not only 3-ionizable enzyme groups, hence, two differently active forms of the enzyme but also one substrate ionizable group. Three groups that influence catalysis include histidine residues at the catalytic centre that bind with the iron ion, two other histidine residues which are in close proximity of the active site cavity and bind with the substrate, His439 and His529. These residues can be tentatively identified as residues with certain pK_a values. The third group includes a lysine, a cysteine and a tyrosine, Lys177, Cys355, Tyr153 and pK_a values can be assigned to these residues. Based on these pK_a values, the optimal pH can be determined with higher degree of precision. This would allow assuming that residues like His439, His529 and Tyr153 are catalytic

residues and their functions are related to their ionization states and the contribution of these residues in the activity at different pH conditions. However, understanding the functions of these residues at different pH conditions only cannot provide enough clues as to the identification of their exact roles in the catalytic mechanism, although it may be suggestive of their involvement in the extraction and/or the release of a proton from/to the substrate. In addition, electron withdrawing groups which increase the acidity of the phenolic hydroxyl groups, reduce the catalytic activity but cause further reduction of His439. This finding suggests that His529 is involved in initial deprotonation, while His439 promotes the reaction between oxygen and the aromatic ring. The Krebs cycle intermediates are metabolized by initiation of ring cleavage reactions that are generally catalyzed by ring cleavage enzymes like dioxygenases. These enzymes catalyze the addition of two atoms of molecular oxygen into the aromatic ring with a substrate specificity more restricted with respect to the upper pathway enzymes. As the consequences, ring cleavage enzymes like dioxygenases guide and reduce the metabolic flow of aromatic compounds that can be degraded, thus playing a central role in their catabolism. The ring cleavage enzymes like dioxygenases acting on aromatic ring compounds are classified as intradiol and extradiol dioxygenases. Extradiol dioxygenase activity depends on the presence of Fe (II) ion catalytic centre and the catalytic metal is bound to 2-3 His residues and one Phe or Glutamate residue that are well conserved among the members of the family. These residues are essential for metal binding and activity. Residues like His439, His529 and Tyr 153 are also present at the active site on binding of the substrate; this indicates that these residues might be playing roles in substrate binding as well as in the catalysis. Tyr 153 or Tyr164 on binding of different substrates has been proposed to be H-bonded to the anionic oxygen of the aromatic ring acting as hydrogen donor, whereas His439 is H-bonded with non-dissociated hydroxyl group, acting as H-

acceptor. According to these observations, a mechanism is being proposed in which conserved His529 and Tyr153 or their homologous residues, act in concerted way to abstract the first proton from the aromatic ring, whereas His439 at a later stage, acts as a basic catalyst by accepting the second proton from the substrate. Finally, the position of O₂ attack and the mechanism of its insertion into the aromatic ring could be addressed by density functional theory studies and by studies with model compounds and analogues of the hypothetical reaction intermediates. The formyl and nitro groups help in the deprotonation of the substrate and hence, they should increase the efficiency of reaction, it gives to the basis of modifying substrates by adding or substituting groups in the main aromatic ring. This behavior observed can be explained by the hypothesis that these groups can also decrease their activity towards O₂ by their electron withdrawing effects and this hypothesis could be a possible alternative to the hypothesis of a nucleophilic attack of O₂ on the substrate as proposed for catechol -2,3 dioxygenases and supports an electrophilic radical attack as proposed for DHBD. The possible roles of His 439 are as follows, (i) abstraction of second hydrogen atom from the aromatic substrate (ii) stabilization of the partially reduced oxygen intermediates to promote the oxidative attack on the aromatic substrate (iii) and release of a proton at the later stages of the reaction. Excluding the fact that the primary role of His439 is acid-base catalysis might indirectly support the involvement of this residue in the activation of oxygen for the attack on the aromatic ring. Another interesting possibility exists that the positively charged His529 together with the catalytic Fe²⁺, both contribute to the deprotonation of the substrate, the aromatic ring of which rests on the imidazole ring of His529 as mentioned above. Lowering the pK_a, hence, facilitating deprotonation and the proton acceptor would be any solvent molecule bound to the Fe²⁺ and some observations also indicate that the increase in substrate acidity does not compensate the loss of hydroxyl group at position Tyr153, thus unlike His529; Tyr153

does not seem to be directly involved in substrate deprotonation. The role of hydroxyl group could be that of binding the negative oxygen of the substrate, thus stabilizing the substrate-enzyme complex. Moreover, the hypothesis can be advanced that Tyr153 and His439 play a role of inducing asymmetry in the substrate through hydrogen bonds, thus favoring specific deprotonation of the hydroxyl group, which is on the same side of Tyr153. Instead, His439 would be more directly involved in proton abstraction. Also of interest, Lys177, which interacts with the other oxygen atom of 2, 4-D, does not appear to function in the herbicide substrate degrading enzymes; however, Tyr153 is reasonably positioned to perform this function.

4.3. Self Hydroxylation: Herbicide degradation and Flavor biogenesis in Strawberry

As we know that, ripening is a highly complex process, which includes series of coordinated biochemical and physiological changes. Alteration to texture, pigmentation, and soluble solids levels which are often accompanied by the biosynthesis of flavor and aroma compounds. Non-climacteric fruits, such as strawberry does not show rise in ethylene and respiration, and in these fruits the phytohormone auxin has been shown to be of major importance in the regulation of ripening. However, ethylene in the past has been considered not to regulate the ripening of non-climacteric fruits, but there are some evidences that it might influence specific ripening related processes, such as de-greening of citrus fruits (simultaneous degradation of chlorophyll and accumulation of carotenoids) and anthocyanin biosynthesis. Certain types of sugars and organic acids, such as, glucose, fructose, malate and citrates, which accumulate in the cell vacuole, play dual role in ripening, by both providing substrates for respiration and biosynthesis of flavor and fragrance compounds. Cell wall modifications which give rise to major ripening events, which include fruits softening and tissue deterioration, various studies have focused largely as primary reason for these events. The structural components of cell wall include

cellulose, hemicelluloses and pectins. The remaining cell wall components are structural proteins such as hydroxyproline-rich glycoproteins. Expansins are some enzymes, which mediate loosening of cell wall components; another major process is the depolymerization of high molecular mass hemicelluloses. The aroma of fruit is produced by a large mix of molecules. These volatile compounds are derived from several main biosynthetic pathways, such as lipoxygenase pathway (fatty acid metabolism), amino acid degradation pathways, the polypropanoid pathway and isoprenoid pathway. The different flavor and aroma compounds play important role in fruit ripening and also in the protection of fruits against pathogens. The complex process like ripening is well co-ordinated between products of one metabolic pathway to the substrates of another pathway. In order to execute successfully, the complicated program such as ripening of fruit have to produce a new batch of 'ripening associated proteins'. The syntheses of these proteins are regulated, mediated by hormone action and signal transduction cascades. Majority of ripening related studies include identified genes and their expression, cell wall metabolism, pigmentation (anthocyanins and carotenoids), stress and pathogen-related sugar metabolism, chlorophyll degradation, hormonal regulation and signal transduction, fatty acid metabolism, water regulations and regulatory genes including transcription factors. In strawberry fruits plant hormone auxin is the main signaling molecule co-ordinating growth and initiation of ripening. During early development auxin promotes fruit growth. Auxin conjugated with either ester or amide linked and this might be part to determine its levels in the fruit and hence to control fruit development and ripening. It clearly indicates that as the auxin levels decline, it triggers the ripening by inducing de-novo synthesis of mRNAs associated with specific ripening. In similarity to other fruits, softening in strawberry during ripening might be attributed to the enzymatic degradation of cell wall material by cell wall hydrolases, such as polygalacturonases (PGs), pectinmethylesterases (PMEs) and cellulases (CELs),

which could be identified in crude fruit extracts of different fruit developmental stages. While, expansins do not hydrolyse cell wall components but promotes loosening of cell wall and catalyze wall extension by binding to the surface of cellulose microfibriles, disrupting the hydrogen bonds formed with xyloglucan molecules. Apart from changes in texture strawberry fruits undergo changes in contents of different types of phenolic compounds, providing color, flavor and resistance to attacks against pathogens and other environmental conditions. During early stages, non-tannins and mainly condensed tannins accumulate to give strawberry astringent flavor. Later in development, when fruit begin to ripen, other flavonoids such as anthocyanins and flavonols are accumulated, which parallels the rise in activities of both phenylalanine-ammonia-lyase (the initial step of phenylpropanoid pathway) and UDP-glucose: flavonoid-3-O-glucosyltransferase (adding the sugar molecule to anthocyanines). The concentration of sugars and organic acids and the ratio between them play a significant role in the overall flavor of strawberry fruit. Ellagic acid, a phenolic compound constituent of many plants derived as well from polyphenylpropanoid pathway is present in strawberry plant parts and its level increases dramatically during ripening. Major soluble sugars present in strawberry are glucose, fructose, sucrose and their levels increase during fruit maturation. Sucrose is present in lower concentration and starts to accumulate later in fruit development, while, concentrations of minor soluble sugars, e.g. galactose, inositol and xylose decrease during maturation. Invertase, which catalyzes irreversible hydrolysis of sucrose, is present in multiple forms in plant parts, mainly as wall bound and soluble acid types. It is considered, that they may play important roles in determining fruit sweetness. Both type of activities (wall bound and soluble acid types) have been characterized in strawberry. In contrast to soluble sugars (major or minor), levels of total organic acids in strawberry, rise until the turning stage and then they decrease during ripening. The three main organic acids in

