COMPUTATIONAL ANALYSIS OF PROTEIN-PROTEIN INTERACTIONS BETWEEN RICE AND BLAST PLANT PROTEOMICS

A dissertation submitted in the partial fulfilment for the award of Degree of Master of Science in Genomic Science

MASTER OF SCIENCE IN GENOMIC SCIENCE

By

UTTKARSH VERMA

Reg. No. BGS052128

Under the guidance of

Dr. Alagu Manickavelu

professor

Department of Genomic Science

Department of Genomic Science

School of Biological Science

Central University of Kerala

Kasaragod

August 2023

CERTIFICATE

This is to certify that the dissertation entitled "**Computational Analysis of protein-protein interactions Between Rice and Blast plant proteomes**" submitted by Mr. UTTKARSH VERMA, (Reg.No. BGS052128) to the Central University in partial fulfillment of the requirements for the award of the degree of Master of Science in GENOMIC SCIENCE is based on research carried out by his under my guidance and supervision. It is further certified that this research work has not been submitted either partially or fully for any other degree or fellowship of this or any other University.

.....	.....
Dr. A. Manickavelu Professor Dept of Genomic Science School of Biological Sciences Central University of Kerala	Dr. V. B. Sameer Kumar Professor (HoD) Dept of Genomic Science School of Biological Sciences Central University of Kerala

DECLARATION

I, Mr. UTTKARSH VERMA(BGS052128) hereby declare that the Dissertation work Entitled "**Computational Analysis of protein-protein interactions Between Rice and Blast plant proteomes**" submitted to the Central University of Kerala in partial fulfillment of the requirements for the award of the degree of Master of Science in Genomic Science is a bonafide record of original research work done by me under the supervision and guidance of Dr. Alagu Manickavelu, professor, Department of Genomic Science, Central University of Kerala, Kasaragod.

I also hereby declare that this work in part or full has not been submitted to any other University/Institution for the award of any Degree/ Diploma before.

Central University of Kerala
August, 2023

Mr. UTTKARSH VERMA
BGS052128

ACKNOWLEDGEMENT

It is very pleasant for me to acknowledge my indebtedness to all those who have helped me in completing the project. The success and outcome of this project required a lot of guidance and assistance from many people and I am extremely fortunate to have got this all along with the completion of my project work. Whatever I have done is only due to such guidance and assistance. It fills my heart with joy unspeakable to express my gratitude to everybody who contributed to the successful accomplishment of this thesis.

First and foremost, with intense gratitude, I extend my gratefulness to my project guide Prof. (Dr.) Alagu Manickavelu, professor, Department of Genomic Science, for allowing me to do this project work and providing me all the support and guidance which made me complete the project on time. I am extremely grateful for his constant inspiration and timely advice all along.

I would like to present my heartfelt thanks to Prof. (Dr.) V. B. Sameer Kumar, Head of the Department of Genomic Science, Central University of Kerala, for providing me all the facilities and support throughout my work. I also express my gratitude to Dr. Padmesh Pillai, Dr. M Nagarajan, Dr. Ranjith Kumavath, Dr. Tony Grace, for their support and cooperation throughout my research work.

I express my sincere gratitude to Dr. Preethi and Dr. Md Shanhid, for their generous efforts in encouraging and guiding me throughout my work. I feel immense pleasure in thanking Ms. Jyothsna, S, Ms. Gayathri KS (Research Scholars) and Dr. Maya P for their unlimited support and cooperation.

My gratitude could never be complete if I didn’t mention Ms. Nikhila, the lab assistant in the Department laboratory, for his continuous cooperation which helped me to progress in the right way throughout the work.

I reserve my special thanks for Ms. Manjima Manoj, Ms. Sree Lakshmi A V, Mr. Awirat Uttkarsh Jha, Mr. Uttkarsh Verma and Mr. Srinivas Poldas (lab mates), for being the constant source of inspiration and comfort throughout the course work. I would like to thank the entire CUK community for letting me work in a comfortable environment with excellent facilities and research infrastructure.

Above all, I have to thank my parents and family for their support and trust which gave me confidence and faith to face the challenges.

TABLE OF CONTENT

ABERRATION OF WORDS

BLAST *Xanthomonas oryzae* pv. *oryzae*

Xoo BLAST

BLB bacterial leaf blight

DPF peptide deformalize

BE binding energy

MD molecular dynamic

RMSD root mean square deviation

Rg radius of gyration

SASA solvent accessible surface area

RMSF root mean square fluctuation

ABSTRACT

As (***Xanthomonas oryzae* pv. *Oryzae***) BLAST is the bacteria that basically cause disease in rice (Oryza) that is basically host specific also it plays a role in for fundamental understanding of pathogen biology such as the rice production worldwide. Due to the economic importance, extensive genomic studies have been conducted to elucidate the molecular mechanism of rice response to Xoo and Xoc in the last two decades. As recent study uses to find the interaction between rice and Xoo and R protein seq or their product and effectors. These diseases lead to 16% of global crop yield losses, Plants have evolved sophisticated innate ability of each cell to fend off the attack, as we all know every disease is caused by some microorganism whether in plant or in animals as the infection of Rice (Oryza) by Xoo is causing bacterial blight (BB) Disease all the disease is caused by some protein that play a role in suppression of growth of rice as well as the protein also increase toxicity in the leaf of rice as by this it's really challenging part to understand molecular level the infection of leaf is done by effector protein which have the specific binding domine of receptor protein of Oryza, as its very challenging part to understand the working and finding the exact molecular interaction is now a days done by PPI(protein-protein interaction) or by Ai ML(machine learning) tools as this part of biology basically deals with computational part of biology with great understanding of both molecular part as well as computational part of science, for this process we have chosen the AI(artificial intelligence) ML(machine learning) .

