

**Genome-wide analysis of 3D-Domain swapped predicted protein in
*"Sesamum indicum"***

A dissertation
Submitted in partial fulfilment of the requirement
For the award of a degree of
Masters of Technology
in
Biotechnology
(June 2022)

Under the guidance of
Dr. Atul Kumar Upadhyay
Assistant Professor

Submitted by
Payal Singh Kashyap
Roll No. 602004016



THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

DEPARTMENT OF BIOTECHNOLOGY
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY,
PATIALA

ACKNOWLEDGEMENTS

I express my special gratitude to HOD, Dr. M S Reddy, Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, for his kind suggestions and foresightedness. With great reverence, I express my warmest feeling with deep gratitude to Dr. Atul Kumar Upadhyay, Assistant Professor, TIET, who agreed to take upon and guide this dissertation & training, and made this possible. I am also thankful to Tanmayee Basu, Ph.D. scholar (TIET), and Arshwinder Singh for their guidance during the whole dissertation work. Ultimately, I would like to thank my family and friends for giving me the courage to keep going.

DECLARATION

I, at this moment, declare that the project work entitled ("Genome-wide analysis of 3D-Domain swapped predicted protein sequences in *Sesamum indicum*") is an authentic record of my work carried out at Thapar Institute of Engineering and Technology as requirements of 12 months dissertation for the award of the degree of Masters of Technology in Biotechnology, under the guidance of Dr. Atul Kumar Upadhyay, during July 2021 to June 2022.



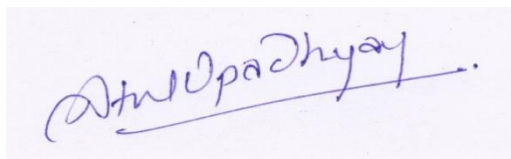
(Signature of student)

Payal Singh Kashyap
M.Tech, Biotechnology
602004016

Date: 28/07/2022

CERTIFICATE

This is to certify that the thesis entitled “**Genome-wide analysis of 3D-Domain swapped predicted protein sequences in *Sesamum indicum***” and submitted by **PAYAL SINGH KASHYAP**, Roll No. **602004016** for the award of M.Tech (Biotechnology) of Thapar Institute of Engineering & Technology, Patiala embodies original work done by her under my supervision. This dissertation has not been submitted to award of any degree or certificate in any other university/institute.



Signature of the supervisor
ATUL KUMAR UPADHYAY
Assistant Professor
Department of Biotechnology

Date: 28/07/2022

CONTENTS

Chapters	Contents	Page No
	Abstract	1
Chapter 1	Introduction	2-5
	1.1 Research Gap	4
	1.2 Aim of the study	5
Chapter 2	Review of Literature	6- 11
	2.1 <i>Sesamum indicum</i> : Evolutionary History, Genome analysis, and Medicinal Uses	6
	2.1.1 Evolutionary History	6
	2.1.2 Genome sequencing of the critical oilseed crop <i>Sesamum indicum L</i>	7
	2.1.3 Characterization of Thioesterase Gene in <i>S. indicum</i> using Bioinformatics tools	7
	2.1.4 Genome Analysis in Abiotic stress	7
	2.1.5 Medicinal Uses	8
Chapter 3	Materials and methods	12-18
	3.1 Protein sequence resources	12
	3.2 Identification of 3D-Domain swapping	12
	3.3 Proteomes in <i>S. indicum</i>	12
	3.4 Homology searches & Structure selection	13
	3.5 Hinge Identification	14
	3.5.1 HingeProt	14
	3.5.2 PACKMAN- Packing and Motion analysis	14

	3.5.3 PyMol	15
	3.6 Molecular Dynamic Simulation	15
	3.6.1 Molecular Dynamics Simulation Steps	15
Chapter 4	Results	19-34
	4.1 Protein sequence resource	19
	4.2 Identification of 3D-Domain swapping	21
	4.3 Homology searches & Structure selection	23
	4.3.1 PDB Structures of 2F1K: Crystal structure of <i>Synechocystis aerogenate</i> dehydrogenase	23
	4.4 Hinge Prediction	23
	4.5 PyMol Visualization of Hinges	26
	4.6 Molecular Dynamics Simulation Studies	28
	4.6.1 Root mean square deviation	28
	4.6.2 Radius of Gyration	30
	4.6.3 Root mean square fluctuation	32
	4.6.4 Solvent accessible surface area	34
Chapter 5	Discussions	36
Chapter 6	Conclusions	37
Chapter 7	References	38
	Plagiarism report	43

LIST OF FIGURES

S. No.	Title	Page No
1.	Structural Representation of 3D-Domain swapping	3
2.	PDB structure of “Crystal structure of <i>Synechocystis</i> aerogenate dehydrogenase: 2F1K dimer	23
3.	PDB structure of “Crystal structure of <i>Synechocystis</i> aerogenate dehydrogenase: 2F1K monomer	23
4.	PyMol visualization of hinge region (HingeProt) in 2F1K dimer	26
5.	PyMol visualization of hinge region (HingeProt) in 2F1K monomer	26
6.	PyMol visualization of hinge region (PACKMAN) in 2F1K dimer	27
7.	PyMol visualization of the hinge region (PACKMAN) in 2F1K monomer	27
8.	RMSD analysis in 2F1K dimer for 30 ns	28
9.	RMSD analysis in 2F1K monomer for 30 ns	29
10.	Radius of Gyration analysis in 2F1K dimer	30
11.	Radius of Gyration analysis in 2F1K monomer	31
12.	RMSF analysis in 2F1K dimer	32
13.	RMSF analysis in 2F1K monomer	33
14.	SASA analysis in 2F1K dimer	34
15.	SASA analysis in 2F1K monomer	35

LIST OF TABLES

S. No.	Title	Page No.
1.	Classification of 3D-Domain swapping	8
2.	List of plants previously studied in 3D-Domain swapping	10
3.	Reviewed sequences in the Proteome of <i>Sesamum indicum</i> in UniProt Database	19
4.	3D-Domain swapped cases in reviewed sequences in <i>Sesamum indicum</i>	21
5.	HingeProt showing hinge residues in the slowest two modes of GNM in 2F1K Dimer	24
6.	HingeProt showing hinge residues in the slowest two modes of GNM in 2F1K monomer	24
7.	PACKMAN showing hinge residues in 2F1K dimer	25
8.	PACKMAN showing hinge residues in 2F1K monomer	25

LIST OF ABBREVIATIONS USED

SGWG	Sesame Genome Working Group
RFO	Raffinose family oligosaccharides
NCBI	National Center for Biotechnology Information
BLAST P	Basic Local Alignment Search Tool, Protein
PDB	Protein Data Bank
ENM	Elastic Network Model
GNM	Gaussian Network Model
ANM	Anisotropic Network models
MD	Molecular Dynamics
GROMACS	GRoningen MACHine for Chemical Simulations
CHARMM36	Chemistry at Harvard Macromolecular Mechanics
PACKMAN	Packing and Motion Analysis
RMSD	Root Mean Square Deviation
RG	Radius of Gyration
RMSF	Root Mean Square Fluctuation
SASA	Solvent Accessible Surface Area

ABSTRACT

3D-Domain swapping is a naturally occurring phenomenon found in every organism, including plants. It is a process of protein oligomerization usually seen in dimers or higher oligomers where similar domains or structural element gets exchanged with the help of swapped and hinge region, which is also considered the most flexible part of the protein and acquires functional importance. When these structural domains gets switched, the appearance of a "Domain swapped Interface" indicates that the structure is engaged in the domain swapping. This dissertation examine predictions of 3D-domain swapped cases and hinge region involved in the 3D-Domain swapping in the proteome of "Sesamum indicum". The comparison of the hinge fluctuations through molecular dynamics simulation was performed using the GROMACS 2020.1 software. Analysis plots against Root Mean Square Deviation (RMSD), the Radius of Gyration, Root Mean Square Fluctuation (RMSF), and Solvent Accessible Surface Analysis (SASA) were carried out.

A recent analysis of proteins shows that approximately 35% of the proteins in a cell are oligomeric (*Goodsell & Olson, 2000*). Because of the advantages, protein oligomers have over their monomers, such as the increased likelihood of allosteric control, the emergence of new active sites at oligomer subunit interfaces, and the expansion in the local concentration of active sites in oligomeric structures as compared to monomers, which have the potential to reserve larger binding surfaces, protein oligomers have emerged (*Upadhyay & Sowdhamini, 2019*).

One of the protein oligomerization mechanisms is 3D-Domain swapping (*Bennett, Schlunegger & Eisenberg, 1995*). 3D-Domain swapping is a process of protein oligomerization where two or more protein chains form dimer or higher oligomers by exchanging an identical structural element between the monomers (*Upadhyay & Sowdhamini, 2019*). 3D-Domain swapping is mainly expedited by some parts of the polypeptide that is immersed in inter-chain interactions; the hinge region and swapped regions. A hinge region can be defined as a short stretch of amino acids that connects the swapped region and the remaining structural core of the protein. A swapped region is the structural segment following the hinge region that the other chain shares (*Shameer et al., 2011*).

3D-Domain swapping was reported for the first time in the structure of Diphtheria toxin in 1992, even though the prediction of a similar structural phenomenon was also seen over four decades earlier with dimers of ribonuclease (RNase) in 1962. The mechanism has gained much attention in neurodegenerative diseases (Alzheimer's) and prions because of its role in functional regulation, the evolution of higher oligomers, protein misfolding, etc. (*Shameer et al., 2010*).

The phenomenon of 3D-Domain swapping is also seen in several cases of plants, such as bleomycin resistance, the light-harvesting complex of photosystem II in *Ocimum tenuiflorum* found to be involved in many diverse functions (*Upadhyay & Sowdhamini, 2019*).

3D-Domain swapping is a protein oligomerization process in which the monomer's subunits get interchanged with the help of Hinges to form either a Dimer or any higher oligomer.

From the below (Figure 1), it is visible that there is a monomeric structure with its domains linked with the Hinges. When the Domains or structural element gets exchanged with each other, it is called "Swapped Domain" in the presence of Domain swapped interface and is said to be involved in 3D-Domain swapping. On the other side, if the Domains or structural elements are not getting exchanged, then that is the case of "Non-Domain swapping."

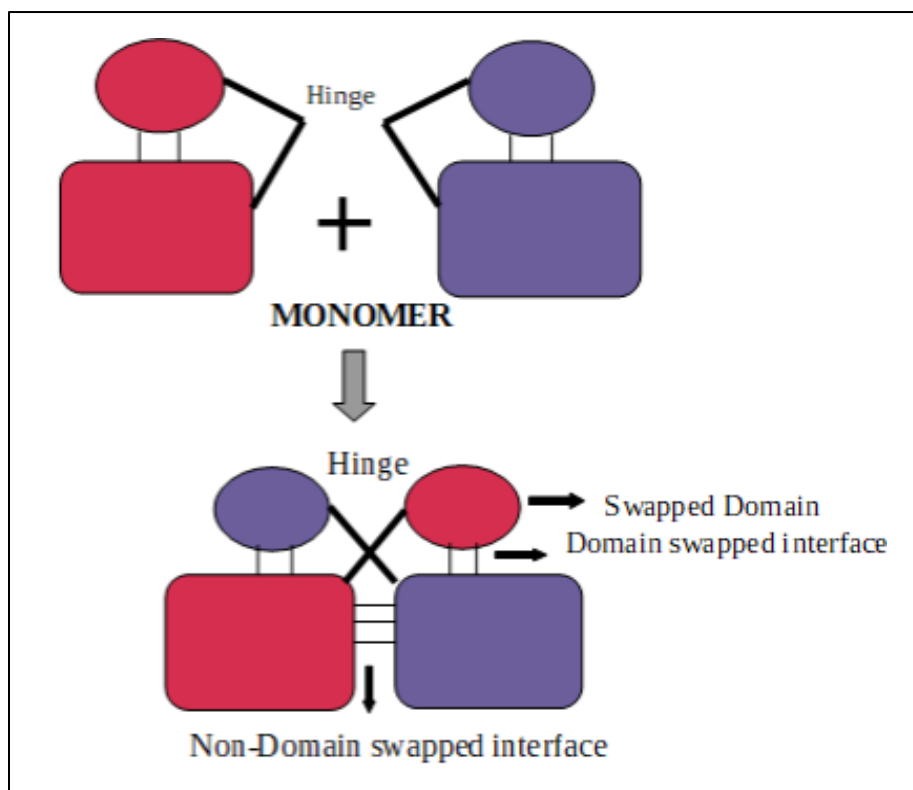


Figure 1: Structural Representation of 3D-Domain Swapping

1.1 Research Gap

Many studies have explored the aspects of 3D-Domain swapping in plants, such as the light-harvesting complex of photosystem II, bleomycin resistance, and SET domain protein methyl-transferase (Upadhyay & Sowdhamini, 2019). Recent studies have shown the 3D Domain Swapping Dimerization of the Receiver Domain of Cytokinin Receptor CRE1 From *Arabidopsis thaliana* and *Medicago truncatula* (Tran, Urbanowicz, Jasiński, Jaskolski & Ruszkowski, 2021). Prior genetic work has been carried out in the whole genome of *Ocimum tenuiflorum* (holy basil) to study the domain-swapped sequences involved in the plant's genome. Approximately 25% of the proteins of *Ocimum tenuiflorum* were anticipated to undergo 3D-Domain swapping, and 12% of the sequences engaged in diverse roles in plants, such as the Defence mechanism of plants in biotic and abiotic stress (Upadhyay & Sowdhamini, 2019). In the present study, the proteome of “*Sesamum indicum*” has been selected, and tried to compare hinges of the monomeric and dimeric structures of the proteome using Molecular Dynamics simulation.