strawberry are citrate, quinate and malate minor organic acids in strawberry, include isocitrate, succinate, fumarate, acetate, oxalate and aconitate. Citrate is the main organic acid present in strawberry, which determines the changing pH of strawberry fruits, during developmental stages. The pH of the fruit changes slightly from 4.6 in early stages to 3.3 at the turning stage, rising to 3.7 at the ripe stage. Other soluble components include aromatic acids which significantly contribute to fruit flavor, either imparting flavor from themselves or through the supply of precursors for the biosynthesis of aroma components. In addition, amino acids, serve as building blocks for the synthesis of proteins and as a source of nitrogen. The total concentration of free amino acids in strawberries is much higher during maturation and this might be a result of extensive protein metabolism during ripening. Asparagines are the dominant amino acid in all strawberry developmental stages and together with glutamine serve as major nitrogen-transport compounds. Lipoxygenases and hydrogenperoxide lyases activities, which are part of the biosynthesis of aldehydes, alcohols and esters are important contributors to strawberry aroma. Both activities are shown to increase coordinately, during ripening. Only a minor set of genes are supposed to be involved in the biosynthesis of aroma and flavor compounds in strawberry. Some genes encoding ester forming enzymes in strawberry and a putative pyruvate decarboxylase showing elevated expression during ripening fruit. The formation of esters and the function of amino acids as their precursors for the effects of flavor development through ripening can be due to the amino acid composition and the contents of different esters in strawberry aroma. The amino acids are individually selected which are involved in flavor formation. The products of amino acid metabolism generate aliphatic and branched chain alcohols, carbonyls, acids and esters. The free amino acids responsible include aspartate, glutamate, asparagines, glutamine, serine, alanine, proline, valine, tryptophan and histidine. On the basis of structural model and ligand binding certain conserved amino acid residues have

been identified, which are predicted to be involved in flavor development in strawberry fruits, these residues include Ala341, Asn159, Asn155, Asn340, Pro293, Pro405, Glu469, Ile283, Leu172, Phe284, Phe275, Val419, Lys177, Asp176, His339, His529, Ser152, Phe404, Val407 and Trp344. The amino acids such as asparagines, glutamine and alanine are the most abundant but the most significant changes occur in alanine contents during ripening. However, isoleucine, lysine, leucine, phenylalanine are present in less quantity in strawberry fruits. It is observed that the concentration of alanine significantly increases during early days of strawberry fruit ripening but its concentration decreases when fruit is ripen. The concentration of alanine is directly proportional to ester concentration and therefore ester biosynthesis is increased when alanine concentration reaches its maximum. Thus, the study proves the importance of amino acids in the formation of different type of ester compounds in the strawberry aroma. There is a kind of a self-preservation mechanism in strawberry, which is due to the consequence of retaining solvent in the sixth coordination site of the metal ion in the absence of prime substrate, which increases the activation energy required for O₂ attack but solvent must first be displaced. These reactions are also known as self hydroxylation reactions, considerably slower than the prime substrate into products transformation, which minimizes the destructive effects of these aberrant side effects. Structural models of α -KG dependent dioxygenases combine with ligands or substrates, such as phenylamineimidazole (PIM) which are also considered as IDO inhibitors, 2, 4-D or Dicamba, suggest, that the conversion of α -ketoglutarate (α -KG) into succinate in the absence of substrate, and an outer-sphere tyrosine residue act as the metal ligand, which could prevent or significantly limit the exposure of alternative oxidation targets to the high valent intermediates. This self- hydroxylation reaction can prevent more destructive and non-selective oxidations such as peptide chain cleavage by radicals, resulting in permanent loss of catalytic activity. In fact, self hydroxylation of

residues Phe404 and Phe284, allows the catalytic functions of enzyme to recover, since, Tyr153, performs an important structural and catalytic role. Significantly, this study of dioxygenases could provide new insights, that two mechanistic pathways in the self-hydroxylation of a single non-heme iron enzyme are involved. This demonstrates that mononuclear iron centres with 2-His-1-carboxylate motifs are versatile (Bugg *et al.*, 2003; Hegg *et al.*, 1997; Liu *et al.*, 2001) in activating dioxygen with the cosubstrate α -ketoglutarate (α -KG) in a bidentate coordination; and preassumably, remaining coordination site is for O₂. However, this study focuses on dioxygenases, an iron containing (ferrous) enzyme, that may catalyze the biodegradation of herbicides, such as 2, 4-dichlorophenoxyacetate (2, 4-D) or Dicamba. The enzyme complex, in absence of the herbicide substrate 2, 4-D or dicamba, could react with O₂ to form a compound, which arises due to the hydroxylation of an active site residue Tyr153 and provides insight into the type of reactive species involved in oxygen activation. As ligand-to-metal charge transfer transition and the enhanced ligand vibrations, which is mainly caused by a hydroxylated arene and it differs from the previously characterized Tyrosine residue. This study can be done by resonance raman spectroscopy, which would indicate some chromophore and loss of the color, on addition of the primary substrate 2,4-D or dicamba and subsequent exposure to O₂, products, e.g.2,4-dicholorophenol, glyoxalate, succinate and CO₂. In absence of substrates, 2, 4-D or Dicamba the enzyme complex reacts with O₂, to form a compound, which is a high spin Fe (III) complex. This high spin which can be detected from EPR studies would indicate the presence of iron, and that the compound is formed by ligand-to-iron (III) charge transfer transition. The metal-ligand vibration studies can be done for more amino acid residues, which would clearly indicate the modifications (hydroxylation) of which residue are actually taking place, in such cases, Phe, Tyr or Trp are some amino acids, which are usually involved in this kind of hydroxylations.

Comparison of other model compounds, such as indoleamines can also prove to be useful. And, it suggests, that the product ion spectra can be searched against theoretical digest of the enzyme complex *in silico*. The signatory peak of the corresponding modified compound can be traced at nanoscale capillary LC-MS /MS analysis. This MS/MS analysis could indicate the amino acid hydroxylation, such as Tyr-OH, or Trp-OH, and it also shows, the surrounding or nearby residues, such as His, a metal ligand. But, it has been observed in some studies that the hydroxylated compound-enzyme complex was inactive, however treatments with other stronger reductants recover enzymatic activity, so, the catalytic activity of the enzyme can be regained presumably, by displacing the hydroxylated-amino acid residue. The hydroxylation of Trp82 and Tyr153 occur at the expense of α -KG, which on decarboxylation produces succinate and CO₂, in an O₂ dependent reaction. This class of enzymes has the active oxidant e.g. either an iron-peroxo or a high valent iron-oxo species derived from the decarboxylation of α -KG. Without primary substrate, this oxidant attacks an aromatic residue nearby in the active site to form the OH-Tyr or OH-Trp. An intermediate that allows water to fully incorporate into Tyr-OH or Trp-OH in dioxygenases, which excludes the possibility of direct hydroxylation of Trp82 or Tyr153 by iron-peroxo species. Many α -KG dependent dioxygenases undergo uncoupling and self inactivate. The first instances of enzyme self-hydroxylation are visualized as the formation of a chromophore (Liu *et al.*, 2001). Such self-hydroxylation occurring in other α -KG dependent enzyme i.e. Hydroxyphenylpyruvate-dioxygenases (HPPD), which was reported to have a blue chromophore (λ_{max} ~595 nm) associated with tyrosinate ligand. However, recently reported crystal structures of the enzyme show no tyrosine is coordinated to the Fe-centre. But, Tyr residue in the structure model perform this function, so, the reasons for assuming that this tyrosine residue might be involved in self-hydroxylation of the strawberry dioxygenase, which is presumably 9-cis-

epoxycarotenoid or α -Ketoglutarate dependent dioxygenase. So, based on the structure models, the contradictions can be resolved by suggesting that the blue chromophore could result from uncoupled self hydroxylation of a phenylalanine residue in the active site. Such self hydroxylations are also found in tyrosine hydroxylases and these results lead to consider that in the non-heme iron enzymes the “uncoupled” self-hydroxylation reactions are not an unusual side reactions. These side reactions have biochemical significance that it serves to protect the enzyme from a more destructive, irreversible oxidation, such as peptide chain cleavage due to radicals (i.e. inhibitor substrates), which could result in permanent loss of the catalytic activity. The mechanism of this fascinating class of dioxygenases includes rates of uncoupled reaction relative to coupled reactions which provide an opportunity to investigate intermediate steps. This hydroxylation reactions could also be involved in herbicide substrates, such as 2,4-dichlorophenoxyacetate (2,4-D) or 3,6-dichloro,2-methoxybenzoic acid (Dicamba) degradation. Although, monooxygenases or dioxygenases are not completely involved in herbicide degradation but these enzymes can play an important role in herbicide metabolism and hydroxylation reactions mainly catalyze this metabolic process. As we know that Fe(II) and α -KG dependent dioxygenases metabolize the co-substrate into succinate and CO_2 , the enzyme can use different phenoxyacetate or α -ketoacids but most efficient catalytic activities were shown using 2,4-D and α -ketoglutarate. The structure model of dioxygenase possesses multiple essential histidine, cysteine, lysine, alanine residues, which are supposed to be involved in herbicide degradation. Herbicides such as 2, 4-Dichlorophenoxyacetic acid (2, 4-D) and Dicamba (3, 6-dichloro, 2-methoxybenzoic acid) that are considered broadleaf herbicides and HPPD inhibitors could be involved in side chain removal, forming 2, 4-dichlorophenol and muconate, succinate and subsequent metabolism of this intermediate. The first step of herbicide degradation is catalyzed by α -ketoglutarate dependent dioxygenase. It requires α -