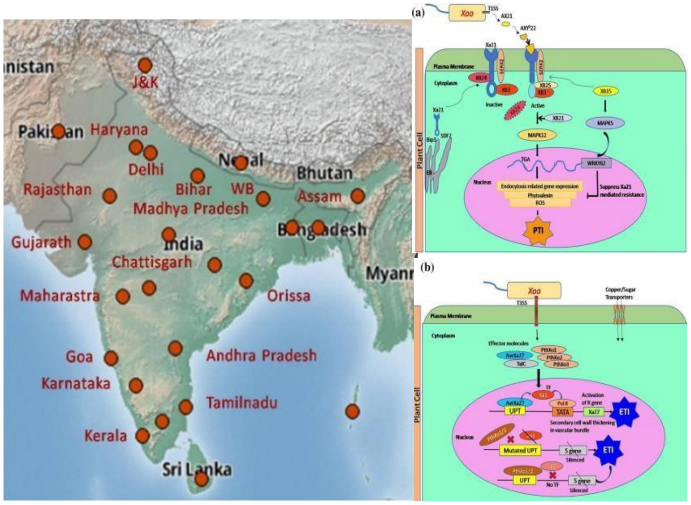
These abstract reviews the state-of-the-art bioinformatics tools designed for predicting and characterizing protein-protein interactions. We discuss different computational approaches employed by these tools, such as machine learning, network analysis, and structural modelling. Furthermore, we explore the challenges and limitations associated with PPI prediction and

highlight the potential future developments in this field. We highlight the importance of integrating multiple tools and data sources to enhance the accuracy and reliability of PPI predictions. By harnessing the power of bioinformatics tools, researchers can decipher the complex landscape of PPIs, uncover novel interaction networks, and gain insights into cellular processes and disease mechanisms.

+++

INTRODUCTION

Protein-protein interactions (PPIs) are fundamental to the functioning of biological systems, regulating various cellular processes and mediating complex molecular mechanisms. Understanding the landscape of PPIs is crucial for unravelling the intricate relationships between proteins and their roles in biological pathways. Traditional experimental methods for studying PPIs have limitations in terms of coverage and scalability. However, recent advances in bioinformatics tools and computational approaches have revolutionized our ability to predict, analyze, and understand PPIs on a large scale.



In 2020, a groundbreaking development in protein structure prediction and PPIs occurred with the introduction of AlphaFold2. AlphaFold2 is an advanced deep learning system that uses artificial intelligence (AI) and deep neural networks to predict protein structures accurately and infer potential protein interactions. The significance of AlphaFold2 lies in its ability to provide accurate and reliable predictions of protein structures, which are crucial for understanding their functions and interactions. Several computational methods have been developed to predict PPIs based on different principles and algorithms. Machine learning algorithms, such as support vector machines (SVMs) and random forests, have been applied to predict PPIs from protein sequence data **(Jones et al., 2012; Pazos & Valencia, 2002).** Network-based approaches, including graph theory and network propagation methods, analyze the topology of PPI networks to identify functional modules and uncover hidden interactions **(Bader & Hogue, 2003; Vella et**

al., 2017). Structural modelling techniques, such as protein docking and molecular dynamics simulations, provide insights into protein complexes' three-dimensional structure and binding interfaces **(Yan et al., 2020; Schneidman-Duhovny et al., 2005).** The original Alpha Fold was introduced by Senior et al. in 2019, demonstrating impressive progress in protein structure prediction., marked a significant advancement with its remarkable performance in the Critical Assessment of Structure Prediction (CASP) competition. AlphaFold2 outperformed other computational methods and even rivalled experimental techniques in accurately predicting protein structures. Its success has since attracted widespread attention and sparked enthusiasm within the scientific community. The accuracy and reliability of protein structure prediction achieved by AlphaFold2 have profound implications for the study of PPIs. Accurate predictions of protein structures enable researchers to infer potential interactions between proteins and provide insights into their binding interfaces. This information is crucial for understanding the molecular mechanisms underlying biological processes, drug discovery, and the design of therapeutics targeting specific PPIs.The application of AlphaFold2 in the field of PPIs has opened up new avenues for research. By combining AlphaFold2 predictions with existing experimental data and bioinformatics tools, researchers can gain deeper insights into the networks and dynamics of PPIs. This integrated approach enhances our understanding of complex biological systems and facilitates the discovery of novel interactions and functional modules. In this research paper, we aim to explore the impact of AlphaFold2 in the field of PPIs and its implications for protein structure prediction. We will review the development of AlphaFold2, highlighting its performance in protein structure prediction and its potential for predicting PPIs. We will also discuss integrating AlphaFold2 predictions with other bioinformatics tools and experimental approaches to enhance the accuracy and reliability of PPI

predictions.

Furthermore, we have examined protein-protein interaction(ppi) by using selective data from NCBI and differentiate that data into different form to filter it by using in silico tools to take best protein data from NCBI database. Select the data from NCBI by choosing three different matlab (Growth, metabolic, defense pathway) we choose 50 proteins of BLAST (Xoo) and 50 proteins of Oryza sativa(rice) based on all three pathways. We have successfully filter Protein seq by Refseq further filter it with MATLAB to check the stability of protein structure and future we select protein on the basis on stability we got 10 each protein on both effector and Receptor side which are having known function and play a role in growth of plant growth and development process, metabolic factor, transport, defense pathway etc. after selection of these protein seq we just use tool for PPI interaction between two protein rice and Xoo/ Receptor-effector complex protein model and to visualization of this protein structure in different tools in silico method and observation of result of complex protein.

LITERATURE REVIEWS AND HYPOTHESIS

Rice is the staple food for about half of the world population Grown for more than 6000 years (H Pathak, P Samal and M Shahid.2018) It accounts for 35-75% of the calories for more than 3 billion Asians. Globally, it provides 27% of dietary energy, 20% of dietary protein and 3% of dietary fat. Rice fields cover around 160 million hectares, the third largest cereal (H Pathak, P Samal and M Shahid.2018) rice is mainly grown in south region of India as from year we are facing problem growth of rice such conditions are pathogen, stress, soil but mostly the problem occur is Pathogen some pathogen like BLAST(Xoo), is one of the devastating diseases of rice worldwide. The pathogen reported to cause 70% crop loss in some of the susceptible genotypes under disease favoring environments **(Joshi2020. 10.1007/s12038-020-00085-8)**

Advance biotechnology is come forward and find the issue by which we can solve problem of these pathogen but as we all know pathogen have strength of modified and gain resistance towards pesticides so its time to take a step towards protein part and by this, we can actually stop the pathogen infection by closing the binding site to inhibited the protein that is responsible for infection. To study the compatible and incompatible interactions between rice and bacteria, a proteomic approach was applied**.(Tariq Mahmood, Asad Jan, Makoto Kakishima, Setsuko Komatsu Dr. 15 November 2006.https://doi.org/10.1002/pmic.200600470)**