1.2 Aim of the Study

- Prediction of 3D-Domain swapping cases in the proteome of *Sesamum indicum*
- To predict the hinge region in swapped proteins.
- To validate the predictions of 3D-Domain swapping and respective hinge by molecular dynamics simulation studies

The phenomenon of 3D-Domain swapping is the exchange of structural domains between molecules of the same protein to form dimer or higher oligomers (*Rousseau, Schymkowitz & Itzhaki, 2012*). The unit structure of domain-swapped oligomers is constitutive identical to the corresponding monomer structure, excluding the hinge region that links the interchanged regions (*Hirota, Mashima & Kobayashi, 2021*).

In 1992, 3D-Domain swapped structure was reported in the diphtheria toxin for the first time (*Upadhyay & Sowdhamini, 2019*). After that, 3D-Domain Swapping has been reported to occur in many natural proteins and is involved in protein deposition diseases (*Bennett, Sawaya & Eisenberg, 2006*). Since then, several proteins have been known to display 3D-Domain Swapping, increasing. Over 600 3D-Domain Swapping structures have been submitted to the Protein Data Bank (PDB) (*Hirota, Mashima & Kobayashi, 2021*). Most swapped domains are at either the N- or C-terminus. The swapped domains are diverse in their primary and secondary structures (*Liu & Eisenberg, 2002*), indicating that almost any protein may undergo 3D-Domain Swapping.

2.1 *Sesamum indicum*: Evolutionary History, Genome analysis, and Medicinal Uses

2.1.1 Evolutionary History

Sesamum indicum commonly called sesame, which is an annual plant of the family "Pedaliaceae." It is one of the aged oilseed crops extensively grown in Asia for its superior nutritional seeds (*Dossa et al., 2017*) Due to its potency of resistance to oxidation and rancidity, it is also known as the "Queen of Oilseeds" (*Miraj & Kiani, 2016*)

Research history followed three significant periods:

- 1. Germplasm collection and genebank constitution era (<2000):** In the first era before 2000, the collection of genetic materials of cultivated sesame and related wild species were taken from many growing areas, they were morphologically characterized, and different seedbanks were set up in several countries (*Dossa et al., 2017*).

- 2. Classical breeding and genetics era (2000-2013):** The second era in *S. Indicum* research (2000–2013) was distinguished by implementing classical breeding methods. This era involved some attainments like the interspecific crossing, the development of molecular markers such as AFLP, RAPD, SSR, etc., studies of Genetic Diversity, marker-assisted Breeding, and study of sesamum oil nutritional and pharmaceutical properties (Dossa et al., 2017).
- 3. Omics era (>2013):** With the completion of the final era after 2013, numerous research studies have been generated like Chloroplast & Nuclear genome sequencing, RNA sequencing, Genome sequencing, online database construction, Association mapping, Gene family functional study, ultra-dense genetic maps, development of informative molecular markers such as SNPs, etc. which is being used for the improvement of *S. indicum* (Dossa et al., 2017).

2.1.2 Genome sequencing of the critical oilseed crop *Sesamum indicum* L

The Sesame Genome Working Group (SGWG) was created to assemble the sesame (*S. indicum* L.) The main Aim of the Sesame Genome Project is to furnish a fine map of *S. indicum* and aid the learning of global and functional genomic studies (Zhang et al., 2013)

2.1.3 Characterization of Thioesterase Gene in *S. Indicum* using Bioinformatics Tools

After examining obtained sequences of thioesterase-related proteins, it was found that retrieved thioesterase subgroups clusters mainly belonged to dicots as Sesamum is a eudicot species. Code hop and web primer showed amplification for thioesterase. Thioesterase obtained sequences exhibit high homology with *Arabidopsis*, *Brassica* sp, *Mus musculus*, and *Homosapien's* thioesterase genes (B & DSRS, 2017).

2.1.4 Genome Analysis in Abiotic stress

Multiple abiotic stress affects the growth and production of *S.indicum*. Raffinose family oligosaccharides (RFOs), raffinose, and stachyose play a crucial role in plant tolerance and seed development. 7 *SiGols* and 15 *SiRS* genes were identified in the genome of *S.indicum* analyzed under various abiotic stresses indicating that these genes were significantly regulated by drought, osmotic, salt, and waterlogging stresses (You et al., 2018).

2.1.5 Medicinal Uses

Antioxidant studies of sesame show that it enhances Vitamin E's redemption, boost liver functions and protects against oxidative stress induced by alcohol. Sesame decreases cholesterol levels while increasing high-density lipoprotein levels. The antimicrobial studies of sesame show that it is a natural antibacterial for skin pathogens like Streptococcus and Staphylococcus. It helps patients suffering from dry skin concerns and psoriasis. Anti-cancerous studies show that sesame inhibits the growth of malignant melanoma. It is also found that intake of sesame seeds enhances the plasma γ tocopherol and increases the activity of vitamin E, preventing heart diseases and cancer (Ahmed, 2019).

3D-Domain swapping is classified into three types: Bonafide 3D-Domain swapping, quasi domain 3D-Domain swapping, and candidate for 3D-Domain swapping. Table 2.1 illustrates some characteristics, such as its examples with PDB code and their functions.

TABLE 2.1: Classification of 3D-Domain Swapping

	Definition	Proteins	PDB Code	Function	Structure of Domain swapped	References
Bonafide 3D-Domain swapping	In a protein molecule, if both the monomer & dimer exist in a steady configuration in which the dimer acquires a closed configuration (Schlunegger, Bennett & Eisenberg, 1997)	1. Barnase monomer 2. Barnase trimer	1BRN 1YVS	Ribonuclease Ribonuclease	– N-terminal helices	(Buckle, Schreiber & Fersht, 1994) (Zegers, Deswarte & Wyns, 1999)

Quasi-Domain 3D-Domain swapping	If a protein form intertwined, domain-swapped oligomers without a known closed monomer <i>(Schlunegger, Bennett & Eisenberg, 1997)</i>	1. CksHs1 monomer 2. CksHs2 dimer	1DKS 1CKS	Cell cycle regulation Cell cycle regulation	– C-terminal beta-strand	<i>(Arvai, Bourne, Hickey & Tainer, 1995)</i> <i>(Parge, Arvai, Murtari, Reed & Tainer, 1993)</i>
Candidate for 3D-Domain swapping	If a protein forms an oligomer by switching domains, but there is no monomeric configuration <i>(Schlunegger, Bennett & Eisenberg, 1997)</i>	1. Bleomycin resistance protein dimer 2. Simian virus 40 oligomer	1YBL 1SVA	Bleomycin resistance Virus coat protein	N-terminal beta-strand C-terminal helix and strands	<i>(Dumas, Bergdoll, Cagnon & Masson, 1994)</i> <i>(Stehle, Gamblin, Yan & Harrison, 1996)</i>

There are several examples available that are involved in 3D-Domain swapping in plants. Some Plant molecules are listed in Table 2.2 with their PDB code.

TABLE 2.2: List of Plants previously studied in 3D-Domain swapping

S. No.	Molecule names	PDB ID		References
		Monomer	Dimer	
1.	Crystal structure of saposin like the domain of plant aspartic protease from <i>Solanum tuberosum</i>	3RF1	–	(Bryksa et al., 2011)
2.	Crystal structure of SNC1 from <i>Arabidopsis thaliana</i>	5TEC	5H3C	(Hyun, Lee, Yoon, Yi & Song, 2016)
3.	Crystal structure of <i>Arabidopsis</i> WD40 domain in complex with flowering transcription factor homolog	–	6QTS	(Lau, Podolec, Chappuis, Ulm & Hothorn, 2019)
4.	Structure of a plant NLR resistome conferring immunity	–	6J5T	(Wang et al., 2019)
5.	Structure of the calmodulin-binding domain of plant calcium ATPase ACA8	2M73	–	(Bonza et al., 2000)
6.	Cryo-EM structure of plant NLR RPP1 LRR-ID domain in complex with ATR1	–	7CRB	(Ma et al., 2020)
7.	Structure of the plant transcriptional regulator PBF-2	–	1K3A	(Desveaux, Allard, Brisson & Sygusch, 2002)
8.	Native crystal structure analysis on <i>Arabidopsis</i> dUTPase	–	2PC5	(Bajaj & Moriyama, 2007)
9.	Crystal structure of Rice cellulose synthase plant conserved Region (P-CR)	5JNP	–	(Rushton et al., 2016)

10.	Crystal structure of plant Defensin NsD7 bound to PIP2.	_	5VYP	(Järvå, Lay, Hulett & Kvansakul, 2017)
11.	Plant cysteine oxidase PCO5 from <i>Arabidopsis thaliana</i>	6SBP	_	(White, Kamps, East, Taylor Kearney & Flashman, 2018)
12.	Structure of pseudokinase domain of BIR2, immune regulator of RLK/Pelle family	4L68	_	(Blaum et al., 2014)
13.	Cryo-EM structure of plant NLR RPP1 tetramer core part	_	7DFV	(Ma et al., 2020)
14.	Cryo-EM structure of plant NLR RPP1 tetramer in complex with ATR1	_	7CRC	(Ma et al., 2020)
15.	Crystal structure of the <i>Arabidopsis thaliana</i> O-acetyl serine Sulphydrylase K46A mutant.	1Z7Y	_	(Bonner, Cahoon, Knapke & Jez, 2005)

3.1 Protein sequence resources

The *S. indicum* (*Sesamum orientale*, cv. Zhongzhi No.13) proteome was administered from a publicly available UniProt database (<https://www.uniprot.org>). It contains a wide range of proteomes including Bacteria- 346,247; Viruses- 105,437; Archae- 3667; Eukaryota- 3758. UniProt is a Protein Database containing many details regarding the biological function of proteins. The total protein count in the complete proteome of “*Sesamum indicum*” is 24,109.

3.2 Identification of 3D-Domain swapping cases

3dswap-pred web server (<http://caps.ncbs.res.in>) was used for analyzing the “3D-Domain swapping” sequences in *S. indicum*. 3Dswap-pred is a web server specifically designed to predict the 3D-Domain swapping in a protein structure. The FASTA sequence needs to be uploaded to the server, and the results could be expected. From the total count of proteomes of *S. indicum*, only reviewed sequences were taken for further analysis in which out of 15 reviewed sequences of *S. indicum*, 10 sequences were found to be involved in the case of 3D-Domain swapping.