KG as co-substrate and converts this compound to carbon dioxide and succinate. And the subsequent metabolism of succinate, as we know that succinate is an organic acid and with other solid or soluble sugars, plays an important role in flavor biogenesis and during fruit ripening in strawberry, thus, it can be assumed that the self hydroxylation of dioxygenases are greatly involved in strawberry flavor biogenesis, as well as herbicide degradation. Although, the enzyme function includes hydroxylation of a wide range of phenoxyacetates and related compounds also shows greater affinity and catalytic efficiency for α -ketoglutarate and 2, 4-D. The enzymatic activity also depends on the substitutions, in the substrate, such as halogenated or non-halogenated, methylated etc. Furthermore, α -ketoglutarate, malonate, succinate and glutarate are not supporting hydroxylation, thus, in absence of 2, 4-D or Dicamba, α -ketoglutarate are not decomposed. As mentioned above, (Table-4.1), Monsanto, Dow Agrosiences and Syngenta patent papers, describe the Dicamba tolerant transgenic plants, expressing polypeptide encoding Dicamba monooxygenase which comprises certain conserved amino-acids, such as arginine, histidine, asparagines, lysine, alanine, tryptophan, tyrosine, glutamine etc. These polypeptide sequences have some specific amino acid residue at specific positions, the positions which are mainly taken into consideration are 2,3 and 112, some specific types of residues at these positions include Ala,Thr,Cys or Leu,Thr,Cys or Leu,Thr,Trp or Ala,Thr,Trp, repectively. But, changing amino acids of a protein is required to create an equivalent or even improved modified polypeptide. In a protein structure certain amino acids can be substituted for other amino acids without much loss for the interactive binding capacity such as binding sites on substrate molecules. It is contemplated, that these changes in the polypeptide sequences and codon usage give rise to herbicide tolerance, without much loss of their activity and biological utility. Hydropathic index of amino acids are being considered and the substitution of amino acids with similar hydropathic index or

score can still obtain a biological functionally equivalent protein. Exemplary substitutions taking these characteristics into consideration, is well known which, include: glutamate and aspartate, arginine and lysine, glutamine and asparagines, serine and threonine, valine, leucine and isoleucine. Diethylpyrocarbonate (DEP) are certain compounds, which are histidine selective, thus the presence of histidines at substrate binding site and active catalytic site can be traced, as after DEP treatment, the enzyme becomes inactive. The substrates such as 2, 4-D, Dicamba, α -KG, Fe (II) and combinations of these substrates can be used to protect the enzymatic inhibition induced by DEP, as binding of these substrates include histidine at substrate binding and catalytic active site and notably, Dioxygenase have histidine residues at the active and substrate binding sites, thus using DEP would be a suitable confirmation test for the presence of histidine residues at these sites. These substrates individually may not be able to protect enzyme from inactivation, but combinations of these compounds can protect enzyme inactivation induced by DEP or there could be a significant decrease in enzyme inactivation. Furthermore, there are some conserved histidine residues at substrate binding sites and some histidine residues are buried in the protein at the Fe (II) binding site. Binding of 2,4-D, Dicamba and α -KG, protects these histidine residues at substrate binding site and additionally, at Fe(II) binding site, Fe(II) ligands by steric constraints. While ascorbate increases the enzymatic activity, thus in 2,4-D or Dicamba degradation and ascorbate induced stimulation of activity are typical characteristics of α -KG dependent dioxygenases and Fe(II) e.g. prolyl, lysyl, aspartyl, deacetoxycephalosporin hydroxylases or mechanistically similar p-hydroxyphenylpyruvate hydroxylase etc. The biodegradation of a plant derived compound containing an aromatic ring in ether linkage to an acidic side chain mainly involve these enzymes. Thus, enzymes which are involved in lignin biodegradation e.g peroxidases degrade the lignin substrate aromatic ring into small pieces and pieces, which retain ether

linkages and subsequently, being degraded by α -KG dependent dioxygenases. Similar, property of the enzyme can also be applicable for synthetic herbicides, e.g. the synthetic version of plant hormone auxin 2,4-D or Dicamba, which have the same aromatic ring in ether linkages to acid side chains. The study discusses mechanism for 2,4-D and Dicamba binding and possible roles for amino-acid residues in these dioxygenases in the degradation of these herbicides on structural basis.

CHAPTER 5

METHODS

METHODS

Over the years many researchers have used X-ray crystallography and homology modeling for solving protein structures; here, in the present study, homology modeling techniques have been used to solve the structural issues related to strawberry dioxygenase. The protein sequence was analyzed before generating the structural model as there are certain specificities required for validating the concept of homology modeling. Any protein sequence must be analyzed as it may provide necessary information, based on which one can study the protein. Therefore, the strawberry dioxygenase amino acid sequence was deduced from NCBI database (www.ncbi.nlm.nih.gov). The sequence analyses began with BLAST analysis of the deduced dioxygenase sequence using BLAST tool of NCBI. Multiple alignments were performed by using CLUSTAL W tool of EMBL-EBI (www.ebi.ac.uk). Thus, the strawberry dioxygenase sequence was analyzed for transit peptides, whether, any transit peptide is present in the dioxygenase sequence iPSORTserver (ipsort.hgc.jp/) was used. Further, the sequence was analyzed for hydropathy and translocon segments. Membrane protein explorer (MPE) software was used for analyzing hydropathy and β -barrel translocon plot. Again, the sequence was analyzed for β -barrel outer membrane protein by using PRED-TMBB server (biophysics.biol.uoa.gr/PRED-TMBB/). The generated 2D representation of dioxygenase membrane position indicated the dioxygenase protein part that is in the outer side of the membrane. To confirm the membrane orientation of dioxygenase PPM server (opm.phar.umich.edu/server.php) was used that indicated the tilt of dioxygenase protein from the membrane and membrane buried residues. The dioxygenase sequence was analyzed for conserved domain in the protein sequence by using CDD and Batch CDD tool of NCBI (www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml); which indicated conserved domains in the sequence such as RPE65 superfamily. The sequence was further analyzed

for substrate binding domain, Phyre2 server (www.sbg.bio.ic.ac.uk/phyre2/) was used for finding substrate binding domains. An approach that uses designed sequences that fold into known secondary structures was implemented to develop a method for identifying substrate binding domains; the method includes generation of small template structures from dioxygenase sequence parts. However, Rosetta @Home (boinc.bakerlab.org/) can also be used for generating small protein structures. The generated small templates, based on template based homology searches provide the possibility of types of substrate, which might bind or interact with the corresponding protein. Furthermore, the dioxygenase sequence was analyzed for folds and motifs. NAD-FAD and Rossmann folds (www.ncbi.nlm.nih.gov) were analyzed and the motif analyses were performed using MyHits server (myhits.isb-sib.ch/cgi-bin/motif_scan). Usually, MyHits server generates motifs along with predictions for possible glycosylation, amidation and phosphorylation sites but for further confirmation, the sequence was analyzed by using LRR finder (www.lrrfinder.com/) which is a server for finding leucine rich repeats in the protein sequence. The LRR finder server generated predictions for all types of glycosylations such as N-, C- and O- glycosylation and phosphorylation sites. However, LRR finder indicated no leucine rich repeats in the sequence but drug substrate binding indicated LRR domains in the protein structure, which have important structural and functional roles. Furthermore, the sequence was analyzed for coil and coiled coil probabilities in the sequence; Marcoil, PSSM (www.mybiosoftware.com/protein-sequence-analysis/) servers were used for such analyses. The dioxygenase sequence was analyzed for secondary structure and topology predictions, using PSI-PRED (128.16.10.201/psipred/) and EMBL-EBI (www.ebi.ac.uk/) tool ProFunc for protein functions. These analyses provided all relevant information related to protein secondary structure such as α -helices, β -sheets, β -strands, β -hairpins, β -turns, β -bulges and γ -turns. With more than 30% similarity and less gap in alignment, the template

3NPE_A served as an ideal template for generating dioxygenase structure. The dioxygenase protein structure was generated using Geno3D server (geno3d-pbil.ibcp.fr/). The models were generated with the multipass model expectation value threshold 0.002 in the expectation value of 10 and the alignment was done in the matrix BLOSUM62. The same analyses for the model sequence were repeated to confirm predictions as generated by different servers for strawberry dioxygenase sequence. Further, the model was analyzed for validation and quality assessment by using QMEAN density plot, Ramachandran plot and ERRAT2 (nihserver.mbi.ucla.edu/ERRATv2/) analysis by using PROCHECK (www.ebi.ac.uk/thornton-srv/software/PROCHECK/) and SWISS-MODEL servers (swissmodel.expasy.org/). Furthermore, reverse template comparison with PDB structures and existing PDB structures was performed using EMBL-EBI (www.ebi.ac.uk/) protein function tool ProFunc. The dioxygenase structural model was further analyzed for electrostatic and hydrophobic patches. The electrostatic patch analysis was performed by using Patch Finder Plus, PF Plus server (pfp.technion.ac.il/) at Technion, Haifa, Israel and hydrophobicities by using server at NIH-MBI (nihserver.mbi.ucla.edu/SAVES/). Cysteine disulphide bonds in the structure model were predicted using DiANNA1 server (www.ncbi.nlm.nih.gov) at NCBI and the pores and pore lining residues in the structural model were predicted by using EMBL-EBI PoreWalker tool, (www.ebi.ac.uk/) which generated structure model PDBs with pores and pore lining residues. The server also generated pore cavity features, such as pore shape and pore diameter plots. The pore visualization model PDBs were also generated by the Porewalker tool of EMBL-EBI. Different substrates and their 2D and 3D representations were selected from databases for chemical compounds, PubChem Compounds (www.ncbi.nlm.nih.gov/pccompound) and PubChem substance (www.ncbi.nlm.nih.gov/pcsubstance) at NCBI. The PDB viewers, such as Raswin,

(rasmol.org/OpenRasMol.html), Chimera (www.cgl.ucsf.edu/chimera/) and PyMol (www.pymol.org/) are mainly used for determining the dioxygenase structure and generation of suitable pictures. The dioxygenase structure model was analyzed for most possible substrate binding pockets; this analysis was performed by using Pocket finder tool at Bioinformatics and Biological Computing Biological Services (BBCU) (bip.weizmann.ac.il/toolbox/structure/binding.htm). The substrate docking studies were performed by using Hex (hex.loria.fr/), ClusPro (cluspro.bu.edu/), RosettaDock (graylab.jhu.edu/docking/rosetta/), Patchdock (bioinfo3d.cs.tau.ac.il/PatchDock/), dockingserver, (www.dockingserver.com/) and GRAMM-X docking servers (vakser.bioinformatics.ku.edu/resources/gramm/grammx).