BLAST(Xoo) causes bacterial leaf blight, a devastating rice disease. The Xoo-rice interaction, wherein wide-ranging host- and pathogen-derived proteins and genes wage molecular arms race, is a research hotspot. **(Guichun Wu,1 Yuqiang Zhang,2 Bo Wang,1 Kaihuai Li,3 Yuanlai Lou,1 Yancun Zhao, 1 and Fengquan Liu 1. Rice (N Y). 2021 Jun 26. doi: 10.1186/s12284-021-00503-x)**

pathogen are having protein that are responsible for infection as so plants have evolved a

sophisticated immune system to detect these effectors using cognate disease resistance proteins, a recognition that is highly specific **(Weiman Xing 1, Yan Zou, Qun Liu, Jianing Liu, Xi Luo, Qingqiu Huang, She Chen, Lihuang Zhu, Ruchang Bi, Quan Hao, Jia-Wei Wu, Jian-Min Zhou, Jijie Chai.10.1038/nature06109)**

The interaction of these protein is mostly responsible for plant pathogen infection in directly or in direct way as in some cases this interaction can make defense towards pathogen **(Shiping Yang 1, Hong Li 1, Huaqin He 2, Yuan Zhou 1, Ziding Zhang. 10.1093/bib/bbx123)**

Bharat Mishra and colleagues conducted a study in 2022 where they explored the rice protein interaction network to identify key nodes and potential targets of pathogen effectors. This study, published in the Computational and Structural Biotechnology Journal, focused on understanding the interactions between rice proteins and the effectors of BLASToryzae, shedding light on the mechanisms underlying the infection process **(Bharat Mishra 1, Nilesh Kumar 1, M Shahid Mukhtar1. DOI: 10.1016/j.csbj.2022.04.027)**

by using this PPI technique help us to understand the the effectiveness of rice resistance against these diseases. It highlighted a decade of research on the interactions between rice and BLASToryzae, emphasizing the molecular aspects that determine rice resistance and BLASToryzae virulence. The study aimed to enhance our comprehension of rice's innate immunity, the coevolution of BLASTvirulence, and potential applications in crop breeding **(2018,19(10),3008; https://doi.org/10.3390/ijms19103008)**

As we all know alpha-Fold2 I the recent AI tools which use for the PPI interaction of two or multiple complex protein structure as AlphaFold2 has shown promising potential in accurately predicting protein-protein interactions. Its AI-driven approach and the ability to provide interaction structure models offer valuable insights into the intricate mechanisms of protein

interactions, contributing to our understanding of biological processes and potentially aiding in drug discovery and other applications**. (Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2)**

protein-protein interactions (PPIs) between rice and the BLASToryzae pathogen. The references mostly discuss AlphaFold2's capabilities in predicting protein structures, interactions, and its application in protein-protein interaction studies, but they do not explicitly mention its use in predicting PPIs between rice and BLASToryzae. Any research paper as this is new approach for us using this pipeline of Alpha-fold2 but using protein-protein complex structure is being done by using alpha fold as after getting the accurate prediction of alphafold2 of PPI interaction we have to check the how these proteins are interacting takes placed. Plant-pathogen interactions, such as those involving rice and the blast fungus Magnaporthe oryzae, are critical in understanding disease resistance. Protein-protein interactions (PPIs) are central to recognizing pathogens. Structure-based approaches, including protein-protein docking, are used to predict PPIs between rice and the blast fungus, contributing to insights into molecular mechanisms of disease resistance **(Front. Plant Sci., 23 July 2021 Sec. Plant Biophysics and Modeling Volume 12 - 2021 | https://doi.org/10.3389/fpls.2021.690124).**

protein-protein docking, are used to predict plant-pathogen protein-protein interactions. These methods aim to enhance our understanding of the interactions that play crucial roles in disease progression and resistance. As we all know plant–pathogen PPIs is time-consuming and labor-intensive, computational methods are emerging as an important strategy to complement the experimental methods. In this work, we first evaluated the performance of traditional computational methods such as interolog, domain–domain interaction and domain–motif

interaction in predicting known plant–pathogen PPIs. Owing to the low sensitivity of the traditional methods, we utilized Random Forest to build an inter-species PPI prediction model based on multiple sequence encodings and novel network attributes in the established plant PPI network. So on between plants and pathogens. So far, two levels of plant immune responses to pathogens have been well established. Briefly, pattern recognition receptors located on the plant cell surface first recognize pathogen-associated molecular patterns (PAMPs) from pathogens and activate the first tier of plant immunity called PAMP-triggered immunity (PTI). To sabotage the PTI response, pathogens secrete virulence molecules called effectors into plant cells. In response, plants use intracellular resistance proteins (R-proteins) to specifically recognize effectors and trigger the second tier of immune response named effector-triggered immunity (ETI) Molecular docking is employed to study resistance mechanisms in the oomycete pathogen Peronophythora litchii. This research involves investigating the effects of specific amino acid substitutions in the pathogen's target protein, providing insights into the mechanisms underlying **resistance (Zhou, Y., Chen, L., Hu, J. et al. Resistance Mechanisms and Molecular Docking Studies of Four Novel QoI Fungicides in Peronophythora litchii. Sci Rep 5, 17466 (2015). https://doi.org/10.1038/srep17466),**

Molecular docking is a computational technique used to predict the binding interactions between a protein and a small molecule ligand. In the context of rice (Oryza sativa) and the pathogen BLAST(Xoo), molecular docking can be employed to understand the interactions between specific proteins from both organisms and potential ligands. This approach is valuable for drug discovery, understanding protein-ligand interactions, and designing targeted interventions. As Protein structures from rice and Xoo, obtained through experimental methods or computational prediction, need to be prepared for docking. Ligand structures are also prepared, and both