3.3 Proteome of *S. indicum*

➤ 11-beta-hydroxysteroid dehydrogenase A

The enzyme 11-beta-hydroxysteroid dehydrogenase A is an enzyme that is involved in turning active cortisol into inactive cortisone. The SOP2 gene, also known as steroleosin, is present in the enzyme and is known in the oil bodies of *Sesamum indicum* and expressed in developing seeds. Dehydrogenases/reductases, which are involved in the transmission of plant signals controlled by various sterols, are classified as preferable steroleosin proteins (*Lin & Tzen, 2004*).

➤ Peroxygenase

Triacylglycerols form the matrix of the plant seed oil bodies, which are encased in a monolayer of phospholipids that is rich in oleosins and contains a few trace proteins. Sesame oil bodies contain three small proteins that have been discovered and are currently known as Sops 1-3. Storage lipid in oil bodies is broken down by a peroxygenase that binds to calcium. Perhaps involved in the interaction between oil bodies and vacuoles during seed germination (*Chen, Tsai & Tzen, 1999*)

➤ 11S globulin seed storage protein 2

The two main storage proteins in sesame make about 80–90% of the total seed proteins: insoluble 11S globulin and soluble 2S albumin, sometimes known as alpha- and beta-globulin in common usage. Sesame 11S globulin and 2S albumin precursors were identified by sequencing two full-length cDNA clones, respectively. According to the inferred amino acid composition, 2S albumin, but not 11S globulin, is a protein with a high sulphur content (*Tai, Wu, Chen & Tzen, 1999*).

3.4 Homology searches & Structure selection

Reviewed sequences of selected proteins of *S. indicum* were submitted to the NCBI (National Centre for Biotechnology Information) website performing BLAST P (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) against Protein Data Bank (<https://www.rcsb.org>).

“Prephanate dehydrogenase” was selected amongst the 15 reviewed sequences of *S. indicum* and searched against PDB, various conformations were seen of the protein having different PDB structures. The PDB structure was chosen following the desired criteria, including confirmation of the structure (i.e., monomer, dimer, or any higher oligomer), visibility of chain swapping, and structure resolution.

Synechocystis aerogenate dehydrogenase crystal structure (PDB ID: 2F1K), with a Dimer conformation, was selected considering all the requirements.

Synechocystis sp. (PCC 6803) is one of the majorly considered species segregated in 1968 from freshwater. The whole genome of the species has been sequenced, comprising its four endogenous plasmids, & about 3000 genes have now been interpreted. Moreover, biochemical likewise between the plant chloroplasts and *Synechocystis* 6803 produce a supreme system for analysing the molecular procedure for primary stress responses in higher plants (*Yu et al., 2013*).

Aerogenate dehydrogenase is an enzyme responsible for tyrosine formation using the aerogenate route (*Legrand et al., 2006*) found in the cyanobacterium *Synechocystis* sp. PCC6803; a TyrA protein family enzyme involved in the biosynthesis of L-tyrosine via the aerogenate route, with prephenate aminotransferase that acts as a tyrosine precursor.

Reaction catalysed:



3.5 Hinge Identification

Hinge identification was performed using the HingeProt server (<http://bioinfo3d.cs.tau.ac.il>) and PACKMAN (<https://PACKMAN.bb.iastate.edu/>) for a better comparison.

3.5.1 HingeProt Server

HingeProt mainly shows a focal point on the rigid part & Hinge region of a protein structure. The hinge region is considered an essential part of any structure as it is involved in motions that are likely to have functional importance (*Emekli, Schneidman-Duhovny, Wolfson, Nussinov & Haliloglu, 2007*)

It is based on Elastic network models (ENM) using the Gaussian Network Model (GNM), which is responsible for mean square fluctuations of residues in the slowest two modes (*Haliloglu, Bahar & Erman, 1997*), and the Anisotropic network model (ANM), which is responsible for the direction of fluctuations (*Atilgan et al., 2001*). The PDB file of the selected proteins was uploaded along with the chain selection and the cut-off distance for GNM & ANM, respectively, to get the final HingeProt output. Generally, the GNM model is preferred over the ANM model because it offers more robust results.

3.5.2 PACKMAN- Packing and Motion analysis

PACKMAN is another online tool that stands for Packing and motion analysis. It is an online web server designed to predict the hinge region. It is based on three different parameters such as alpha values, clustering parameter (k) values, and a minimum hinge length parameter for the prediction of hinge residues.

The web server requires a PDB file with an alpha value of 2.8, a clustering parameter value of 4, and a minimum hinge length value of 5 which are set as default, and can be changed according to the requirement of a closed or open form of the hinges (*Khade, Kumar & Jernigan, 2020*)

3.5.3 PyMol

To visualize the Hinge region of the 3D-protein structure manually, PyMol (<https://pymol.org>) was also used. It is a graphic tool that accesses the information of macromolecules such as proteins in the PDB file and then provides the perks to modify the structures accordingly as per the desired requirement.

3.6 Molecular Dynamic Simulation

Molecular dynamics (MD) is a computational simulation approach for examining the physical movements of atoms and molecules (*Patodia, 2014*). MD simulation of the proteins was performed using GROMACS 2020.1 software package. The elements of the simulation system were protein and water. All the MD simulations were accomplished operating force field, CHARMM36 (*Huang et al., 2016*).

MD simulation was performed for “2F1K: Crystal structure of Synechocystis aerogenate dehydrogenase” and its monomer to compare the fluctuations of hinges.

3.6.1 Molecular Dynamics Simulation Steps:

Molecular Dynamics simulation comprises seven steps from creating the initial state such as generating the topology, adding box and solvation to the biomolecule (protein), and ion addition to the solvated system. Then there is an introduction to the interaction potential, such as minimizing the energy and equilibration. Finally, the prediction of the movement of the biomolecule was analyzed based on MD Run (*“Lysozyme in Water”, 2022*).

1. Generating the Topology:

Topology is generated as it reserves the information which is necessary to define within a simulation.

2. Add Box and solvation to the system:

A box is defined around a protein and then water molecules are added to the system.

3. Add Ions to the solvated system

Ions are added to bring equilibrium to the system when any protein does not have particular charges.

4. Energy Minimization

It is used to bring down the overall energy of the protein to make it stable.

5. Equilibration of system

It is done concerning the constant pressure, and constant temperature to equilibrate the solvent and ions around the protein.

6. Molecular Dynamics Production run

The MD simulation of the protein is performed for a specific time (ns).

7. Analysis

The protein is simulated and the final analysis is performed on the system such as RMSD, RMSF, Radius of Gyration, and SASA plot

Gromacs statement used in Molecular Dynamics simulation.

1. GENERATE TOPOLOGY

➤ `pdb2gmx -f protein.pdb -o protein_processed.gro -water spce -ignh`

2. DEFINE BOX DIMENSION & ADD SOLVATE

➤ `editconf -f protein_processed.gro -o protein_newbox.gro -c -d 1.0 -bt triclinic`

➤ `genbox -cp protein_newbox.gro -cs spc216.gro -o protein_solv.gro -p topol.top`

3. ADD IONS

Download mdp file

➤ `gmx grompp -f ions.mdp -c protein_solv.gro -p topol.top -o ions.tpr`

➤ `gmx genion -s ions.tpr -o protein_solv_ions.gro -p topol.top -pname NA -nname CL -neutral`

NOTE: choose 13 for embedded ions

4. ENERGY MINIMIZATION

Download .mdp file

- `gmx grompp -f minim.mdp -c protein_solv_ions.gro -p topol.top -o em.tpr`
- `gmx mdrun -v -deffnm em`
- `gmx energy -f em.edr -o potential.xvg`

Press “10 0” to select potential

To visualise the plot

- `xmgrace potential.xvg`

5. EQUILIBRATION

PHASE 1

Download .mdp file

- `gmx grompp -f nvt.mdp -c em.gro -p topol.top -o nvt.tpr`
- `gmx mdrun -v deffnm nvt`
- `gmx energy -f nvt.edr -o temperature.xvg`

Type “16 0” to select temperature

To visualise the plot

- `xmgrace temperature.xvg`

PHASE 2

Download .mdp file

- `gmx grompp -f npt.mdp -c nvt.gro -t nvt.cpt -p topol.top -o npt.tpr`
- `gmx mdrun -v deffnm npt`
- `gmx energy -f npt.edr -o pressure.xvg`

Type “18 0” to select pressure

6. PRODUCTION MD

Download .mdp file

- `gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md_0_30.tpr`
- `gmx mdrun -v deffnm md_0_30 -nt 16`

ANALYSIS

- `trjconv -s md_0_30.tpr -f md_0_30.xtc -o md_0_30_noPBC.xtc -pbc mol -center`

NOTE: Select "1" for protein and "0" for system for the output

1. RMSD

- `gmx rms -s md_0_30.tpr -f md_0_30_noPBC.xtc -o rmsd.xvg -tu ns`

NOTE: Select "4" for backbone for both least square & RMSD

2. RADIUS OF GYRATION

- `gmx gyrate -s md_0_30.tpr -f md_0_30_noPBC.xtc -o gyrate.xvg`

3. RMSF

- `gmx rmsf -s md_0_30.tpr -f md_0_30_noPBC.xtc -o rmsf.xvg`

4. SASA

- `gmx sasa -s md_0_30.tpr -f md_30_noPBC.xtc -o sasa.xvg`

4.1 Protein sequence resource

The UniProt search count of *S. indicum* produced a protein count of 24,109, with 24,094 as unreviewed and 15 as reviewed sequences. Reviewed sequences were taken for the further analysis of 3D-Domain swapping. Here is the list of reviewed sequences for the proteomes of *S. indicum* available in the UniProt database.

TABLE 4.1: Reviewed sequences in the Proteome of *Sesamum indicum* in the UniProt database

S. No.	Protein Names	Genes Names	Organism	Length
1.	Prephenate dehydrogenase	TyrA	<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	279 AA
2.	11-beta-hydroxysteroid dehydrogenase A	SOP3	<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	348 AA
3.	Peroxygenase	SOP1	<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	245 AA
4.	Oleosin L		<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	145 AA
5.	Ribulose biphosphate carboxylase large chain	rbcL	<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	443 AA
6.	Oleosin H1		<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	166 AA
7.	2S seed storage protein 1		<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	148 AA

8.	Cytochrome c		<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	111 AA
9.	(+)-piperitol/(+)-sesamin synthase CYP81Q1	CYP81Q1	<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	506 AA
10.	NAD(P)H-quinone oxidoreductase subunit 5, chloroplastic	ndhF	<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	684 AA
11.	Oleosin H2		<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	144 AA
12.	Inositol-3-phosphate synthase		<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	510 AA
13.	Omega-3 fatty acid desaturase, chloroplastic	FAD7	<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	447 AA
14.	Antimicrobial protein 2		<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	99 AA
15.	11S globulin seed storage protein 2		<i>Sesamum indicum</i> (Oriental sesame) (<i>Sesamum orientale</i>)	459 AA

4.2 Identification of 3D-Domain swapping

The 3Dswap-pred server was used to identify 3D-Domain swapped sequences of *S. indicum*. The FASTA sequences of reviewed cases of *S. indicum* were submitted to the 3Dswap-pred server, and amongst the 15 reviewed cases- 10 cases were found to be involved in 3D-Domain swapping.