CHAPTER 6

CONCLUSION

CONCLUSION

There are striking differences between studies pertaining to protein structures as determining protein crystal structure using X-ray diffraction or Laser diffraction and Neutron scattering methods, in the present study, dioxygenase structure generated from strawberry dioxygenase amino acid sequence was defined by a homology modeling approach. Perhaps most intriguing are the sequence analyses using various servers and tools which are as important as solving crystal structures. The sequence analysis provided required informations about its orientation in the membrane and its functions. The study started from sequence analysis, structural model was generated and subsequently dioxygenase active site and substrate binding pockets were studied. The main part of protein structural analysis have been identifying and studying substrate binding pockets; which comprises different secondary structural symbols so called helical domains, β -sheets and long loops. However, long or short loops have their own significance in the protein structure and identifying amino acid residues that play important roles in providing rigid or flexible conformations to the structure and also the interacting residues provided clues to investigate the potential implications of hydroxyl, carbonyl, ammonium, nitro and guanidine groups of different amino acids, that how, these groups are involved in forming bonds and salt bridges; their weak or strong hydrophobic or polar interactions mediated by electron charge transfer or weak vanderwaal forces also helped to proposing mechanisms of various possible reactions including identifying products and intermediate states. Furthermore, various substrate binding domains were identified and different substrates were analyzed, including the most ideal substrate for dioxygenases, dioxygen, demonstrating the mode of oxidative transformations, which is an important function of α -ketoglutarate (α -KG) dependent dioxygenase. Furthermore, various substrates and protein interactions with dioxygenase were investigated, which may

provide new insights into the interacting propensity of dioxygenase in terms of coordinating the signal transduction sensed by different receptors, however, the study remain focused on Integrin receptors. Integrin receptors are receptors which play very important role in sensing mechanical cues such as extracellular matrix and in relaying such transmission to different responsive molecules which in turn are activated and channelize signals downstream performing various cellular functions. Several such proteins which may interact with dioxygenase were identified that includes Nematode Allergen and Lipids, α -Epoxycholesterol, Chorismate mutase or Hydroxyphenylpyruvate synthase and proteins related to cell cytoskeleton such as Actin, Myosin ATPase motors, cell adhesion proteins such as Focal adhesion kinase (FAK), which also binds to Integrin receptors and perform many regulatory functions. Other proteins for example, Electron transport protein Cytochrome 'c', Erythrocyte band 3 protein were also identified, which may solve structural issues related to potential dioxygenase role in energy related or energy derived functions for instance stress and diseased conditions where ATP levels are crucial for proper functioning of different enzymes. Recently, solved RNA interference, RNAi complex structure provided insights to using dicer protein for determining, its structural interactions and their functional significance. Moreover, Immunoglobulin repeats were identified in the dioxygenase sequence, which helped to assuming that dioxygenase plays very significant role in the immune responses triggered by various pathogens in plants as well as in mammals. The role of MAMP, PAMP and DAMP alongside the R-proteins present on the plant cell surfaces are mentioned. In terms of determining dioxygenase interactions with Immunoglobulin with studied dioxygenase and Immunoglobulin kappa (κ) light chain dimer structural complex, which may also provide insights into the potential implications of dioxygenase and Immunoglobulin complexes in various innate and adaptive immune responses in mammals. Specific amino acid residues

have their own functional importance along with their structural roles; as different amino acids have been reported to be involved in various cellular processes in eukaryotes as well as in prokaryotes but the main focus of the present study has been to determine the possible role of plant hormones such as Absciscic acid in the strawberry fruit ripening and identifying potential new targets for herbicide inhibitors. Some herbicide inhibitors were selected from the herbicide database, which many commercial companies including Dow, Monsanto and Syngenta etc. are currently developing as potential hydroxyphenylpyruvate dioxygenase (HPPD) inhibitors and planning to commercialize them by patent filing and marketing them probably in 2014 or 2015. These herbicides and potential dioxygenase inhibitors include Cyclohexadione Ca which is a dioxygenase substrate α -ketoglutarate analog, which may serve as dioxygenase inhibitor as well, Abamine, Indoleamine imidazole, NDGA and synthetic auxin herbicides such as 2,4-D and Dicamba. Their structural and functional interactions indicated their significance in their functions of inhibiting the dioxygenase enzymatic activity and also in the herbicide degradation process in the absence of herbicides. The amino acid interactions also indicated the possibility of self hydroxylation and phosphorylation reactions by specific amino acids such as tyrosine and phenylalanine, whereas alanine, leucine, glutamine and also phenylalanine may provide some stereo-selectivity to the enzyme, distinguishing R-enantiomers from S- enantiomers types. Self hydroxylation reaction is an important survival mechanism in this class of α -ketoglutarate dependent dioxygenases that leads to the assumption that coupled or uncoupled self hydroxylation reactions are not very uncommon reactions in these groups of enzymes. Furthermore, strawberry health benefits were investigated, as strawberry fruit and leaf extracts contain several natural chemical compounds including alcohol, flavonoids, tannins, organic acids such as succinate, fumarate, citrate etc. and anthocyanins which have many functional implications. The

flavonoids present in strawberry extracts help lowering cholesterol levels and other dietary fats. Moreover, the strawberry also contains allergens as found in glutenin and wheat. Current findings are all consistent with the recent developments in crystallography, several crystal structures of different biomolecules have been determined that show the architecture of functional complexes, however, the present study deals with homology modeling, therefore, to gain more structural information about the rigid and flexible regions within the protein and how this flexibility redistributes itself upon complex formation predicting large scale conformational change between two protein states, further crystallizing strawberry dioxygenase is required but here, the present is able to evaluate conformational ensembles of the ligand and protein structures that are consistent with binding and predicting ligand binding by docking into the dioxygenase structure model considering an ideal enzyme conformation that approximate larger scale protein flexibility. Molecular dynamics simulations are other options; however, it is generally difficult to assess substantially different ligand orientation and protein conformations in molecular dynamics simulation. In this case, the situation is further complicated as the study has been focused on homology modeling, but the study is able to predict the possible influence of subunit interfaces, amino acid residues and substrates on active site conformations. Therefore, the accuracy of homology modeling and predictions of amino acid interactions and their structural and functional roles within the dioxygenase structure and in various pathways mediated by dioxygenases and other interacting proteins and factors or activators may be limited by not anticipating medium- to large- scale protein conformational changes. The predictive value of this modeling is identifying residues likely to be responsible for activation and functioning of dioxygenase-substrate complexes and dioxygenase and other protein complexes that can be assessed by consistency with the binding effects of mutation of these residues. Site specific variants of dioxygenase and

interacting proteins can be created, their activities can be measured and in selective cases detailed kinetic experiments can be performed. These results can confirm the residues involved in substrate and cosubstrate binding and can provide useful insights into the structural basis of dioxygenase functions. A molecular packing analysis of strawberry dioxygenase crystals can provide insights into the surface of proteins that interacts with the membrane and other proteins. It is speculated that the predictions made in the present study for residues involved in such interactions can be observed in similar packing arrangements in crystals of this class of monotopic membrane proteins, such as RPE65, Prostaglandin H₂ synthases and squalene- hopene cycalse that are crystallized in the presence of detergents. The surprising feature of the present study have been the substrate binding pockets and the motifs, which are identified and speculated that most of these substrate binding pocket and motif residues are charged and polar ones that form an intricate hydrogen bond- network, mediating the interactions between the transmembrane and the intracellular domain. The substrate binding residues indicated some QREN and LR domains which seem to be important as they may have structural and functional significance. The catalytic triads of Asp-His-Glu, Cys-His-Asp- and His-Glu/Asp-His residues indicated consistency with the enzyme mechanism. The ortho-rearrangements mediated by His residues and π -stacking interactions between Phe residues are also consistent with the mechanism involved in reactions of α -ketoglutarate dependent dioxygenases. Protonation and deprotonation of His residues may affect hydrogen bond formations and the hydroxylation of tyrosine residues which is an important outcome in such reactions. Nonetheless, further interpretation is complicated as these essential modifications must be confirmed by LC-MS/MS analysis against the hypothetical protein digest.

Overall, structural and functional informations related to dioxygenase presented in the study may contribute to understanding the physiological conditions in which dioxygenase plays a key role as an important node in various related signaling pathways. From an evolutionary standpoint, it might be expected that dioxygenase catalytic mechanism would resemble that of other carotenoid oxygenases and mechanisms that involve participation of molecular oxygen and other factor or activators in the isomerization reaction and oxidative transformations cannot be definitively ruled out but would require especially complex chemistry that is not justified by experimental evidences available to date. The present study can provide such integration of biochemistry, chemistry, plant physiology on structural basis which requires time to be the part of currently available documents and in the flow stream of knowledge.

REFERENCES

REFERENCES

- Abu-Omar M M, Loaiza A, Hontzeas N, Reaction Mechanisms of Mononuclear Non-Heme Iron Oxygenases. *Chemical Reviews*, 2004.
- Adams-Phillips L, Barry C, Giovannoni J, Signal Transduction Systems Regulating Fruit Ripening, *Trends Plant Sci.*, 9, 331–338, 2004.
- Adhikari A N, Peng J, Wilde M, Xu J, Freed K F, Sosnick T R, Modeling Large Regions in Proteins: Applications to Loops, Termini, and Folding, *Protein Science*, 2012.
- Ahrens W H, Herbicide Textbook, 7th edition, Champaign, IL, *Weed Sci Soc. of Amer.*, 352, 1994.
- Alvarez-Saurez Jose M, Dekanski D, Ristic S, Radonjic N V, Petronijevic N D, Giampieri, F, Astolfi, P, Gonzalez-Paramas A M, Santos-Buelga C, Tulipani S, Quiles J L, Mezzeti B, Battino M, Strawberry Polyphenols Attenuate Ethanol-Induced Gastric Lesions in Rats by Activation of Antioxidant Enzymes and Attenuation of MDA Increase, *Plos One*, 6, 10, e25878, 2011.
- Andersen J M, Turley S D, Dietschy J M, Low and High Density Lipoproteins and Chylomicrons as Regulators of Rate of Cholesterol Synthesis in Rat Liver in vivo (Cholesterol Esters/Lipoprotein Clearance/Lipoprotein Transport). *Proc. Natl. Acad. Sci. USA*, 76, 1, 165-169, 1979.
- Baldwin J E, Adlington R M, Crouch N P, Pereira I A C, Incorporation of ¹⁸O Labeled Water into Oxygenated Products Produced by the Enzyme Deacetoxy Deacetylcephalosporin C-Synthase, *Tetrahedron*, 49, 7499-7518, 1993.
- Ballesteros J A., Deupi X, Olivella M, Haaksma E E J, Pardo L, Serine and Threonine Residues Bend α -Helices in the $\alpha_1 5_g 2$ Conformation. *Biophysical Journal*, 79, 2754–2760, 2000.

Baker D, Sali A, Protein Structure Prediction and Structural Genomics, *Science*, 294, 93-96, 2001.