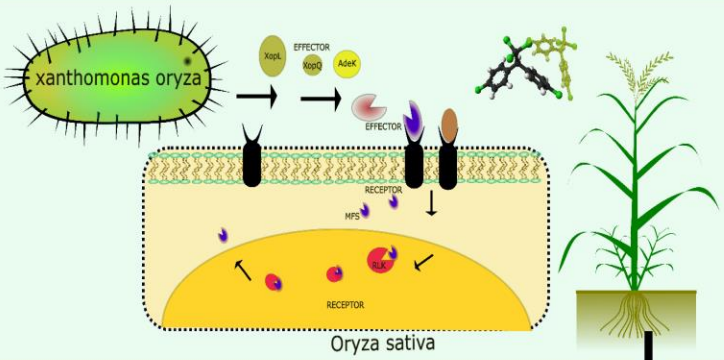
proteins and ligands are optimized to ensure accurate representation of their conformations by using such docking software tools are available, such as AutoDock, AutoDock Vina, and GOLD. Researchers choose an appropriate algorithm based on the specific research objectives and the characteristics of the proteins and ligands it is based on The binding site on the protein where the ligand is likely to bind needs to be identified. This can be based on known functional sites, protein-ligand complex structures, or prediction methods The ligand is docked into the protein's binding site using the selected algorithm by using The docking software calculates a score that reflects the binding energy between the protein and ligand. Lower scores indicate stronger binding. They analyze the interactions between the protein and ligand to understand the binding mechanism and key residues involved. Use of protein-ligand that help in Identifying small molecules that can inhibit essential Xoo proteins can lead to the development of new therapies against bacterial blight in rice. And also Molecular docking helps elucidate how Xoo proteins interact with rice proteins, providing insights into the pathogenicity and defense mechanisms. accurate protein and ligand structures are essential for reliable results. As research in this area progresses, molecular docking contributes to our understanding of the molecular interactions between rice and Xoo**,(DOI:10.1080/07391102.2020.1719200. Rashida Perveen)**

Docking plays a crucial role in inhibiting the interaction between proteins from BLAST(Xoo) and rice, thereby contributing to the development of strategies for disease control. By utilizing computational docking techniques, researchers can identify and analyze potential binding interactions between proteins and various compounds. researchers investigated the role of the pathogenicity factor F (RpfF) protein in Xoo and its potential as a drug target [3]. Through gene-gene interaction and pathway analysis, the 3D structure of the RpfF protein was modeled and refined. Virtual screening and docking identified compounds with strong binding and low

toxicity that could potentially inhibit Xoo's pathogenicity **(Mundla SRILATHA a, Naina PATYAL b, Madhu Sudhana SADDALA. https://doi.org/10.1016/S2095-3119(19)62813-3)** Bacterial leaf blight (BLB) caused by BLAST(Xoo) is one of the most damaging diseases to rice across the world. Various chemicals have been employed so far for the management of bacterial leaf blight. On the other hand, these compounds are damaging to the ecosystem and have an impact on non-target species such as humans and animals. As a result, there is a need to create a new natural inhibitor for BLB management. Deformylase (PDF) enzyme is present in all eubacteria and its necessity in bacterial protein synthesis reveals it as an attractive target for drug development. In this study, the active components of Nigella sativa have been selected based on their previously reported antimicrobial activity and screened on the active site of bacterial PDF by the in silico art of techniques. Among these compounds, dithymoquinone and p-cymene strongly bind with the PDF enzyme with binding energy values of 7.77 kcal/mol and 7.26 kcal/mol, respectively, which is comparatively higher than the control compound (−6.73 kcal/mol). **(Pravej Alam* and Thamer H. Al balawi. DOI: 10.32604/phyton.2022.021334)** docking techniques play a significant role in inhibiting the interaction between proteins of Xoo and rice. By simulating the binding interactions between potential inhibitors and key proteins, researchers can identify compounds that have the potential to disrupt essential processes in Xoo's pathogenicity, ultimately contributing to the development of effective strategies to protect rice crops from diseases like bacterial leaf blight.

PIPELINE FOR PROTEIN ANALYSIS

Xanthomonas oryzae is a bacterial pathogen that causes diseases in rice plant by Effector protein.



selected the effector protein(Xoo) and receptor protein(rice) from the NCBI database

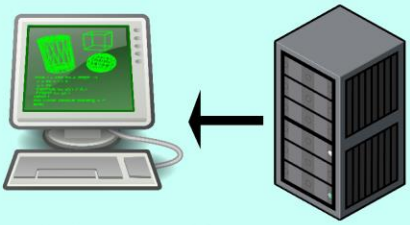
and filter it based on Protein stability

$$X = [x_1, x_2, x_3, \ldots, x_r]$$

$$h(X) = 1 / (1 + e^{(-z)})$$

$$z = \theta_0 + \theta_1 * x_1 + \theta_2 * x_2 + \theta_3 * x_3 + \ldots + \theta_r * x_r$$

$$J(\theta) = (-1/m) * \Sigma[Y_i * \log(h(X_i)) + (1 - Y_i) * \log(1 - h(X_i))], \text{ where } i = 1 \text{ to } m$$

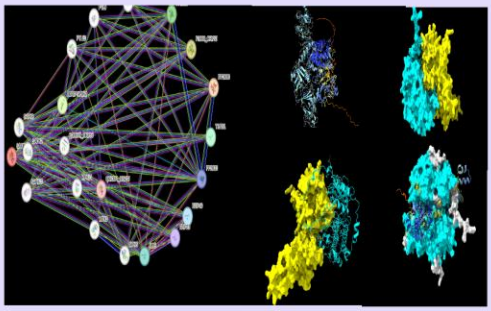


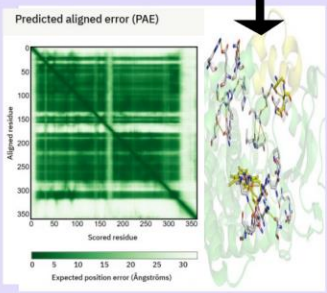
ALPHAFOLD STRUTURE PREDICTION

by using alphafold2 model for protein protein intrecation

Network analysis of PPI

functional annotation and GO analysis

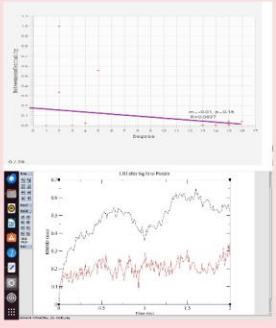


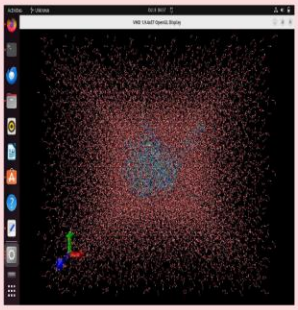


ALPHAFOLD2 PROTEIN-PROTEIN INTRECATION PREDICTION

Ligand selection and molecular docking(inhibition)