TABLE 4.2: *Sesamum indicum*: 3D-Domain swapped cases in reviewed sequences

S. No.	Entry	Protein Names	Length	Domain/Non- -Domain	References
1.	O04925	Oleosin H2	144	NO	<i>Chen J.C., et al., 1997; Tai S.S. et al., 2002</i>
2.	Q9SQ57	Peroxygenase	245	NO	<i>Chen, Tsai & Tzen, 1999; Lin et al., 2005; Jiang et al., 2008; Chen et al., 1998; Hsiao & Tzen, 2011.</i>
3.	B3EWE9	Antimicrobial protein 2	99	YES	<i>Maria-Neto S. Et al., 2011</i>
4.	Q9XHP1	2S seed storage protein 1	148	YES	<i>Tai S.S.K et al., 1999; Wolff N. Et al., 2003</i>
5.	Q9XHP0	11S globulin seed storage protein 2	459	YES	<i>Tai S.S. et al., 1999; Wallowitz M.L. et al., 2007; Beyer K et al., 2007</i>
6.	P48620	Omega-3 fatty acid desaturase, chloroplastic	447	YES	<i>Shoji K, 1995</i>
7.	Q9FYV1	Inositol-3-phosphate synthase	510	YES	<i>Jin U.-H and Chung C.-H., 2000</i>

8.	P73906	prephenate dehydrogenase	279	YES	<i>Legrand, P. Et al., 1994</i>
9.	-Q9FUJ9	Oleosin H1	166	NO	<i>Tai S.S et al.,2002; Leduc V. Et al., 2006</i>
10.	P00054	Cytochrome c	111	NO	<i>Thompson E.W. et al., 1971</i>
11.	Q8LKV5	11-beta-hydroxysteroid dehydrogenase B	362	YES	<i>Lin L.J. et al., 2002; Chen E.C. et al., 1998; Lin L.J., Tzen J.T., 2004</i>
12.	Q9XHP2	Oleosin L	145	NO	<i>Tai S.S. et al., 2002; He F., Wu Y., Liu Z, 2008; Leduc V. Et al., 2006</i>
13.	P36487	Ribulose biphosphate carboxylase large chain	443	YES	<i>Olmstead R.G et al., 1993</i>
14.	Q33DYO	(+)-piperitol/(+)-sesamin synthase CYP81Q1	506	YES	<i>Ono E et al., 2006</i>
15.	Q33113	NAD(P)H-quinone oxidoreductase subunit 5, chloroplastic	684	YES	<i>Olmstead R.G & Reeves P.A, 1995</i>

4.3 Homology Identification & Structure selection

Amongst the reviewed “3D-Domain swapped” cases, “Prephenate dehydrogenase” was selected, and then Blast P was performed against the Protein data bank (PDB) to explore the homology sequences. After performing BLAST P, some sequences have shown similarities concerning the target protein.

The selected PDB structure was "Crystal structure of *Synechocystis* arogenate dehydrogenase (PDB ID: 2F1K)". The protein conformation of this protein is Dimer with the interacting ligands like NADP Nicotinamide-Adenine dinucleotide phosphate (NAP) & 2-Amino-2-Hydroxymethyl-propane-1,3-Diol (TRS), in which swapping of the domains can be seen easily within chains of the protein.

4.3.1 PDB Structures of 2F1K: Crystal structure of *Synechocystis* arogenate dehydrogenase

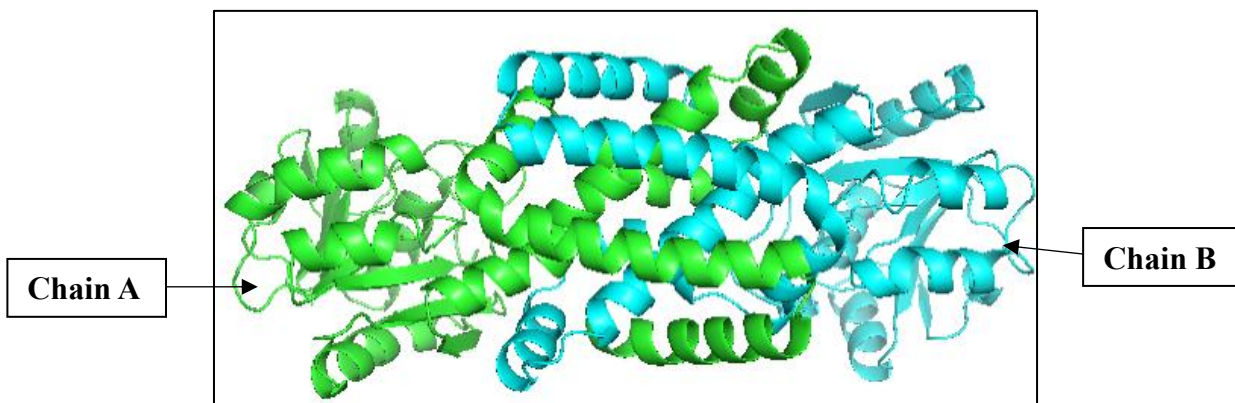


Figure 2: PDB of crystal structure of *Synechocystis* arogenate dehydrogenase: 2F1K dimer

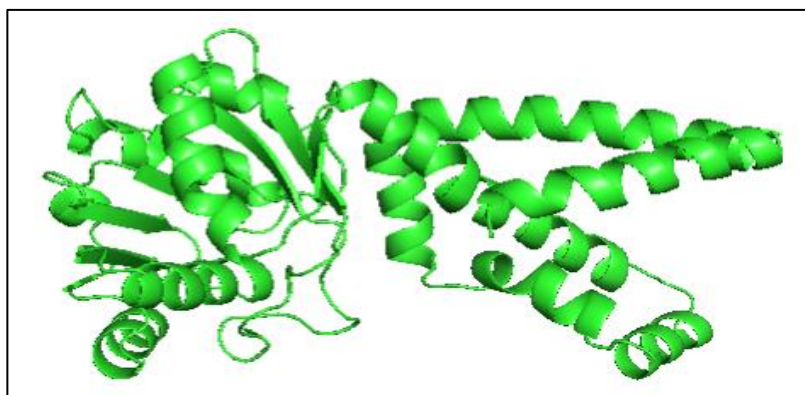


Figure 3: PDB of crystal structure of *Synechocystis* arogenate dehydrogenase: 2F1K monomer

4.4 Hinge Prediction

4.4.1 HingeProt

Using the HingeProt server, the hinge residues were identified based on the GNM & ANM methods and chain selection. It has two modes in GNM, including slowest mode 1 & slowest mode 2. After uploading the PDB file to the server, the output was displayed.

Table 4.4.1: HingeProt showing hinge residues in the slowest two modes of GNM in 2F1K dimer

Slowest Mode 1

Rigid Part No	Residues
1	A:1-180,224-243
2	A:181-223
3	A:244-278; B:1-191,213-279
4	B:192-212

Hinge residues: 180A 223A 243A 191B 212B

Slowest Mode 2

Rigid Part No	Residues
1	A:1-162
2	A:163-278; B:168-279
3	B:1-167

Hinge residues: 162A 278A 167B

Table 4.4.2: HingeProt showing hinge residues in the slowest two modes of GNM in 2F1K monomer

Slowest Mode 1

Rigid Part No	Residues
1	A:1-171
2	A:172-278

Hinge residues: 171A

Slowest Mode 2

Rigid Part No	Residues
1	A:1-167
2	A:168-245
3	A:246-278

Hinge residues: 167A 245A

4.4.3 PACKMAN Web Server

The web server requires a PDB file and an alpha value of 2.8 set as default, it can be changed according to the requirement of a closed or open form of the hinges.

Table 4.4.3.1: PACKMAN showing hinge residues in 2F1K Dimer

Protein Name	PDB ID	Hinge	Position
Crystal structure of <i>Synechocystis</i> aerogenate dehydrogenase	2F1K	1 (Chain A)	THR-166 to VAL-184
		2	ASN-222 to ASP-249
		3 (Chain B)	PRO-167 to MET-183
		4	LEU-225 to ASP-249

Table 4.4.3.2: PACKMAN showing hinge residues in 2F1K Monomer

Protein Name	PDB ID	Hinge	Position
Crystal Structure of <i>Synechocystis</i> aerogenate dehydrogenase	2F1K	1	PRO-167 to MET-183
		2	LEU-225 to LEU-248

4.5 PyMol Visualization of hinges

PyMol is a protein visualization tool specifically designed to view the 3D conformation of the protein molecule. The PDB file of the proteins was uploaded in the PyMol software, and then the sequence was selected in the display option to check for the hinge region involved in 3D-Domain swapping. Therefore, the PyMol can manually show the hinge regions involved in 3D-Domain swapping.