Barclay A N, Membrane Proteins with Immunoglobulin-like Domains—A Master Superfamily of Interaction Molecules, *Genomics and Immunology*, 15, 4, 215–223, 2003.

Bateman A, Eddy S R, Chothia C, Members of the Immunoglobulin Superfamily in Bacteria, *Protein Sci.*, 5, 1939-1941, 1996.

Beyer E M, Duffy M J, Hay J V, Schlueter D D, Sulfonylurea Herbicides. in P.C. Kearney and D.D. Kaufman, ed. *Herbicides: Chemistry, Degradation and Mode of Action*, Marcel Dekker, 3, 117-189.

Bomgardner M M, War On Weeds. C & E N. *American Chemical Society*, 90, 21, 20-22, 2012.

Bood K G, Zabetakis I, The Biosynthesis of Strawberry Flavor (II): Biosynthetic and Molecular Biology Studies. *Journal of Food Science*, 67, 1, 2002.

Brigelius R, Spottl R, Bors W, Lengfelder E, Saran M, Weser U, *FEBS Lett.*, 47:72, 1974.

Britton G, Barry P, Young A J, Carotenoids and Chlorophylls: Herbicidal Inhibition of Pigment Biosynthesis, In Dodge A D, ed., *Herbicides and Plant Metabolism*, *Soc. for Exp. Bot. Sem. Series*, 38, 50-72, 1989.

Buckingham S, Picking the Pockets of Protein- Protein Interactions, *Charting Chemical Space*, 2004.

Bugg T D H, Dioxygenase Enzymes: Catalytic Mechanisms and Chemical Models, *Tetrahedron*, 59, 7075–7101, 2003.

Buhler D D, Factors Influencing Fluorochloridone Activity in No-Till Corn. (*Zea mays*). *Weed Sci.*, 36, 207-214, 1988.

Carkeet C, Clevidence B A, Novotny J A, Antocyanin Excretion by Humans Increases Linearly with Increasing Strawberry Dose, *J. Nutr.*, 138, 5, 897-902, 2008.

Chai Y M, Jia H F, Li C, Dong Q H, Shen Y Y, FaPYR1 is Involved in Strawberry Fruit Ripening, *J Exp. Bot.*, 62, 5079-89, 2011,

Chalifoux W A, Reznik S K, Leighton J L, Direct and Highly Regioselective and Enantioselective Allylation of β -Diketones, *Nature*, 487, 86-89, 2012.

Chapman M J, Ginsberg H N, Amarenco P, Andreotti F, Borén J C, Alberico L, Descamps O S, Fisher E, Kovanen P T, Kuivenhoven J A, Lesnik P, Masana L, Børge G, Ray K K, Reiner Z, Taskinen M, Tokgözoğlu L, Tybjærg-Hansen A, Watts G F and on the European Atherosclerosis Society Consensus Panel, Triglyceride-Rich Lipoproteins and High-Density Lipoprotein Cholesterol in Patients at High Risk of Cardiovascular Disease: Evidence and Guidance for Management, *European Heart Journal*, 32, 11, 1345-1361, 2011.

Crouch N P, Adlington R M, Baldwin J E, Lee M H, Mackinnon C H, A Mechanistic Rationalization For The Substrate Specificity of Recombinant Mammalian 4-Hydroxyphenylpyruvate Dioxygenase (4-HPPD), *Tetrahedron*, 53, 6993-7010, 1997.

Crow G R *Org. React.*, 43, 251, 1993.

Cunningham FX, Gantt E, Genes and Enzymes of Carotenoids Biosynthesis in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 49, 557-583, 1998.

Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR, Abscisic Acid: Emergence of A Core Signaling Network. *Annu Rev Plant Biol.*, 61, 651-679, 2010.

Dasari V R, Kaur K, Velpula K K, Dinh D H, Tsung A J, Mohanam S, Rao J S. Downregulation of Focal Adhesion Kinase (FAK) by Cord Blood Stem Cells Inhibits Angiogenesis in Glioblastoma. *Aging*, 2, 11, 791-803, 2010.

Daszkowska-Golec A, Szarejko I, Open or Close the Gate – Stomata Action Under the Control of Phytohormones in Drought Stress Conditions. *Front Plant Sci.*, 4, 138, 2, 1, 2013.

Dean M, Rzhetsky A, Allikmets R, The Human ATP-binding Cassette (ABC) Transporter Superfamily. *Genome Res.*, 11, 7, 1156-66, 2001.

Dejana E, Raiteri M, Resnati M, Lampugnani, M G. Endothelial Integrins and Their Role In Maintaining The Integrity of The Vessel Wall. *Kidney International*, 43, 61-65, 1993.

DeMali K A, Wennerberg K, Burridge K. Integrin Signaling To The Actin Cytoskeleton. *Current Opinion in Cell Biology*, 15, 572–582, 2003.

Deshmukh S S, Tang, K. Kalman, L. Lipid Binding to the Carotenoid Binding Site in Photosynthetic Reaction Centers. *J. Am. Chem. Soc.*, 133, 16309– 16316, 2011.

Dharmasiri N, Dharmasiri S, Estelle M. The F-box Protein TIR1 is An Auxin Receptor. *Nature*, 435, 441-5. 2005.

Dowd P, Ham S W, Geib S J, *J. Am. Chem. Soc.*, 113,7734, 1991.

Edwards RJ, Davey NE, Shields DC Slimfinder: A Probabilistic Method for Identifying Over-Represented, Convergently Evolved, Short Linear Motifs in Proteins. *Plos one*, 3, e967, 2007

Eleniste PP, Bruzzaniti A, Focal Adhesion Kinases in Adhesion Structures and Disease. *Journal of Signal Transduction*, Article ID 296450, 1-12, 2012.

Eleniste M D, Cellular Functions of FAK Kinase: Insight into Molecular Mechanisms and Novel Functions. *Journal of Cell Science*, 123, 1007-1013, 2010.

Ellis M K, Whitfield A C, Gowans L A, Auton T R, Provan W M, Lock E A, Lee D L, Smith L L, Characterization of the Interaction of 2-[2-Nitro-4-(trifluoromethyl) benzoyl]-4,4,6,6-tetramethylcyclohexane-1,3,5-trione With Rat Hepatic 4-Hydroxyphenylpyruvate dioxygenase. *Chem Res Toxicol.*, 9, 24-27, 1996.

Engh R A, Huber R, Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement. *Acta Cryst.* A47, 392-400, 1991.

Eswar N, Eramian D, Webb B, Shen MY, Sali A: Protein Structure Modeling With Modeller. *Methods Mol Biol.*, 426,145-159, 2008.

Eyre T A, Partridge L, Thornton, J M, Computational Analysis of α -helical Membrane Protein Structure: Implications For the Prediction of 3D Structural Models. *Protein Engineering, Design & Selection*, 17, 8, 613–624, 2004.

Faucher E P, Vogel V, Integrin Activation Dynamics between the RGD-binding Site and the Headpiece Hinge. *The Journal of Biological Chemistry*, 284, 52, 36557–36568, 2009.

Fernández B G, Jezowska B, Janody F. Drosophila Actin-Capping Protein Limits JNK Activation by the Src Proto-Oncogene. *Oncogene*, 2013.

Finkelstein R R, Rock C D. Abscisic Acid Biosynthesis and Response. The Arabidopsis Book, *American Society of Plant Biologists*. 2002.

Flaus A, Martin DMA, Barton GJ, Owen-Hughes T. Identification of Multiple Distinct Snf2 Subfamilies With Conserved Structural Motifs. *Nucleic Acids Res.*, 34, 2887–2905, 2006.

Friedberg I: The Interplay of Fold Recognition and Experimental Structure Determination in Structural Genomics. *Curr Opin Struct Biol.*, 14, 307-312, 2004.

Fritze I M, Linden L, Freigang J, Auerbach G, Huber R, Steinbacher S, The Crystal Structures of Zea mays and Arabidopsis 4-Hydroxyphenylpyruvate Dioxygenase. *Plant Physiology*, 134, 1388–1400, 2004.

Fujii H, Chinnusamy V, Rodrigues A, Rubio S, Antoni R, Park SY, Cutler SR, Sheen J, Rodriguez PL, Zhu JK, In vitro Reconstitution of An Abscisic Acid Signalling Pathway. *Nature*, 462, 660–664, 2009.

Gambetta GA, Matthews MA, Shaghasi TH, McElrone AJ, Castellarin SD. Sugar and Abscisic Acid Signaling Orthologs are Activated at the Onset of Ripening in Grape. *Planta*, 232, 219-34, 2010.

Garber K, New Drugs Target Hypoxia Response in Tumors. *Journal of the National Cancer Institute*, 97, 15, 2005.

Garcia I, Rodgers M, Lenne C, Rolland A, Sailland A, Matringe M. Sub Cellular Localization and Purification of a p-Hydroxyphenylpyruvate Dioxygenase from Cultured Carrot Cells and Characterization of the Corresponding cDNA. *Biochem J*, 325, 761– 769. 1997.

Garcia I, Job D, Matringe M. Inhibition of p-Hydroxyphenylpyruvate Dioxygenase by the Diketone nitrile of Isoxaflutole: A case of Half-Site Reactivity, *Biochemistry*, 39, 7501-7507, 2000.

Ginalski K, Comparative Modeling for Protein Structure Prediction. *Curr Opin Struct Biol.*, 16, 172-177, 2006.

Giovannoni J, Molecular Biology of Fruit Maturation and Ripening. *Annu Rev Plant Physiol Plant Mol Biol.*, 52, 725–749, 2001.

Goodman R M, Kishi Y, *J Am. Chem. Soc.*, 120, 9392, 1998.

Gleason C, Foley R C, Singh K B, Mutant Analysis in Arabidopsis Provides Insight into the Molecular Mode of Action of the Auxinic Herbicide Dicamba. *Plos one*; 6, 3, e17245, 2011.

Grossmann K, Mediation of Herbicide Effects by Hormone Interactions. *J Plant Growth Regul.*, 22, 109-22. 2003.

Hagen G, Guilfoyle T. Auxin-Responsive Gene Expression: Genes, Promoters and Regulatory Factors. *Plant Mol Biol.*, 49, 373-85, 2002.