Molecular Dynamics (behaviour of protein(ppi complex)





MOLECULAR DYNAMICS

MOLECULAR DOCKING

METHOD AND METHODOLOGY

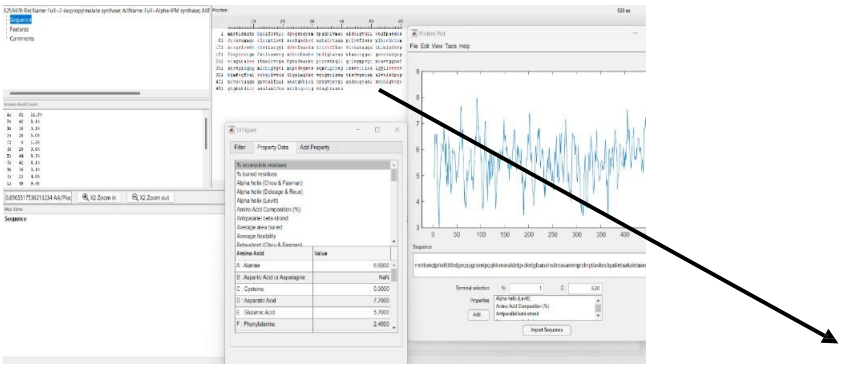
SAMPLE COLLECTION

The data was taken by NCBI website on the bases of various articles majorly which are having role in defense and metabolic pathway with proper literature and function of each protein was known: the data was taken by based on three pathways related directly or indirectly in both the organism Xoo and RICE

MATLAB SOFTWERE

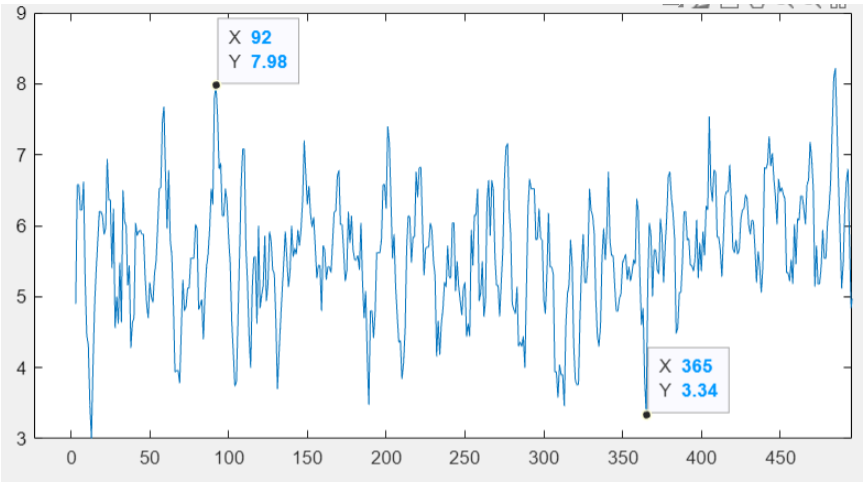
Was developed by MathWorks it is most versatile programming language and numerical computing environments most widely use in many fields like mathematics, biological science, engineering. It was providing a user-friendly interface for performing complex computational algorithms that help to analysis data also it provides a large library for data interpretations.

We have use SEQVIEWER is MATLAB tool use to visualize and explore the biological protein or nucleotide sequences in respect of protein stability analysis, this function allows for the visualization and Exploration of protein stability as it has various customizable option. Vt using "AA" Amino acids seq we can stability of protein by using various tools we have choose three characteristics for protein stability (%accessible residues, HPLC retention PH 7.4(Meek), Hydrophobicity (Aboderin)) for stability of protein



Amino Acid Color Scheme	Color Legend
Charge	• Acidic – Red • Basic – Light Blue • Neutral – Black
Function	• Acidic – Red • Basic – Light Blue • Hydropobic, nonpolar – Black • Polar, uncharged – Green
Hydrophobicity	• Hydrophilic – Light Blue • Hydrophobic – Black
Structure	• Ambivalent – Dark Green • External – Light Blue • Internal – Orange

We get the result of 50(Receptor): 50(Effector) we have chosen top 10 predicted protein which have the highest stability as for that we have choose the protein plot of all three features and by using graph took the average of one highest and one lowest value and take the average of that by that we put out the value average protein value for stabilization purpose



- $\frac{7.98 + 3.34}{2} = 5.66$
- $\frac{92 + 365}{2} = 228.5$
- (X:228.5, Y:5.66)

LIST OF SELECTED PROTEINS IN EFFECTOR AND RECEPTOR

Effector protein				Receptor protein			
	PROTEIN NAME	NCBI Reference Sequence:	ORGANISM		Protein name	NCBI Reference Sequence:	Organism
1	YopJ family type III secretion system effector XopJ [Xanthomonas]	WP_109292224.1	BLASToryzae pv. oryzae	1	receptor like kinase, partial [Oryza sativa Indica Group]	ABR25930.1	Oryza sativa Indica Group
2	xopx	WP_285957449.1	BLASToryzae pv. oryzae	2	ethylene receptor 3 isoform 2 precursor [Oryza sativa Japonica Group]	NP_001388942.1	Oryza sativa Indica Group
3	xopq	WP_258532520.1	BLASToryzae pv. oryzae	3	wall-associated receptor kinase-like 3 [Oryza sativa Japonica Group]	XP_025880762.1	Oryza sativa Indica Group
4	xopl	WIX07272.1	BLASToryzae pv. oryzae	4	putative receptor protein kinase ZmPK1 [Oryza sativa Japonica Group]	XP_025877832.1	Oryza sativa Indica Group
5	Avrbs2	ABG23670.1	BLASToryzae pv. oryzae	5	TPA_inf: WRKY transcription factor 30 [Oryza sativa Japonica Group]	DAA05095.1	Oryza sativa Indica Group
6	ABC transporter ATP-binding protein [BLASToryzae]	UNE64853.1	BLASToryzae pv. oryzae	6	RecName: Full=Calcium-binding protein CBP; Short=OsCBP	Q2QY10.1	Oryza sativa Indica Group
7	nucleoside hydrolase [BLASToryzae]	UNE64851.1	BLASToryzae pv. oryzae	7	subtilisin-like protease Pr1B, partial [Metarhizium anisopliae]	AAC49831.1	Oryza sativa Indica Group
8	IS3 family transposase [BLASToryzae]	UNE64850.1	BLASToryzae pv. oryzae	8	serine/threonine-protein kinase SAPK9 [Oryza sativa Japonica Group]	NP_001391667.1	Oryza sativa Indica Group
9	MFS transporter [BLASToryzae]	UNE64796.1	BLASToryzae pv. oryzae	9	type-1 proteins geranylgeranyltransferase subunit beta [Pyricularia oryzae Y34	ELQ34401.1	Oryza sativa Indica Group
10	AdeC/AdeK/OprM family multidrug efflux complex outer membrane factor [BLASToryzae]	UNE64758.1	BLASToryzae pv. oryzae	10	bidirectional sugar transporter SWEET2b	NP_001395931.1	Oryza sativa Indica Group