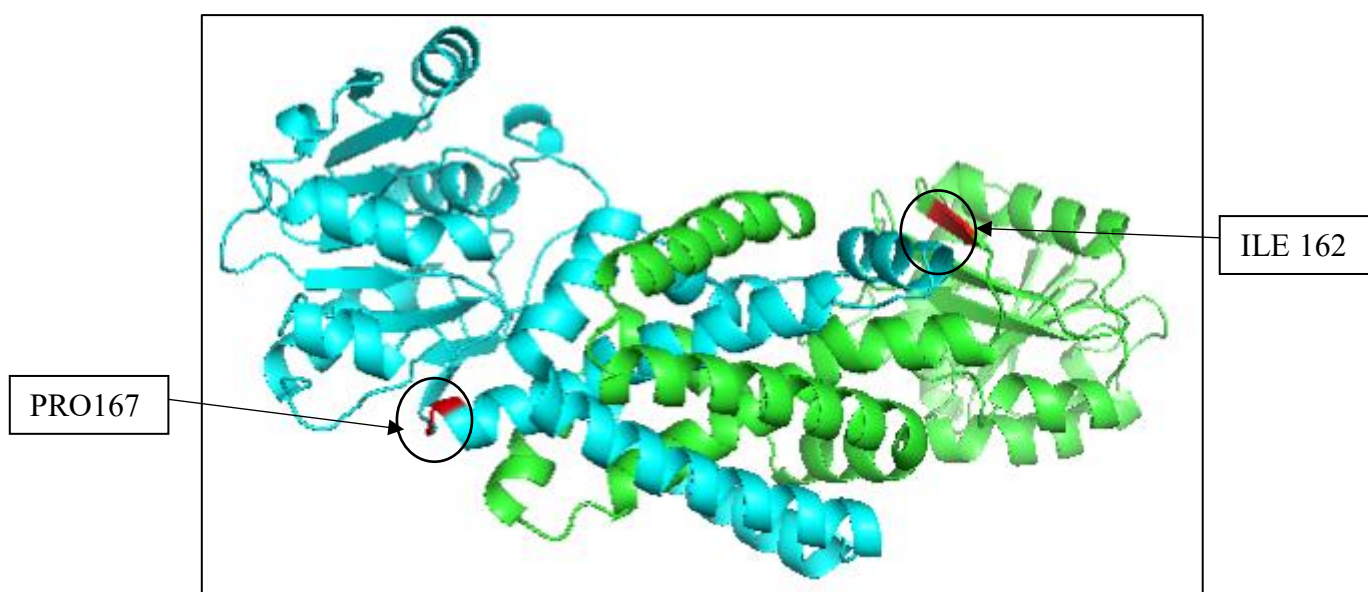


Figure 4: PyMol visualization of hinge region (HingeProt) in 2F1K dimer

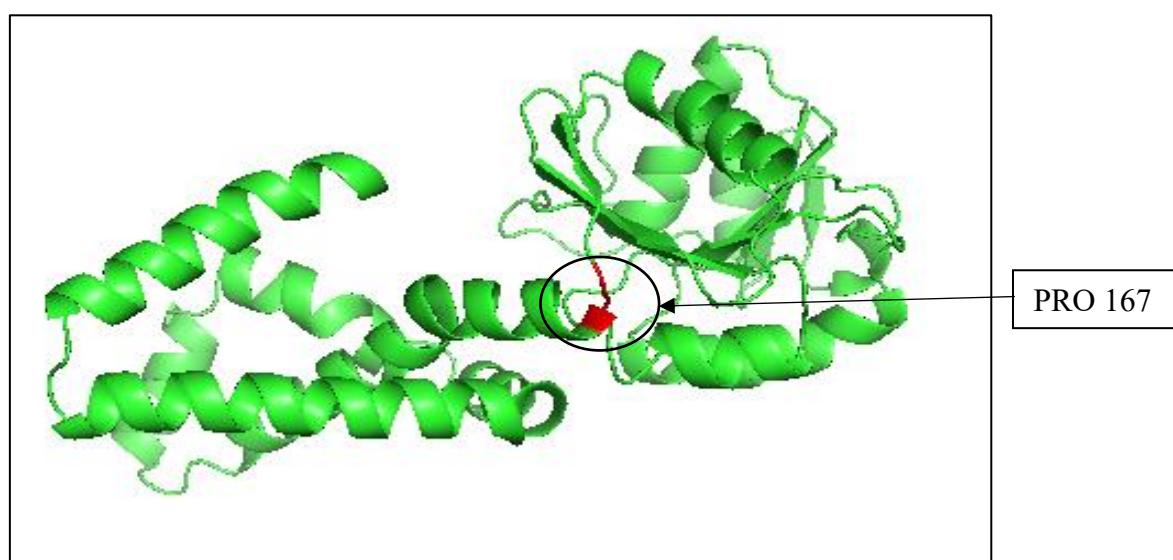


Figure 5: PyMol visualization of hinge region (HingeProt) in 2F1K monomer

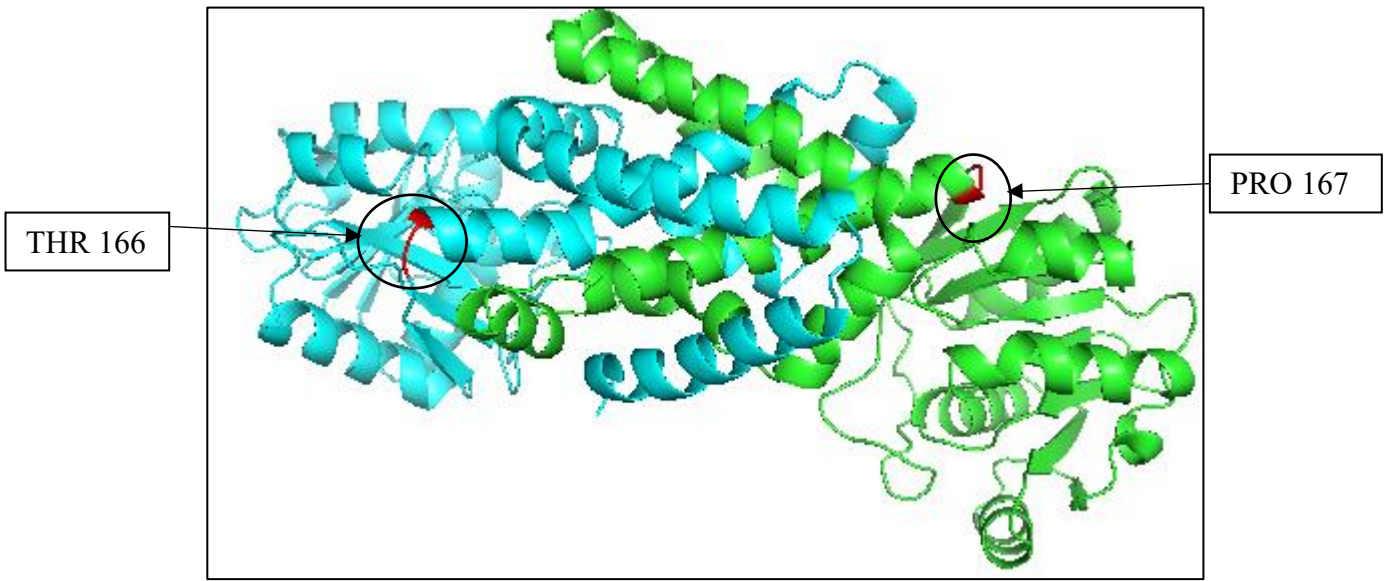


Figure 6: Pymol visualization of Hinge region (PACKMAN) in 2F1K dimer

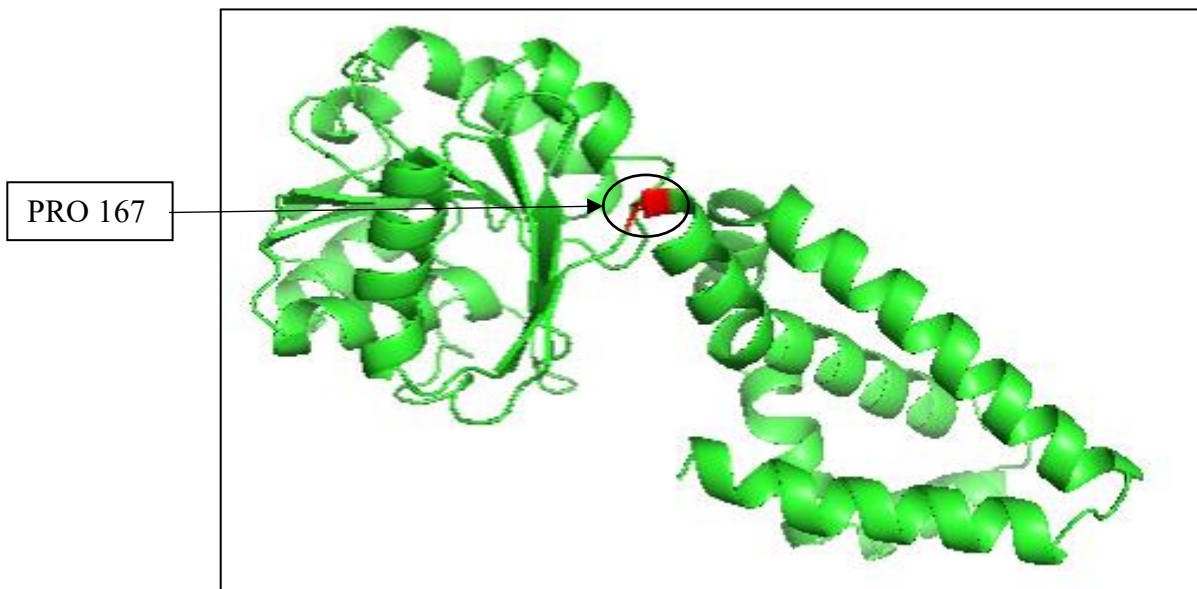


Figure 7: Pymol visualization of Hinge region (PACKMAN) in 2F1K monomer

4.6 Molecular Dynamics Simulation Studies

The Molecular Dynamics simulation was carried out for the proteins "2F1K: Crystal structure of *Synechocystis aerogenate* dehydrogenase" and its monomeric structure.

MD simulation was performed for 30 nanoseconds and analysed for Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Radius of Gyration, and Solvent Accessible Surface Area (SASA).

4.6.1 RMSD

Root Mean Square Deviation (RMSD) is a measure in molecular dynamics in which the structural distance of a protein can be depicted over the whole duration of time, and it tells the average distance between a group of atoms (*"RMSD/RMSF Analysis |BioChemCoRe 2018", 2022*) RMSD analysis showing Molecular Dynamics Simulation for 30 ns of protein 2F1K Dimer and its monomer respectively, with its Time (ns) on the x-axis and RMSD (nm) on y-axis.

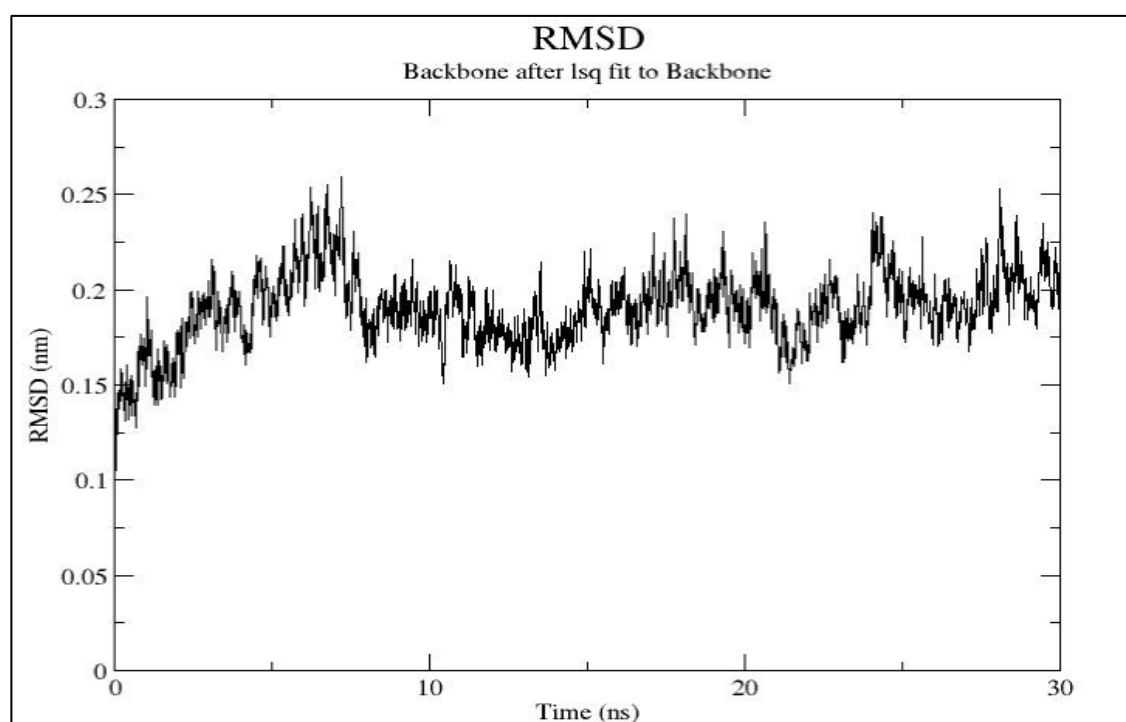


Figure 8: RMSD analysis in 2F1K dimer

The graph analyzes the Root mean square deviation (RMSD) which was calculated for backbone atoms for 30 ns in a 2F1K dimer. The fluctuations were observed at 0.1nm and reached 0.2nm between the time intervals of 0-5 ns.

The next high jump of the fluctuations was observed between the time intervals of 5-10 ns at 0.25nm, where the system reached its equilibrium. Between the intervals of 10-20 ns, the stability of fluctuations was observed at 0.2 nm. After 20ns a mild drop in fluctuation was noticed and again a jump which fluctuates till 30ns at 0.2 nm.

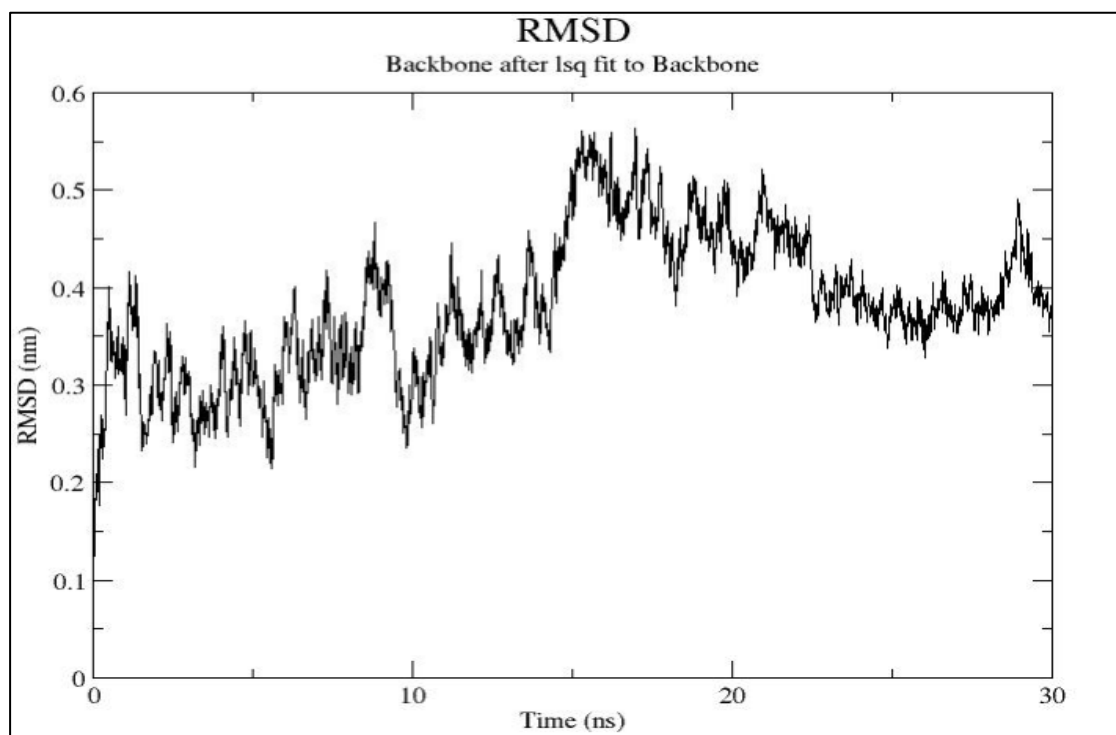


Figure 9: RMSD analysis in 2F1K monomer

The graph analyzes the Root mean square deviation (RMSD) which was calculated for backbone atoms for 30ns in a 2F1K monomer. In this case, the initial fluctuation can be seen at 0.2 nm and goes up to 0.4nm. The fluctuations kept changing and dropped again around 0.2 nm at 10 ns. There was no stability observed in the fluctuations. Between 10-20ns a huge jump in fluctuations were observed where the system reached its equilibrium above 0.5 nm. After 20 ns slight fluctuations were seen and it ended above 0.3 at 30ns.