Hagood E.S, Swann CW, Wilson HP, Ritter RL, Majek BA, Curran WS, Chandran R, Pest Mangement Guide: Field Crops, Grain crops, Soybeans and forages. Virginia coop. Ext serv. publ. Blacksberg: Virginia polytech, Inst and State Univ., 456-016, 2001

Haldimann P, Effects of Changes in Growth Temperature on Photosynthesis and Carotenoids Composition in *Zea mays* Leaves. *Physiol. Plant.*, 97, 554-562, 1996.

Hanhineva Kati, Kärenlampi Sirpa, Aharoni Asaph, Recent Advances in Strawberry Metabolomics. Institute of Public Health and Clinical Nutrition, Department of Clinical Nutrition, University of Eastern Finland, 2011.

Hansen H, Grossmann K. Auxin-Induced Ethylene Triggers Abscisic Acid Biosynthesis and Growth Inhibition. *Plant Physiol.*,124, 1437-48, 2000.

Hashimoto T, Perlot T, Rehman A, Trichereau, J, Ishiguro H, Paolino M, Sigl V, Hanada T, Hanada R, Lipinski S, Wild B, Camargo S M. R., Singer D, Richter A, Kuba K, Fukamizu A, Schreiber S, Clevers H, Verrey F, Rosenstiel P, Penninger J M. ACE2 Links Amino Acid Malnutrition to Microbial Ecology And Intestinal Inflammation. *Nature*, 487, 477-481, 2012.

Hause B, Weichert H, Hohne M, Kindl H, Feussner I, Expression of Cucumber Lipid-Body Lipxygenases in Transgenic Tobacco: Lipid-body Lipxygenase is Correctly Targeted to Seed Lipid Bodies. *Planta*, 210, 708-714, 2000.

Heap I, Herbicide Resistant Weeds, *Weed Sci Soc. of Amer.* Online, 2002.

Hegg EL, Que L: The 2-His-1-Carboxylate Facial Triad — an Emerging Structural Motif in Mononuclear Non-Heme Iron (II) Enzymes. *Eur J Biochem.*, 250, 625-629, 1997.

Higgins C F, ABC Transporters: Physiology, Structure and Mechanism-an Overview. *Res. Microbiol.*, 152, 205-210, 2001.

Hirayama T, Shinozaki K. Perception and Transduction of Abscisic acid Signals: Keys to the Function of the Versatile Plant Hormone ABA. *Trends Plant Sci.*, 12, 343–351, 2007.

Jamroz M, Kolinski A. Modeling of Loops in Proteins: a Multi-Method Approach. *BMC Structural Biology*, 10, 5, 2010.

Jia HF, Chai YM, Li CL, Lu D, Luo JJ, Qin L, Shen YY. Abscisic Acid Plays an Important Role in the Regulation of Strawberry Fruit Ripening. *Plant Physiology*, 157, 188–199, 2011.

Jiang Y, Joyce DC ABA Effects on Ethylene Production, PAL Activity, Anthocyanins and Phenolic Contents of Strawberry Fruit. *Plant Growth Regul.*, 39, 171–174, 2003.

Johnson-Winters K, Purpero VM, Kavana M, Nelson T, Moran GR. (4-Hydroxyphenyl)pyruvate Dioxygenase from *Streptomyces avermitilis*: The Basis for Ordered Substrate Addition. *Biochemistry*, 42, 2072–2080, 2003.

Jones R D, Summerville D A, Basolo F, *Chem. Rev.*, 79, 139, 1979.

Jones T A, Zhou J Y, Cowan SW, Kjeldgaard M, Improved Methods of Building Protein Models in Electron Density Maps and the Location Errors in These Methods. *Acta Crystallogr.*, A47, 110-119, 1991.

Kano Y, Asahira T Roles of Cytokinin and Abscisic Acid in the Maturing of Strawberry Fruits. *J Jpn Soc Hortic Sci.* 50, 31–36, 1981.

Kavana M, Moran G R, Interaction of (4-hydroxyphenyl) Pyruvate Dioxygenase with the Specific Inhibitor 2-[2-nitro-4-(trifluoromethyl) benzoyl]-1, 3-Cyclohexanedione, *Biochemistry*, 42, 10238-10245, 2003.

Kende H, Zeevaart J A D The Five “Classical” Plant Hormones. *The Plant Cell*, 9, 1197-1210, 1997.

Kenneth NS, Mudie S, van Uden P, Rocha S. SWI/SNF Regulates the Cellular Response to Hypoxia. *J Biol Chem.*, 284, 4123–4131, 2009.

Kepinski S, Leyser O, The Arabidopsis F-box Protein TIR1 is an Auxin Receptor. *Nature*, 435, 446-451. 2005.

Kiefer C, Hessel S, Lampert JM, Vogt K, Lederer MO, Breithaupt DE, von Lintig J, *J. Biol. Chem.*, 276, 14110-14116, 2001.

Kiser P D, Golczaka M, Lodowska D T, Chanceb M R, Palczewska K. Crystal Structure of Native RPE65, the Retinoid Isomerase of the Visual Cycle. *PNAS*, 106, 41, 17325–17330, 2009.

Kiser P D, Palczewski K, Membrane-Binding and Enzymatic Properties of RPE65. *Prog Retin Eye Res.*, 29, 5, 428–442, 2010.

Koehntop KD, Marimanikkuppam S, Ryle MJ, Hausinger RP, Que L Jr. Self Hydroxylation of taurine/alpha-Ketoglutarate Dioxygenase: Evidence for More than One Oxygen Activation Mechanism. *J Biol Inorg Chem.* 11, 1, 63-72, 2006.

Koga N, Koga R T, Liu G, Xiao R, Acton T B, Montelion G T, Baker D, Principles For Designing Ideal Protein Structures. *Nature*, 491, 222-228, 2012.

Koivunen P, Lee S, Duncan C G, Lopez G, Lu G, Ramkissoon S, Losman J A, Joensuu P, Bergmann U, Gross S, Travins JW, Samuel L, Ryan L, Keith L, Verhaak R G W, Yan Hai, Kaelin Jr, William G. Transformation by the (R)-Enantiomer of 2 –Hydroxyglutarate Llinked to EGLN Activation. *Nature*, 483, 484-487, 2012.

Kolinski A, Bujnicki JM: Generalized Protein Structure Prediction Based on Combination of Fold-Recognition With de novo Folding and Evaluation of Models. *Proteins*, 61, 84-90, 2005.

Kloer DP, Ruch S, Al-Babili S, Beyer P, Schulz GE, The Structure of a Retinal-Forming Carotenoid Oxygenase. *Science*, 308, 267–269, 2005.

Kloer DP, Schulz GE Structural and Biological Aspects of Carotenoid Cleavage. *Cell Mol Life Sci.*; 63, 2291–2303, 2006.

Kmiecik S, Gront D, Kolinski A: Towards the High-Resolution Protein Structure Prediction. Fast Refinement of Reduced Models With All-Atom Force Field. *BMC Struct. Biol.*, 7, 43, 2007.

Kraft M, Kuglitsch R, Kwiatkowski J, Frank M, Grossmann K. Indole-3-acetic acid and Auxin Herbicides Upregulate 9-cis-epoxycarotenoid Dioxygenase Gene Expression and Abscisic acid Accumulation in Cleavers (*Galium aparine*): Interaction With Ethylene. *J Exp Bot.*, 58, 1497-503, 2007.

Latrasse A.. Fruits III. In: Maarse H, editor. Volatile Compounds in Food and Beverages. *New York: Marcel Dekker*, 329-387, 1991.

Lau PW, Potter C S, Carragher B, MacRae I J, Structure of the Human Dicer-T Complex by Electron Microscopy. *Structure*, 17, 1326–1332, 2009.

Leone A, Bleve-Zacheo T, Gerardi, C, Mellilo M T, Leo L, Zacheo G. Lipoxygenase Involvement in Ripening Strawberry. Consiglio Nazionale delle Ricerche, Istituto di Scienze delle Produzioni Alimentari, 73100 Lecce, Italy, and Consiglio Nazionale delle Ricerche, Istituto per la Protezione delle Piante, 70126 Bari, Italy.

Li C, Wallace S, Polymer-Drug Conjugates: Recent Development in Clinical Oncology. *Adv Drug Deliv Rev.*, 60, 8, 886–898, 2008.

Lee D L, Prisbylla M P, Cromartie T H, Dagarin D P, Howard S W, Provan W M, Ellis M K, Fraser T, Mutter L C, The Discovery and Structural Requirements of Inhibitors of p-Hydroxyphenylpyruvate Dioxygenase, *Weed Sci.*, 45, 601-609, 1997.

Lee D L, Knudsen C G, Michaely W J, Chin H L, Nguyen N H, Carter C G, Cromartie T H, Lake B H, Shribbs J M, Fraser T. The Structure-Activity Relationships of the Triketone Class of p-Hydroxyphenylpyruvate Dioxygenase Inhibiting Herbicides, *Pestic. Sci.*, 54, 377-384, 1998.

- Lee C W, Vitriol EA, Shim S, Wise A L, Velayutham R P, Zheng J Q, Dynamic Localization of G-Actin during Membrane Protrusion in Neuronal Motility. *Current Biology*, 23, 12, 1046–1056, 2013.
- Leung J, Controlling Hormone Action by Subversion and Deception, *Plant Science, Science*, 335, 6064, 46-47, 2012
- Lietha D, Eck M J, Crystal Structures of the FAK Kinase in Complex with TAE226 and Related Bis-Anilino Pyrimidine Inhibitors Reveal a Helical DFG Conformation. *Plos one.*, 3, 11, e3800, 2008.
- Lock E A, Ellis M K, Gaskin P, Robinson M, Auton T R, Provan W M, Smith L L, Prisbylla M P, Mutter L C Lee D L From Toxicological Problem to Therapeutic use: The Discovery of the Mode of Action of 2-(2-nitro-4 trifluoromethylbenzoyl)- 1,3-Cyclohexanedione (NTBC), its Toxicology and Development as a Drug, *J. Inherited Metab Dis.*, 21, 498-506, 1998.
- Lloyd MD, Merritt KD, Lee V, Sewell TJ, Whason B, Baldwin JE, Schofield CJ, Elson SW, Baggaley KH, Nicholson NH: Productsubstrate Engineering by Bacteria: Studies on Clavamate Synthase, a Trifunctional Dioxygenase. *Tetrahedron*, 55, 10201-10220, 1999.
- Li C, Jia H, Chai Y, Shen Y, Abscisic acid Perception and Signaling Transduction in Strawberry. A Model for Non-Climacteric Fruit Ripening. *Plant Signaling & Behavior*, 6, 12, 1950-1953, 2011.
- Li H, Ren Z, Kang X, Zhang L, Li X, Wang Y, Xue T, Shen Y, Liu Y. Identification of Tyrosine—Phosphorylated Proteins Associated With Metastasis and Functional Analysis of FER in Human Hepatocellular Carcinoma cells. *BMC Cancer*, 9, 366, 1-16, 2009.