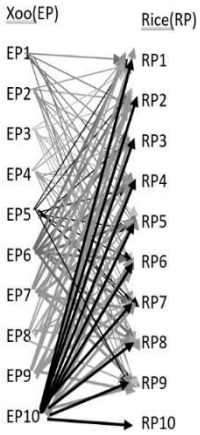
DATA WAS COLLECTED BY NCBI: National Center for Biotechnology Information (nih.gov)

ALPHAFOLD-2

AlphaFold2 is an advanced AI model developed by DeepMind that predicts protein structures with high accuracy. Its application in predicting protein-protein interactions (PPIs) has garnered significant attention in the scientific community. That help to finding the correlation between computational prediction of interface residue protein as alphafold2 is give a high-confidence score. focused on high-confidence predictions to delve deeper into the implications of PPIs, showcasing AlphaFold2's potential in understanding stable interactions. While the study acknowledged limitations of AlphaFold2. As we get 10:10 protein which having most stable we use alphafold2 Api to get the PPI interaction result.

BLAST(*BLAST*(Xoo))

YopJ Family Type III Secretion System
Effector (XopJ):
XopX:
XopQ:
XopL:
Avrbs2:
ABC Transporter ATP-Binding Protein:
Nucleoside Hydrolase:
IS3 Family Transposase:
MFS Transporter (Major Facilitator Superfamily Transporter):
AdeC/AdeK/OprM Family Multidrug Efflux Complex Outer Membrane Factor

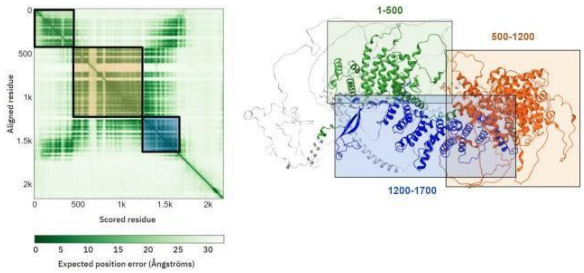


RICE(*ORYZA SATIVA*)

Receptor-Like Kinase (RLK)
Ethylene Receptor 3 Isoform 2 Precursor
Wall-Associated Receptor Kinase-Like 3
Putative Receptor Protein Kinase ZmPK1
WRKY Transcription Factor 30
Calcium-Binding Protein CBP
Subtilisin-Like Protease Pr1B, Partial
Serine/Threonine-Protein Kinase SAPK9
Type-1 Proteins Geranylgeranyl Transferase Subunit Beta
Bidirectional Sugar Transporter SWEET2b

Input protein sequence(s), then hit Runtime -> Run all (query sequence) and jobname	We have to paste our sequence in input sequence (if seq is dimer then protein: protein) Job name: name of that complex PPI interaction
MSA	Multiple sequence Alignment as alphfold2 took data from all the data base and use that database for prediction of protein structure
Advance setting	Tell the amount of cycle use to run for protein structure formation
Run prediction and 3d structure	Get the structure observation of chain of protein and its accuracy
Plots and result	Get the result download as Zip file observation of result

1(receptor) by 1,2,3,4,5,6,7,8,9,10(effector)



Visualization of Predicted Aligned Errors. Protein-specific pages contain an interactive 2D plot of the PAE values. This tool interacts with the 3D molecular viewer to facilitate the identification of domains whose relative positions and orientations Alpha Fold predicts with confidence. In this example (https://alphafold.ebi.ac.uk/entry/Q93074), Alpha Fold has high confidence in the relative position of domains at residues 1–500 (green) and residues 1200–1700 (blue), but not with the region between 500–1200 (orange) nor the C-terminus

API USE IN IT

Alphafold API profile

API STYLES
-

WEBHOOKS
-

AUTHENTICATION

OAUTH PLAYGROUND
-

POSTMAN / INSOMNIA COLLECTIONS
Run in Postman | Run in Insomnia

APP LISTING REQUIREMENTS
-

QUERY LANGUAGE
-

CUSTOM OBJECT SUPPORT
-

ECOSYSTEM
-

API CHANGELOG
-

TUTORIALS
-

GRAPHQL ENDPOINT
-

API REFERENCE
-

SANDBOX ENVIRONMENT
-

SSO / SOCIAL LOGIN
-

API EXPLORER
-

FREE DEVELOPER ACCOUNT
-

CLI
-

GITHUB
https://github.com/deepmind/alphafold

INTEGRATIONS
-

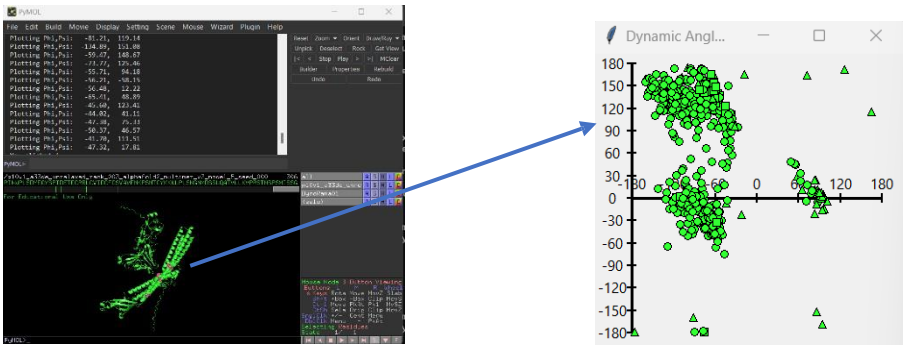
API RATE LIMITS
-

OPEN-SOURCE
-

BASE ENDPOINT
-

ESTIMATED DEVELOPERS
-

Alphafold2 is used in prediction of two protein interaction which have complex structure certainly we have gone through prediction of its protein complex structure and also study about the Ramachandran plot which tell amount the PHI and SHI angle of amino acid residues,