4.6.2 Radius of Gyration

In MD simulation, the radius of gyration of a protein is calculated to determine its compactness. If the protein structure is folded stably, there are maximum chances of an enduring value for the radius of gyration (Rg). On the other side, if a protein unfolds, the changes can be seen in the value of the Radius of Gyration (Rg) gradually.

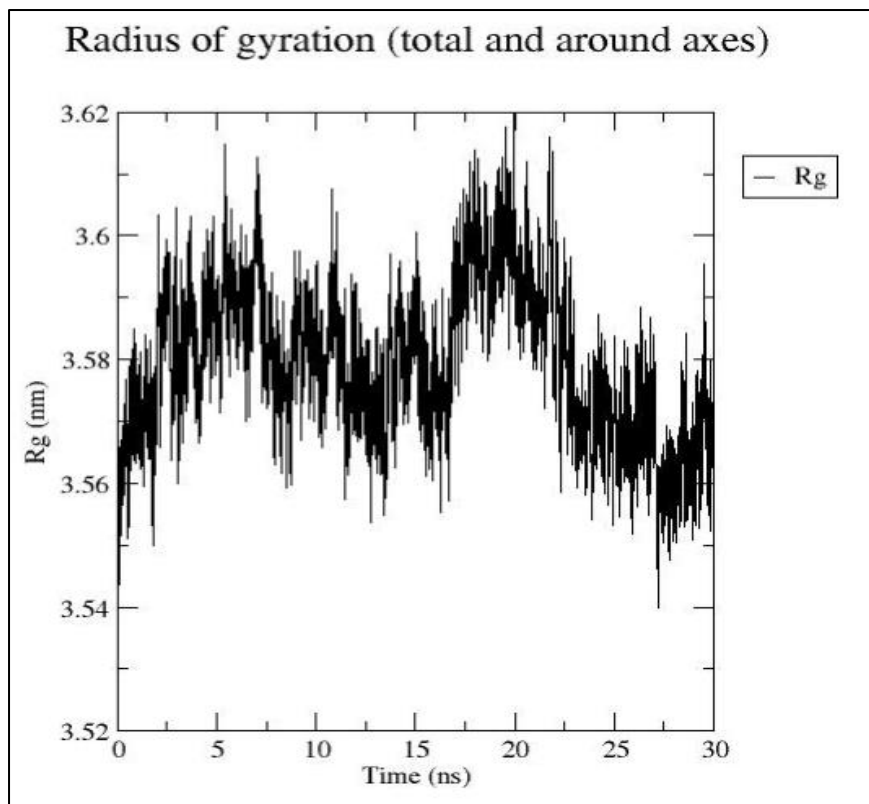


Figure 10: Radius of Gyration analysis in 2F1K dimer

The graph analyzes the Radius of Gyration (Rg) for 30ns in 2F1K dimer. Here the initial fluctuations were begun just after 3.54 nm and during the whole simulation the fluctuations can be observed and no stability could be seen. This suggests that the protein is not folded properly as it was not able to maintain its steady state.

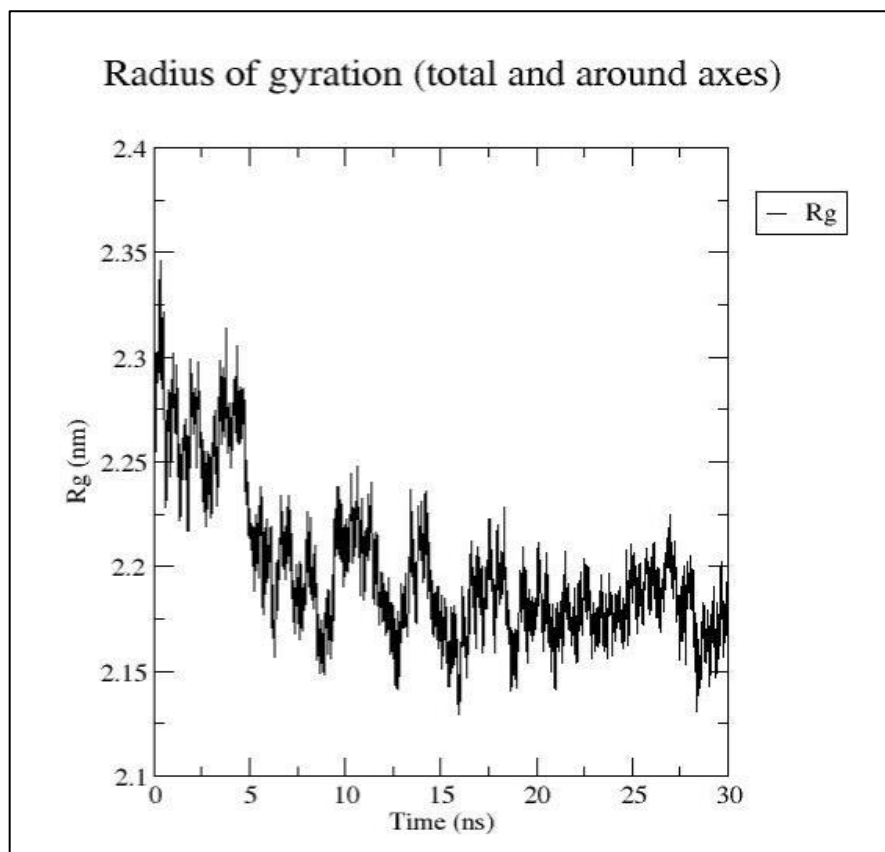


Figure 11: Radius of Gyration analysis in 2F1K monomer

The graph analyzes the Radius of Gyration (Rg) for 30ns in 2F1K monomer. In this case the initial fluctuation was begin at 2.25 nm and reached up to 2.35 nm. After 5ns the fluctuations came down and throughout the whole simulation till 20 ns the fluctuations got stable, showing some compactness of the protein.

4.6.3 RMSF

Root Mean square Fluctuation (RMSF) is a measure generally used to check the flexibility of individual residues in the protein. We can say how an individual residue fluctuates or moves during a simulation (*"RMSD/RMSF Analysis | BioChemCoRe 2018", 2022*) RMSD is exceptionally plotted for residue number vs. Rmsf (nm). From plotting the graph, it could be easily depicted which amino acid residue in a protein is responsible for the motion.

RMSF analysis showing Molecular Dynamics Simulation of protein 2F1K Dimer and its monomer respectively with its residues on the x-axis and RMSF (nm) on the y-axis.

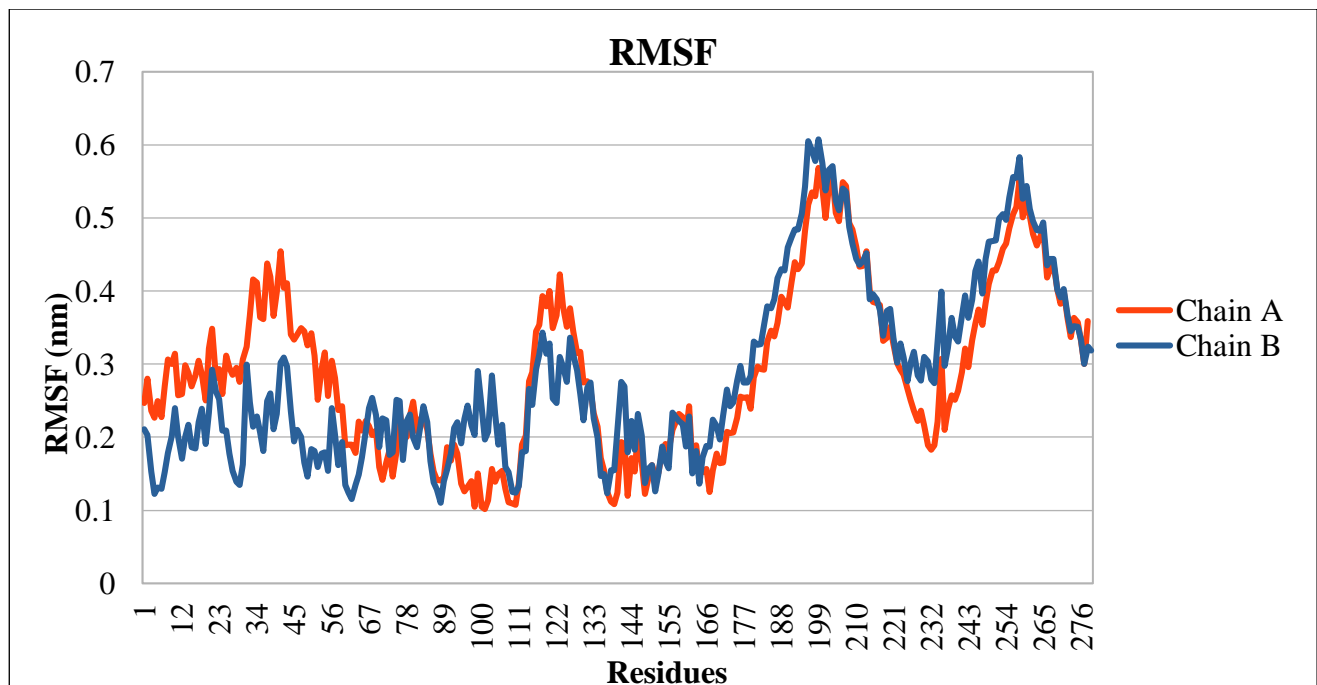


Figure 12: RMSF analysis in 2F1K dimer

The graph analyses the Root mean square fluctuations (RMSF) in 2F1K dimer. The fluctuations for both the chains (A&B) have shown in parallel. The only difference was the initial fluctuation values on different residues. The initial fluctuation of chain A was observed at just below 0.3nm and for chain B it was observed at 0.2 nm at residue no. 1. At different residues the fluctuations were varying. Between the residues 109-140 again both the chains (A & B) had a spike in the fluctuation, for chain A it was observed at above 0.4 nm and for chain B it was observed at above 0.3 nm. After residue no. 140 again it became stable and remained

stable to the residue no. 169. After residue no. 169 the highest fluctuations were observed for both the chains, for chain A it was 0.6 nm and for chain B it was just below the 0.6 nm and remains this way throughout the simulation till the end.