Liu A, Ho R Y N, Que Jr. L. Alternative Reactivity of an α -Ketoglutarate-Dependent Iron(II) Oxygenase: Enzyme Self-Hydroxylation. *J. Am. Chem. Soc.*, 123, 5126-5127, 2001.

MacRae IJ, Zhou K, Li F, Repic A, Brooks A N, Cande W Z, Adams P D, Doudna JA, Structural Basis for Double-Stranded RNA Processing by Dicer. *Science*, 311, 195-197, 2006.

Ma Y, Szostkiewicz I, Korte A, Moes D, Yang Y, Christmann A, Grill E Regulators of PP2C Phosphatase Activity Function as Abscisic acid Sensors. *Science*, 324, 1064–1068, 2009.

Manning K, Changes in Gene Expression During Strawberry Fruit Ripening and Their Regulation by Auxin. *Planta*, 94, 62–68, 1994.

Massa G D, Santini J B, Mitchell C A, Minimizing Energy Utilization for Growing Strawberries During Long-Duration Space Habitation. Agriculture News Page. Purdue University.2010.

Massey V, *J. Biol. Chem.*, 269, 2245, 1994.

McCapra F, *Acc. Chem. Res.*, 9, 201, 1976.

Meazza G, Scheffler B E, Tellez M R, Rimando A M, Romagni J G, Duke S O, Nanayakkara D, Khan I A, Abourashed E A, Dayan F E, The Inhibitory Activity of Natural Products on Plant p-Hydroxyphenylpyruvate Dioxygenase, *Phytochemistry*, 60, 281-288, 2002.

Messing SAJ, Gabelli S B, Echeverria I, Vogel J T, Guan JC, Tan B C, Klee HJ, McCarty D R, Amzela L M, Structural Insights into Maize Viviparous14, a Key Enzyme in the Biosynthesis of the Phytohormone Abscisic Acid. *The Plant Cell*, 22, 2970–2980, 2010.

Mitchell WC, Jelenkovic G, Characterizing NAD- and NADP- Dependent Alcohol Dehydrogenase Enzymes of Strawberries. *J Amer Soc Hort Sci.*, 120, 798-801, 1995.

- Mortenson A, Skibsted L H, Importance of Carotenoids Structure in Radical Scavenging Reactions, *J. Agric. Food Chem.*,45, 2970-2977, 1997.
- Moult J, Fidelis K, Kryshchak A, Rost B, Hubbard T, Tramontano A: Critical Assessment of Methods of Protein Structure Prediction-Round VII. *Proteins*, 69, 3-9, 2007.
- Nakanishi K, Weinberg D E, Bartel D P, Patel D J. Structure of Yeast Argonaute with Guide RNA. *Nature*, 486, 368-374, 2012.
- Nambara E, Marion-Poll A. Abscisic acid Biosynthesis and Catabolism. *Annu Rev Plant Biol.*, 56, 165–185, 2005.
- Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, et al. Systematic Discovery of New Recognition Peptides Mediating Protein Interaction Networks. *PLoS Biol.*, 3, e405, 2005.
- Neidig M L, Kavana M, Moran G R, Solomon E I, CD and MCD Studies of the Non-Heme Ferrous Active Site in (4-Hydroxyphenyl)pyruvate Dioxygenase: Correlation between Oxygen Activation in the Extradiol and R-kg-dependent Dioxygenases,. *J. Am. Chem. Soc.*, 126, 4486-4487, 2004.
- Norris SR, Shen X, Della Penna D, Complementation of The Arabidopsis pds1mutant with the Gene Encoding p-Hydroxyphenylpyruvate Dioxygenase. *Plant Physiol.*, 117, 1317-1323, 1998.
- Olias JM, Sanz C, Rios JJ, Perez AG. Substrate Specificity of Alcohol Acyltransferase from Strawberry and Banana Fruit. In: Rouseff RL, Editor. Fruit Flavors. Washington: *American Chemical Society*, 134-139,1995.
- Ort D R, Energy Transduction in Oxygenic Photosynthesis, An Overview of Structure and Mechanism. In: Staeheli L A, Arntzen C J (Eds.), *Encyclopedia of Plant Physiology*. New Series, vol.9. Springer, New York, 142-196, 1986.

Osanai T, Imashimizu M, Seki A, Sato S, Tabata S, Imamura S, et al. CHLH, the H subunit of the Mg-chelatase, is an Anti-sigma Factor for SigE in *Synechocystis* sp PCC 6803. *Proc Natl Acad Sci., USA*; 106, 6860-5, 2009.

Pallett K E, Little J P, Sheekey M, Veerasekaran P, The Mode of Action of Isoxaflutole I. Physiological Effects, Metabolism and Selectivity, *Pestic Biochem Physiol.*, 62, 113-124, 1998.

Pallett K E, Cramp S M, Little J P, Veerasekaran P, Crudace A J, Slater A E, Isoxaflutole: the Background to its Discovery and the Basis of its Herbicidal Properties, *Pest Manage. Sci.*, 57, 133-142, 2001.

Pan QH, Li MJ, Peng CC, Zhang N, Zou X, Zou KQ, Wang XL, Yu XC, Wang XF, Zhang DP Abscisic acid Activates Acid Invertases in Developing Grape Berry. *Plant. Physiol.*, 125, 157–170, 2005.

Pandey AK, Somvanshi S, Singh V P. Focal Adhesion Kinase: An Old Protein with New Roles. *On Line Journal of Biological Sciences*, 12, 1, 11-14, 2012.

Pandey S, Nelson DC, Assmann SM Two Novel GPCR-type G Proteins are Abscisic Acid Receptors in Arabidopsis. *Cell*, 136, 136–148, 2009.

Park SY, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow TF, et al. Abscisic Acid Inhibits Type 2C Protein Phosphatases via the PYR/PYL Family of START Proteins. *Science*, 324, 1068–1071, 2009.

Perkins-Veazie P, Growth and Ripening of Strawberry Fruit. *Hortic Rev. Am Soc Hortic Sci.*, 17, 267–297, 1995.

Perez AG, Rios JJ, Sanz C, Olias JM. Aroma Components and Free Amino Acids in Strawberry Variety Chandler During Ripening. *J Agric Food Chem.*, 40, 2232-2235, 1992.

Perez AG, Sanz C, Olias J.M.. Partial Purification and Some Properties of Alcohol Acyltransferase from Strawberry Fruits. *J Agric Food Chem.*, 41, 1462-1466, 1993.

- Perez AG, Sanz C, Olias R, Rios JJ, Olias J M, Evolution of Strawberry Alcohol Acyltransferase Activity during Fruit Development and Storage. *J Agric Food Chem.*, 44, 3286-3290, 1996a.
- Petsalaki E, Starka A, Urdialesb E G, Russell, R B. Accurate Prediction of Peptide Binding Sites on Protein Surfaces. *PLoS Computational Biology.*, 5, 3 , e1000335, 2009.
- Pfuhl M. Pastore A. Tertiary Structure of an Immunoglobulin-Like Domain from the Giant Muscle Protein Titin: A New Member of the I Set. *Curr Biol.*, 3, 391-401, 1995.
- Pokarowski P, Kloczkowski A, Jernigan RL, Kothari NS, Pokarowska M, Kolinski A: Inferring Ideal Amino Acid Interaction Forms from Statistical Protein Contact Potentials. *Proteins*, 59, 49-57, 2005.
- Procko E, O'Mara ML, Bennett WF, Tieleman DP, Gaudet R, The Mechanism of ABC Transporters: General Lessons From Structural and Functional Studies of an Antigenic Peptide Transporter. *FASEB J.*, 23, 5, 1287-302, 2009.
- Qin H, Chen F, Huan X, Machida S, Song J, Yuan YA. Structure of the Arabidopsis thaliana DCL4 DUF283 Domain Reveals a Noncanonical Double-Stranded RNA-binding Fold for Protein–Protein Interaction. *RNA*, 16, 474-481, 2010.
- Qin X, Zeevaart JAD. The 9-cis-epoxycarotenoid Cleavage Reaction is the Key Regulatory Step of Abscisic Acid Biosynthesis in Water-Stressed Bean. *Proc Natl Acad Sci. USA.*, 96, 15354–15361, 1999.
- Que L Jr., Ho R Y N *Chem. Rev.*, 96, 2607,1996.
- Raghavan C, Ong EK, Dalling MJ, Stevenson TW Regulation of Genes Associated with Auxin, Ethylene and ABA Pathways by 2,4-Dichlorophenoxyacetic Acid in Arabidopsis. *Funct. Integr. Genomic*, 6, 60–7,. 2006.
- Ramachandran GN, Sasisekharan V. Conformation of Polypeptides and Proteins. *Adv Protein Chem.* 23, 283–437, 1968.

Rangel M, Machado O L, da Cunha, M, Jacinto T, Accumulation of Chloroplast-Targeted Lipoxygenases in Passion Fruit Leaves in Response to Methyl Jasmonate. *Phytochemistry.*, 60, 619-625, 2002.

Rangarajan E S, Izard T. Dimer Asymmetry Defines α -catenin Interactions. *Nature Structural & Molecular Biology. Advance online publication.* 2013

Rayment L, Smith C, Yount RG, The Active Site of Myosin. *Annual Review of Physiology*, 58, 671-702, 1996.

Rohl CA, Strauss CEM, Misura KMS, Baker D: Protein Structure Prediction Using Rosetta. *Numerical Computer Methods, Part D of Methods Enzymol.* Academic Press Brand L, Johnson ML; 383, 66-93, 2004.