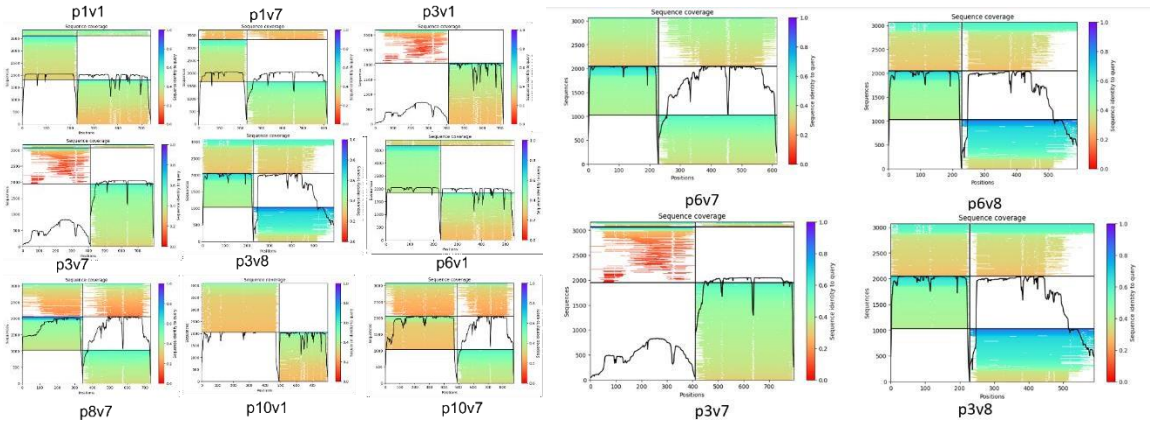
Using Pymol software also this tool help in visulation of predicted protein structure and conversion on PDB to QTPDB format which further use in Molecular docking of protein, BY using Alphafold2 we have done prediction of protein-protein interaction (PPI) we got 100 complex protein structure on which we used to get three output form that result which tell how two complex protein interaction and help in molecular/ cellular mechanism inside the cell in time of infection, we got three output that are used to tell the how these protein have help in making complex.

Xoo (BLAST)

- bidirectional sugar transporter SWEET2b
- xopx
- xopq
- xopl
- avrbs1
- ABC transporter ATP-binding protein [BLASToryzae]
- nucleoside hydrolase [BLASToryzae]
- IS3 family transposase [BLASToryzae]
- MFS transporter [BLASToryzae]
- AdeC/AdeK/OprM family multidrug efflux complex outer membrane factor [BLASToryzae]

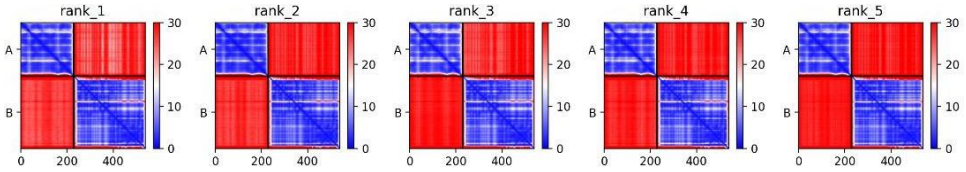
Rice (Oryza sativa)

- receptor like kinase, partial [Oryza sativa Indica Group]
- ethylene receptor 3 isoform 2 precursor [Oryza sativa Japonica Group]
- wall-associated receptor kinase-like 3 [Oryza sativa Japonica Group]
- putative receptor protein kinase ZmPK1 [Oryza sativa Japonica Group]
- TPA_inf: WRKY transcription factor 30
- RecName: Full=Calcium-binding protein CBP; Short=OsCBP
- subtilisin-like protease Pr1B, partial
- type-1 proteins geranylgeranyltransferase subunit beta



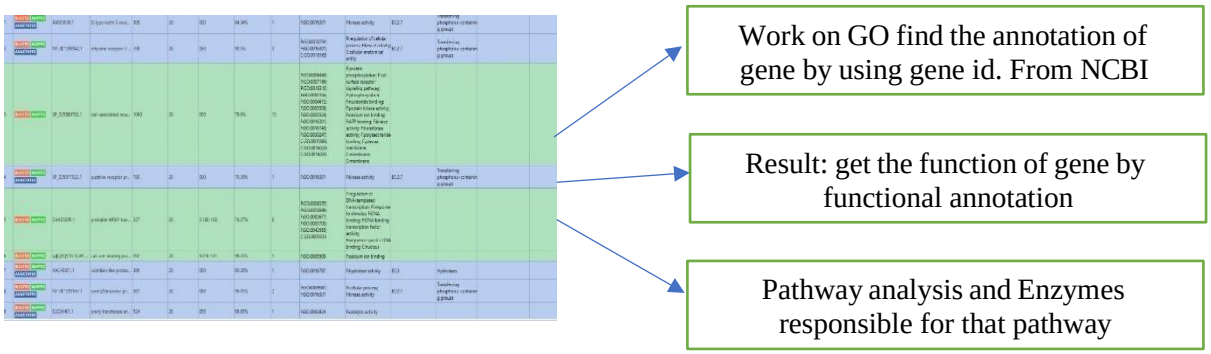
All the graph of one effector and one receptor-based model

- Plddt graph show the error rate value the blue colour shows the amount of accuracy in PPI structure also the site of protein binding
- Generated by google colab pro
- Alpha fold algorithms to predict the interaction and binding sites



BLAST2GO

Blast2GO is a powerful bioinformatics platform specifically designed for functional annotation and analysis of genomic datasets. This user-friendly tool offers a range of features to facilitate the annotation and interpretation of genetic information. It is particularly beneficial for researchers in the field of functional genomics and those working with sequence data that lack Gene Ontology (GO) annotations. As it is extensively used in protein research for functional annotation and analysis, particularly in the context of Gene Ontology (GO) terms and attributes associated with proteins.