The hinge residues (Ile 162; Pro 167 and Thr 166) have not shown any kind of higher fluctuations. So during the initial phase the protein was more stabilized rather than its ending phase

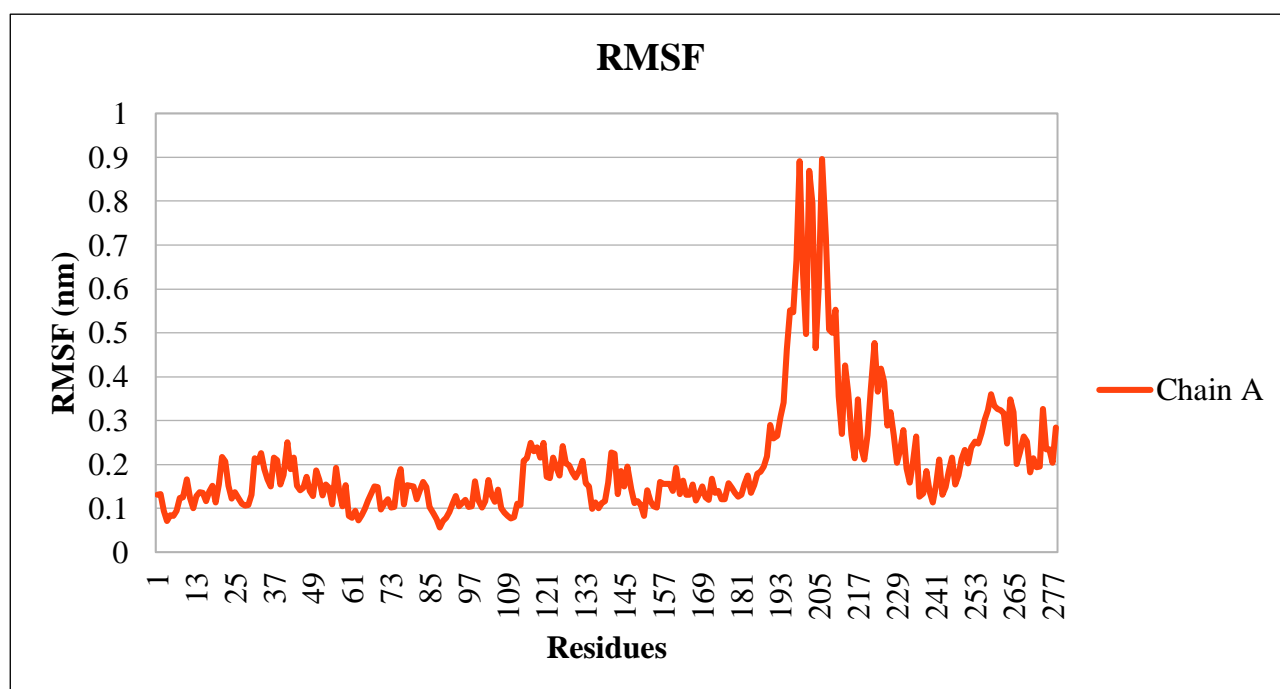


Figure 13: RMSF analysis in 2F1K monomer

The graph analyses the Root mean square fluctuations (RMSF) in 2F1K monomer. The fluctuations were very stable from the beginning till the residues between 181-193 at 0.1 nm. After that a huge spike in fluctuation were observed at 0.9nm between the residues 205-210. And again the fluctuations got a bit stabilized at 0.3. The hinge residue (pro 167) had the least fluctuation.

4.6.4 SASA Plot

Solvent accessible surface Area (SASA) of protein is considered to study protein folding and its stability. Generally, SASA value of a molecule is characterized as the surface area which is accessible by the solvent molecule

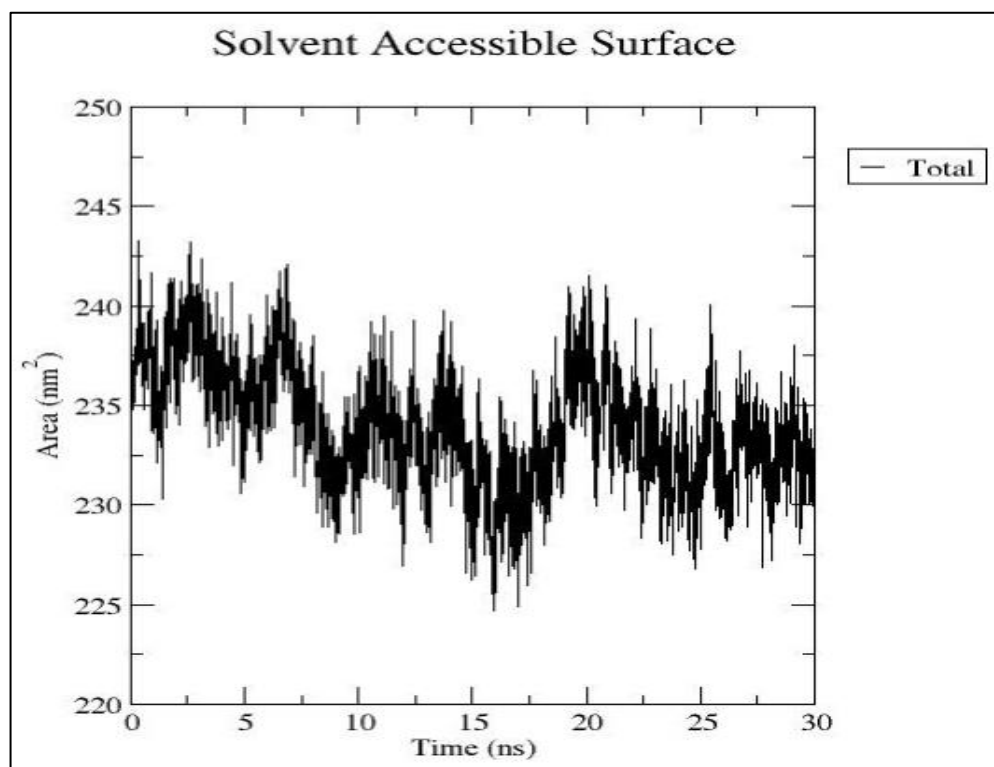


Figure 14: SASA analysis in 2F1K dimer

Solvent accessible surface area (SASA) fluctuations determines the Surface area of a bio molecule that is accessible to the solvent. It is used to evaluate enzyme stability. Solvent accessible surface area for 30ns in 2F1K dimer show hydrophobic interaction that goes up and comes down during the simulation.

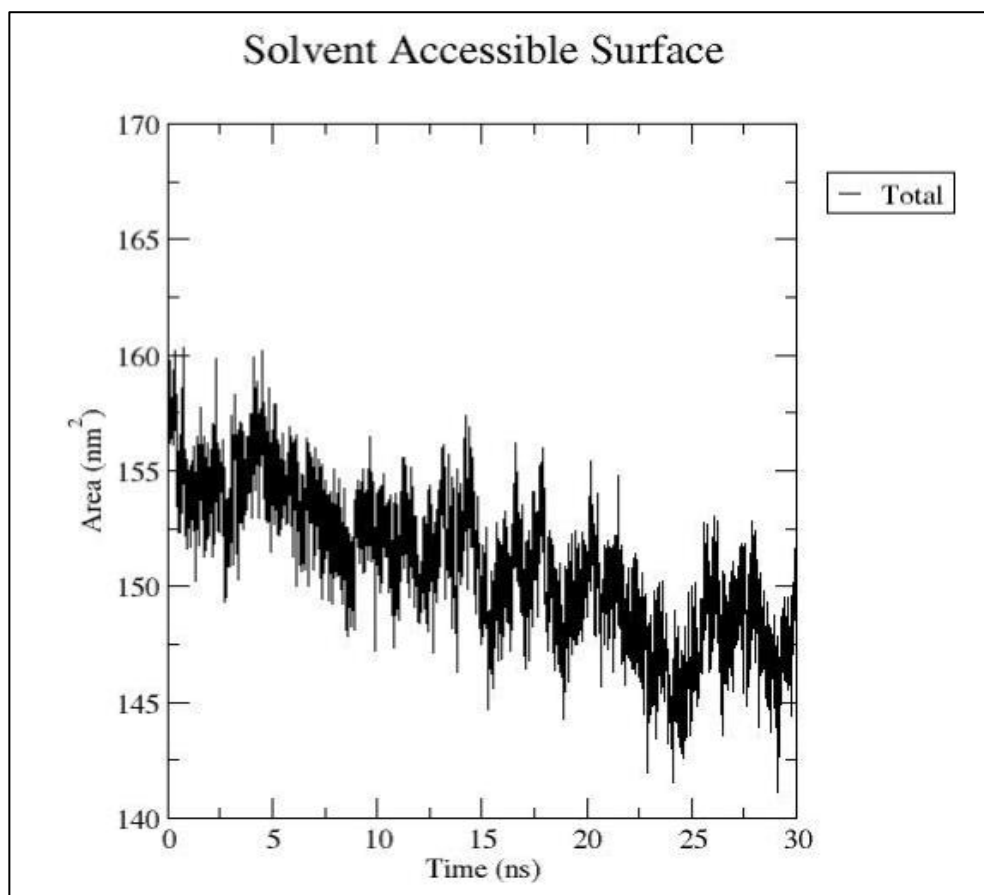


Figure 15: SASA analysis in 2F1K monomer

Solvent accessible surface area for 30ns in 2F1K monomer show hydrophobic interaction that was reduced during the simulation at the end of time

In the present study the reviewed proteomes of “*Sesamum indicum*” was selected and the analysis of hinge region was predicted that was involved in domain swapping. The study was further proceeded with the identification of these reviewed cases that were involved in 3D-Domain swapping with the help of 3Dswap-pred server which gave the outcome results for the positive and negative cases. After the prediction of swapped reviewed cases, the homologous sequences and structures were searched for the “Prephenate dehydrogenase” by using BLASTP against PDB. “*Synechocystis aerogenate dehydrogenase*” was selected as it was showing the visibility of swapped domain within the chain and the availability of its dimer and monomer structures. After getting the structures, the hinge prediction was done with the help of online server such as HingeProt and PACKMAN and for the manual visualization PyMol was used. The Molecular Dynamics (MD) simulation was performed and different analysis were carried out including RMSD where in case of dimer and monomer, the Dimer was much stable as it has shown less fluctuations than its monomer. RMSF analysis have shown that the hinge residues which were involved in domain swapping have a least fluctuation. Radius of gyration analysis shown that the dimer structure was unstable during the simulation as it was showing higher fluctuations than monomer and hence monomer structure was much more compact. SASA plot analysis the hydrophobic interaction, when the results were obtained it was observed that dimer has higher fluctuations again and in case of monomer the fluctuations were coming down at the end of simulation, making it more stable.

The present study analysis the hinge region in the proteome of *Sesamum indicum*, which were involved in 3D-Domain swapping and then a comparison was carried out in the dimer and monomer structure. From the results it was concluded that in case of RMSD analysis the monomer has shown the higher fluctuations as compared to the dimer. Analysis with the Radius of gyration in case of dimer it was observed that it has not shown the compactness of the protein as the fluctuations were higher as compared to the monomer. The hinge residues in RMSF analysis were also observed and concluded that in both the cases the hinge residues have not shown any fluctuations and were stable but throughout the simulation the dimer structure has shown higher fluctuations at the end of simulation as compared to the monomer whereas the fluctuations in monomer were stable and jump at some time but again became stable. The SASA plot analyses the hydrophobic interaction where the dimer has fluctuations showing no stability and the monomer fluctuations were coming down towards the end of time during the simulation.

The current state of knowledge recommends that 3D-Domain swapping occurs under suitable conditions in any protein with a dissipated terminus. This study intends to provide a framework for comprehending future discoveries of 3D-Domain swapping as well as an overview of the researched domain-swapped proteins because domains continue to be swapped (*Liu & Eisenberg, 2002*).