Ryle M J, Hausinger R P. Non-heme Iron Oxygenases, *Curr Opin Chem Biol.*, 6, 193-201, 2002.

Sali A, Blundell TL: Comparative Protein Modeling by Satisfaction of Spatial Restraints. *J Mol Biol.*, 234,779-815, 1993.

Sanz Y, Toldra F, Renault P, Poolman B. Specificity of the Second Binding Protein of the Peptide ABC-transporter (Dpp) of *Lactococcus lactis* IL 1403. *FEMS Mitochondrial Lett.:* 10, 227,1,33-38, 2013

Sawyer DT, Valentine J S *Acc. Chem. Res.*, 14, 393, 1981.

Schaller M D. Cellular Functions of FAK Kinases: Insight into Molecular Mechanisms and Novel Functions. *Journal of Cell Science*, 123, 1007-1013, 2010.

Schofield CJ, Zhang Z, Structural and Mechanistic Studies on 2-Oxoglutarate-Dependent Oxygenases and Related Enzymes. *Curr Opin Struct Biol.*, 9, 722–731, 1999.

Schreiber S L, *J. Am. Chem. Soc.*, 102, 6163, 1980.

Sergeant M J, Li JJ, Fox C, Brookbank N, Rea D, Bugg T D H, Thompson A J, Selective Inhibition of Carotenoid Cleavage Dioxygenases *Phenotypic Effects on Shoot Branching. The Journal of Biological Chemistry*, 284, 8, 5257–5264, 2009.

Shang Y, Yan L, Liu ZQ, Cao Z, Mei C, Xin Q, Wu FQ, Wang XF, Du SY, Jiang T, et al. The Mg-chelatase H subunit of Arabidopsis Antagonizes a Group of WRKY Transcription Repressors to Relieve ABA Responsive Genes of Inhibition. *Plant Cell*, 22, 1909–1935, 2010.

Siddall T L, Ouse D G, Benko Z L, Garvin GM, Jackson J L, McQuiston J M, Ricks M J, Thibault T D, Turner J A, Vanheertum J C, Weimer M R. Synthesis and Herbicidal Activity of Phenyl-Substituted Benzoylpyrazoles, *Pest Manage. Sci.*, 58, 1175-1186, 2003.

Sono M, Roach M P, Coulter E D, Dawson J H, *Chem. Rev.*, 96, 2841, 1996.

Soon FF, Ng, LM, Zhou X., Edward W, Graham M, Kovach A, Tan S, Roanald E. Dioxygen and Hemerythrin, *Chem. Rev.*, 94, 3, 715-726, 1994.

Sun Y, Campisi J, Higano C, Beer T M, Porter P, Coleman I, True L, Nelson P S. Treatment-Induced Damage to the Tumor Microenvironment Promotes Prostate Cancer Therapy Resistance Through WNT16B. *Nature Medicine*, 2012.

Szabados L, Savoure A. Proline: a Multifunctional Amino Acid. *Trends in Plant Science*, 15, 2, 89-97, 2009.

Tan BC, Schwartz SH, Zeevaart JA, McCarty DR Genetic Control of Abscisic Acid Biosynthesis in Maize. *Proc Natl Acad Sci. USA*, 94, 12235–12240, 1997.

Tan X, Calderon-Villalobos LI, Sharon M, Zheng C, Robinson CV, Estelle M, Zheng N. Mechanism of Auxin Perception by the TIR1 Ubiquitin Ligase. *Nature*, 446,640-5, 2007.

Taylor IB, Mulholland BJ, Jackson AC, McKee JM, Hilton HW, Symonds RC, Sonneveld T, Burbidge A, Stevenson P, Taylor IB. Regulation and Manipulation of the Biosynthesis

of Abscisic Acid, Including the Supply of Xanthophylls Precursors. *J Plant Growth Regul.*, 24, 253-73, 2005.

Thornalley P. Game, Set and Match to Strawberries: The Superfruit. University of Warwick (2012, July 4). *Science Daily*. Retrieved July 30, 2013, from <http://www.sciencedaily.com/releases/2012/07/120704124107.htm>

Trabanger T J, Franceschi VR, Hildebrand D F, Grimes H D, The Soybean 94-Kilodalton Vegetative Storage Protein is a Lipoxygenase that is Localized in Paravenial Mesophyll Cell Vacuoles. *Plant Cell*, 3, 973-987, 1991.

Varkuti BH, Yang Z, Kintses B, Erdelyi P, Bardos-Nagy I, Kovacs A L, Hari P, Kellermayer M, Vellai T, Malnasi-Csizmadia A, A Novel Actin Binding Site of Myosin Required for Effective Muscle Contraction. *Nature Structural and Molecular Biology*, 19, 3, 299-306, 2012.

Vianello A, Braidot E, Bassi G, Macri F, Lipoxygenase Activity on the Plasmalemma of Sunflower Protoplasts and its Modulation. *Biochim Biophys Acta.*, 1255, 57-62, 1995.

Viviani F, Little JP, Pallett KE, The Mode of Action of Isoxaflutole. II. Characterization of the Inhibition of Carrot 4-Hydroxyphenylpyruvate Dioxygenase by the Diketone nitrile Derivative of RPA 201772. *Pestic Biochem Physiol.*, 62, 125-134, 1998.

von Lintig J, Vogt K, Filling the Gap in Vitamin A Research. Molecular Identification of an Enzyme Cleaving Beta-Carotene to Retinal. *J. Biol. Chem.*, 275, 11915–11920, 2000.

von Lintig J, Dreher A, Kiefer C, Wernet M F, Vogt K, Analysis of the Blind Drosophila Mutant *ninaB* Identifies the Gene Encoding the Key Enzyme for Vitamin A Formation *in vivo*. *Proc. Natl. Acad. Sci. USA*, 98, 1130–1135, 2001.

Waltz E, Glyphosate Resistance Threatens Roundup Hegemony. *Nature Biotechnology*, 28, 6, 537-538, 2010.

- Wang S Y, Lewers K S, Antioxidant Capacity and Flavonoid Content in Wild Strawberries. *J. Amer. Soc. Hort. Sci.*, 132, 5 , 629-637, 2007.
- Wang XF, Zhang DP. Abscisic Acid Receptors: Multiple Signal Perception Sites. *Ann Bot. (Lond)*; 101, 311–317, 2008.
- Winterhalter K H, *Chimia*, 30, 9, 1976.
- Wise RR, Cook WB, Development of Ultra Structural Damage to Chloroplasts in a Plastoquinone-Deficient Mutant of Maize. *Environ. and Exp. Bot.*; 40, 221-228, 1998.
- Witze E S, Old W M, Resing K A, Ahn N G. Mapping Protein Post-Translational Modifications with Mass Spectrometry. *Nature Methods*, 4, 798 – 806, 2007.
- Wostenberg C, Lary J W, Sahu D, Acevedo R, Quarles K A, Cole J L, Showalter S A, The Role of Human Dicer-dsRBD in Processing Small Regulatory RNAs. *Plos one*, 7, 12, e51829, 2012.
- Wong LL: Cytochrome P450 Monooxygenases. *Curr Opin Chem Biol.*, 2, 263-268, 1998.
- Wu FQ, XinQ, Cao Z, Liu ZQ, Du SY, Mei C, Zhao CX, Wang XF, Shang Y, Jiang T, et al The Magnesium-Chelatase H Subunit Binds Abscisic Acid and Functions in Abscisic Acid Signaling: New Evidence in Arabidopsis. *Plant Physiol.*, 150, 1940–1954, 2009.
- Wu HY, Tseng V S, Chen LC, Chang, HY, Chuang, IC, Tsay YG, Liao PC. Identification of Tyrosine-Phosphorylated Proteins Associated with Lung Cancer Metastasis using Label-Free Quantitative Analyses. *Journal of Proteome Research*, 9, 4102–4112, 2010.
- Xie C, Zhu J, Chen X, Mi L, Nishida N, Springer TA, Structure of An Integrin with an α I Domain, Complement Receptor Type 4. *EMBO J.*, 29, 666-679, 2010.
- Yang C, Pflugrath J W, Camper D L, Foster M L, Pernich D J, Walsh T A. Structural Basis for Herbicidal Inhibitor Selectivity Revealed by Comparison of Crystal Structures of Plant

and Mammalian 4-Hydroxyphenylpyruvate Dioxygenases. *Biochemistry*, 43, 10414-10423, 2004.

Yoneda J, Andou A, Takehana K, Regulatory Roles of Amino Acids in Immune Response. *Current Rheumatology Reviews*, 5, 252-258, 2009.

Young A J, The Photoprotective Role of Carotenoids in Higher Plants. *Plant Physiol.*; 83, 702-708, 1991.

Young VR, Ajami AM., Glutamate: An Amino Acid of Particular Distinction. *Journal of Nutrition*, 130, 892S–900S. 2000

Zhang Y, Kang S A, Mukherjee T, Bale S, Crane B R. Begley T P, Ealick S E, Crystal Structure and Mechanism of Tryptophan 2,3-Dioxygenase, a Heme Enzyme Involved in Tryptophan Catabolism and in Quinolate Biosynthesis. *Biochemistry*, 46, 145-155, 2007.

Zhang Y, Xu W, Li Z, Deng X W, Wu W, Xue Y. F-Box Protein DOR Functions as a Novel Inhibitory Factor for Abscisic Acid-Induced Stomatal Closure under Drought Stress in Arabidopsis. *Plant Physiology*, 148, 2121–2133, 2008.

Zhang M, Leng P, Zhang G, Li X, Cloning and Functional Analysis of 9-cis-epoxycarotenoid Dioxygenase (NCED) Genes Encoding a Key Enzyme during Abscisic Acid Biosynthesis from Peach and Grape Fruits. *J Plant Physiol.*, 166, 1241–1252, 2009a.

Zhou G, Golden T, Aragon I V, Honkanen R E, Ser/Thr Protein Phosphatase 5 Inactivates Hypoxia-induced Activation of an Apoptosis Signal-regulating Kinase 1/MKK-4/JNK Signaling Cascade, *The Journal of Biological Chemistry*, 279, 45, 46595–46605, 2004.