- ❖ Ahmed, S. (2019). Medicinal properties of kunjad (*Sesamum indicum*): A review.
- ❖ Arvai, A., Bourne, Y., Hickey, M., & Tainer, J. (1995). Crystal Structure of the Human Cell Cycle Protein CksHs1: Single Domain Fold with Similarity to Kinase N-lobe Domain. *Journal of Molecular Biology*, 249(5), 835-842. doi: 10.1006/jmbi.1995.0341
- ❖ Atilgan, A., Durell, S., Jernigan, R., Demirel, M., Keskin, O., & Bahar, I. (2001). Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal*, 80(1), 505-515. doi: 10.1016/s0006-3495(01)76033-x
- ❖ Atilgan, A., Durell, S., Jernigan, R., Demirel, M., Keskin, O., & Bahar, I. (2001). Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal*, 80(1), 505-515. doi: 10.1016/s0006-3495(01)76033-x
- ❖ B, S., & DSRS, P. (2017). Characterization of Thioesterase Gene in *Sesamum indicum* using Bioinformatics Tools.
- ❖ Bajaj, M., & Moriyama, H. (2007). Purification, crystallization and preliminary crystallographic analysis of deoxyuridine triphosphate nucleotidohydrolase from *Arabidopsis thaliana*. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, 63(5), 409-411. doi: 10.1107/s1744309107016004
- ❖ Bennett, M., Sawaya, M., & Eisenberg, D. (2006). Deposition Diseases and 3D Domain Swapping. *Structure*, 14(5), 811-824. doi: 10.1016/j.str.2006.03.011
- ❖ Bennett, M., Schlunegger, M., & Eisenberg, D. (1995). 3D domain swapping: A mechanism for oligomer assembly. *Protein Science*, 4(12), 2455-2468. doi: 10.1002/pro.5560041202
- ❖ Blaum, B., Mazzotta, S., Nöldeke, E., Halter, T., Madlung, J., Kemmerling, B., & Stehle, T. (2014). Structure of the pseudokinase domain of BIR2, a regulator of BAK1-mediated immune signaling in *Arabidopsis*. *Journal of Structural Biology*, 186(1), 112-121. doi: 10.1016/j.jsb.2014.02.005
- ❖ Bonner, E., Cahoon, R., Knapke, S., & Jez, J. (2005). Molecular Basis of Cysteine Biosynthesis in Plants. *Journal of Biological Chemistry*, 280(46), 38803-38813. doi: 10.1074/jbc.m505313200
- ❖ Bonza, M., Morandini, P., Luoni, L., Geisler, M., Palmgren, M., & De Michelis, M. (2000). At-ACA8 Encodes a Plasma Membrane-Localized Calcium-ATPase of

Arabidopsis with a Calmodulin-Binding Domain at the N Terminus. *Plant Physiology*, 123(4), 1495-1506. doi: 10.1104/pp.123.4.1495

- ❖ Chen, J., Tsai, C., & Tzen, J. (1999). Cloning and Secondary Structure Analysis of Caleosin, a Unique Calcium-Binding Protein in Oil Bodies of Plant Seeds. *Plant and Cell Physiology*, 40(10), 1079-1086. doi: 10.1093/oxfordjournals.pcp.a029490
- ❖ Buckle, A., Schreiber, G., & Fersht, A. (1994). Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0-Å Resolution. *Biochemistry*, 33(30), 8878-8889. doi: 10.1021/bi00196a004
- ❖ Desveaux, D., Allard, J., Brisson, N., & Sygusch, J. (2002). A new family of plant transcription factors displays a novel ssDNA-binding surface. *Nature Structural Biology*, 9(7), 512-517. doi: 10.1038/nsb814
- ❖ Dossa, K., Diouf, D., Wang, L., Wei, X., Zhang, Y., & Niang, M. et al. (2017). The Emerging Oilseed Crop *Sesamum indicum* Enters the “Omics” Era. *Frontiers in Plant Science*, 8. doi: 10.3389/fpls.2017.01154
- ❖ Dumas, P., Bergdoll, M., Cagnon, C., & Masson, J. (1994). Crystal structure and site-directed mutagenesis of a bleomycin resistance protein and their significance for drug sequestering. *The EMBO Journal*, 13(11), 2483-2492. doi: 10.1002/j.1460-2075.1994.tb06535.x
- ❖ Emekli, U., Schneidman-Duhovny, D., Wolfson, H., Nussinov, R., & Haliloglu, T. (2007). HingeProt: Automated prediction of hinges in protein structures. *Proteins: Structure, Function, and Bioinformatics*, 70(4), 1219-1227. doi: 10.1002/prot.21613
- ❖ Goodsell, D., & Olson, A. (2000). Structural Symmetry and Protein Function. *Annual Review of Biophysics and Biomolecular Structure*, 29(1), 105-153. doi: 10.1146/annurev.biophys.29.1.105
- ❖ Haliloglu, T., Bahar, I., & Erman, B. (1997). Gaussian Dynamics of Folded Proteins. *Physical Review Letters*, 79(16), 3090-3093. doi: 10.1103/physrevlett.79.3090
- ❖ Hirota, S., Mashima, T., & Kobayashi, N. (2021). Use of 3D domain swapping in constructing supramolecular metalloproteins. *Chemical Communications*, 57(91), 12074-12086. doi: 10.1039/d1cc04608j

- ❖ Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., & de Groot, B. et al. (2016). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, *14*(1), 71-73. doi: 10.1038/nmeth.4067
- ❖ Järvå, M., Lay, F., Hulett, M., & Kvensakul, M. (2017). Structure of the defensin NsD7 in complex with PIP2 reveals that defensin: lipid oligomer topologies are dependent on lipid type. *FEBS Letters*, *591*(16), 2482-2490. doi: 10.1002/1873-3468.12761
- ❖ Khade, P., Kumar, A., & Jernigan, R. (2020). Characterizing and Predicting Protein Hinges for Mechanistic Insight. *Journal Of Molecular Biology*, *432*(2), 508-522. doi: 10.1016/j.jmb.2019.11.018
- ❖ Lau, K., Podolec, R., Chappuis, R., Ulm, R., & Hothorn, M. (2019). Plant photoreceptors and their signaling components compete for COP 1 binding via VP peptide motifs. *The EMBO Journal*, *38*(18). doi: 10.15252/embj.2019102140
- ❖ Legrand, P., Dumas, R., Seux, M., Rippert, P., Ravelli, R., Ferrer, J., & Matringe, M. (2006). Biochemical Characterization and Crystal Structure of Synechocystis Arogenate Dehydrogenase Provide Insights into Catalytic Reaction. *Structure*, *14*(4), 767-776. doi: 10.1016/j.str.2006.01.006
- ❖ Liu, Y., & Eisenberg, D. (2002). 3D domain swapping: As domains continue to swap. *Protein Science*, *11*(6), 1285-1299. doi: 10.1110/ps.0201402
- ❖ Lin, L., & Tzen, J. (2004). Two distinct steroleosins are present in seed oil bodies. *Plant Physiology and Biochemistry*, *42*(7-8), 601-608. doi: 10.1016/j.plaphy.2004.06.006
- ❖ Lysozyme in Water. (2022). Retrieved 25 July 2022, from <http://www.mdtutorials.com/gmx/lysozyme/index.html>
- ❖ Ma, S., Lapin, D., Liu, L., Sun, Y., Song, W., & Zhang, X. et al. (2020). Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme. *Science*, *370*(6521). doi: 10.1126/science.abe3069
- ❖ Miraj, S., & Kiani, s. (2016). Bioactivity of Sesamum indicum: A review study.
- ❖ Parge, H., Arvai, A., Murtari, D., Reed, S., & Tainer, J. (1993). Human CksHs2 Atomic Structure: a Role for Its Hexameric Assembly in Cell Cycle Control. *Science*, *262*(5132), 387-395. doi: 10.1126/science.8211159
- ❖ Patodia, S. (2014). Molecular Dynamics Simulation of Proteins: A Brief Overview. *Journal of Physical Chemistry & Biophysics*, *4*(6). doi: 10.4172/2161-0398.1000166

- ❖ RMSD/RMSF Analysis | BioChemCoRe 2018. (2022). Retrieved 25 July 2022, from <https://ctlee.github.io/BioChemCoRe-2018/rmsd-rmsf/>
- ❖ Rousseau, F., Schymkowitz, J., & Itzhaki, L. (2012). Implications of 3D Domain Swapping for Protein Folding, Misfolding and Function. *Advances in Experimental Medicine and Biology*, 137-152. doi: 10.1007/978-1-4614-3229-6_9
- ❖ Rushton, P., Olek, A., Makowski, L., Badger, J., Steussy, C., Carpita, N., & Stauffacher, C. (2016). Rice Cellulose SynthaseA8 Plant-Conserved Region Is a Coiled-Coil at the Catalytic Core Entrance. *Plant Physiology*, 173(1), 482-494. doi: 10.1104/pp.16.00739
- ❖ SCHLUNEGGER, M., BENNETT, M., & EISENBERG, D. (1997). Oligomer Formation by 3D Domain Swapping: A Model for Protein Assembly and Misassembly. *Advances in Protein Chemistry*, 61-122. doi: 10.1016/s0065-3233(08)60319-8
- ❖ Shameer, K., Pugalenti, G., Kandaswamy, K., Suganthan, P., Archunan, G., & Sowdhamini, R. (2010). Insights into Protein Sequence and Structure-Derived Features Mediating 3D Domain Swapping Mechanism using Support Vector Machine Based Approach. *Bioinformatics and Biology Insights*, 4, BBI.S4464. doi: 10.4137/bbi.s4464
- ❖ Shameer, K., Shingate, P., Manjunath, S., Karthika, M., Pugalenti, G., & Sowdhamini, R. (2011). 3DSwap: curated knowledgebase of proteins involved in 3D domain swapping. *Database*, 2011(0), bar042-bar042. doi: 10.1093/database/bar042
- ❖ Stehle, T., Gamblin, S., Yan, Y., & Harrison, S. (1996). The structure of simian virus 40 refined at 3.1 Å resolution. *Structure*, 4(2), 165-182. doi: 10.1016/s0969-2126(96)00020-2
- ❖ Tai, S., Wu, L., Chen, E., & Tzen, J. (1999). Molecular Cloning of 11S Globulin and 2S Albumin, the Two Major Seed Storage Proteins in Sesame. *Journal Of Agricultural And Food Chemistry*, 47(12), 4932-4938. doi: 10.1021/jf990366z
- ❖ Tran, L., Urbanowicz, A., Jasiński, M., Jaskolski, M., & Ruszkowski, M. (2021). 3D Domain Swapping Dimerization of the Receiver Domain of Cytokinin Receptor CRE1 from *Arabidopsis thaliana* and *Medicago truncatula*. *Frontiers in Plant Science*, 12. doi: 10.3389/fpls.2021.756341

- ❖ Upadhyay, A., & Sowdhamini, R. (2019). Genome-Wide Analysis of Domain-Swap Predicted Products in the Genome of Anti-Stress Medicinal Plant: *Ocimum tenuiflorum*. *Bioinformatics and Biology Insights*, 13, 117793221882136. doi: 10.1177/1177932218821362
- ❖ White, M., Kamps, J., East, S., Taylor Kearney, L., & Flashman, E. (2018). The plant cysteine oxidases from *Arabidopsis thaliana* are kinetically tailored to act as oxygen sensors. *Journal of Biological Chemistry*, 293(30), 11786-11795. doi: 10.1074/jbc.ra118.003496
- ❖ You, J., Wang, Y., Zhang, Y., Dossa, K., Li, D., & Zhou, R. et al. (2018). Genome-wide identification and expression analyses of genes involved in raffinose accumulation in sesame. *Scientific Reports*, 8(1). doi: 10.1038/s41598-018-22585-2
- ❖ Yu, Y., You, L., Liu, D., Hollinshead, W., Tang, Y., & Zhang, F. (2013). Development of *Synechocystis* sp. PCC 6803 as a Phototrophic Cell Factory. *Marine Drugs*, 11(8), 2894-2916. doi: 10.3390/md11082894
- ❖ Zhang, H., Miao, H., Wang, L., Qu, L., Liu, H., Wang, Q., & Yue, M. (2013). Genome sequencing of the important oilseed crop *Sesamum indicum* L. *Genome Biology*, 14(1). doi: 10.1186/gb-2013-14-1-401

PLAGIARISM REPORT



Document Information

Analyzed document	thesis_payal plag.docx (D142483824)
Submitted	7/28/2022 8:35:00 AM
Submitted by	Atul Kumar Upadhyay
Submitter email	atul.upadhyay@thapar.edu
Similarity	4%
Analysis address	atul.upadhyay.thapar@analysis.arkund.com

Sources included in the report

SA

Panjab University, Chandigarh / single file without reference.docx

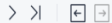
Document single file without reference.docx (D90710816)

Submitted by: chairperson_biotech@pu.ac.in

Receiver: chairperson_biotech.panuni@analysis.arkund.com

 3

1/13



150%