One of its key features is the integration of various biological vocabularies, including Gene Ontology Terms. This means that Blast2GO aids in assigning GO terms to proteins, which helps categorize their functions, cellular locations, and involvement in biological processes. By providing a comprehensive view of protein functions based on GO terms, it aids in functional genomics research by automating annotation, visualization, and analysis processes.
Providing insights into the biological roles and functions of genes or sequences. GO terms categorize genes into biological process, molecular function, and cellular component categories. As we get the information of these 20 gene and in which pathway, they are used for we can go for pathway analysis and get to know how these gene responsible for infector of blight disease to

RICE by Xoo and further KEGG pathway use for analysis allows the mapping of enzyme codes and KEGG pathways to sequences, aiding in the understanding of metabolic pathways and enzyme functions. Blast2GO facilitates data mining by offering various graphical and statistical functions. This helps researchers identify patterns of protein in pathway

NETWORK ANALYSIS

Network analysis of proteins involves studying the interactions and relationships between proteins to gain insights into biological processes, pathways, and the molecular mechanisms underlying complex cellular functions. This analysis can be divided into two main aspects: pathway analysis how o protein plays a role protein-protein interaction pathway by this we can get the conformation analysis which are the top protein play a role in defense or and cell-signaling pathway in rice by this process we can get the understanding of interconnecting pathway of biological system that give n output of network of proteins that are functionally related and participate in common cellular functions.

identify the know function of specific pathways or processes and connect these proteins to form a pathway network that represents the flow of molecular events.

then analysis the pathway networks to get the function, process and annotation and determine which pathway it is on the basis of protein present in that for known function.

PROTEIN- PROTEIN DOCKING

Protein-protein docking is a computational method that plays a crucial role in understanding the intricate mechanisms of biological processes. This approach is designed to anticipate the two distinct protein molecules interact and create a stable molecular complex. Considering the immense number of potential ways these molecules can interact, the docking process relies on complex algorithms and scoring functions to strategically confine the exploration to a manageable space, thereby pinpointing plausible binding orientations. The ultimate objective of protein-protein docking is to make predictions regarding arrangement of the complex, which subsequently provides profound insights into the interface of the interaction, the underlying forces dictating the association, and the potential functional implications that arise from this interaction. As my I have done docking of 100 protein-protein docking and get the TMscore and Docking Score to get the conformation how these protein-protein docking (PPD) helps to understand the rate of binding and find the binding site/ binding energy to get the sequence where most of the possible interaction can happened.

Tools use is ZDOCK, HDOCK, HADDOCK

ZDOCK Server: An automatic protein docking server (umassmed.edu)

HADDOCK Web Server (uu.nl)

ClusPro 2.0: protein-protein docking

PROTEIN LIGAND DOCKING

Molecular docking serves as a powerful computational tool in the realm of structural biology, enabling the exploration of intricate interactions between a small molecule, termed a ligand, and a larger molecule, often a protein. As in contest to protein which majorly play role in majorly in infection or disease causing agent we used to inhibit that by the help of Ligands as ligands emerge as key players. When a ligand binds to a protein, it unleashes a cascade of effects, each shaping the protein's behavior in distinct ways. Among these effects, enzyme inhibition stands as a paramount phenomenon. Ligands have the capability to specifically bind to the active sites of enzymes, thereby obstructing their catalytic activity. This binding interaction acts as a molecular blockade, effectively preventing the substrate from binding to the enzyme's active site. The consequence is a process termed enzyme inhibition, a cornerstone of cellular regulation and a crucial mechanism for therapeutic intervention. Also, when we talk about protein-protein interaction Ligands also play a role in it by Ligands possess the capacity to target specific interfaces where proteins form complexes. By intervening at these critical junctures, ligands obstruct the formation of intricate protein-protein interactions, thereby perturbing cellular communication pathway. we have use ligands for inhibition of pathogenic protein cause by Blast to rice as we use some ligands chooses from research paper for protein ligands docking

INHIBITORS LIGANDS	SMILES OF THESE LIGANDS
Tyrphostin	C1=C(C=C(C(=C1O)O)O)C=C(C#N)C#N
Salicylic Acid	C1=CC=C(C(=C1)C(=O)O)O
Antagonists	CC1=CC=C(C=C1)NC(=O)CN2C(=O)C(=CC3=CC=CN3C4=CC=CC(=C4)C(=O)O)NC2=O
Abscisic acid	CC1=CC(=O)CC(C1(C=CC(=CC(=O)O)C)O)(C)C
Gibberellic acid	CC12C(C=CC3(C1C(C45C3CCC(C4)(C(=C)C5)O)C(=O)O)OC2=O)O

Software I have use for my work is AUTODOCK VINA, PYRX, PYMOL

AUTODOCK VINA: Vina is a widely used computational software tool for molecular docking and virtual screening. it is designed to predict the binding modes and affinities of small molecules (ligands) within the binding sites of target proteins.

PYRX: offers molecular docking capabilities, allowing users to predict the binding modes and affinities of ligands within the binding sites of target proteins.by this e can u multiple ligands docking in ingle tools which give accurate binding its and also has user friendly interface for working Direcotry

RESULT

We have taken 50 protein for effector as so 50 protein for receptor from the NCBI data based the region on the basis of three pathway defense pathway, metabolic pathway, and growth pathway they are majorly responsible for immune defense mechanism in rice by various pathogen interaction as by these pathway only the cell generate immune response for its defense related protein xa2,xa6,xa7 are majorly responsible for defense gene in rice by receptor like kinase enzyme RLK was secreted by these xa2, xa6, xa7 gene only here we are having an objective to block the T3SS pathway of Blast T3SS pathway are basically manipulate host immune responsible for these pathogen can easily infect rice and over take plant for growth and development and kills or damage plant.

On the basis of all these three pathways we have chosen 50/50 protein as for further filtration of this protein we have choose two major parameters

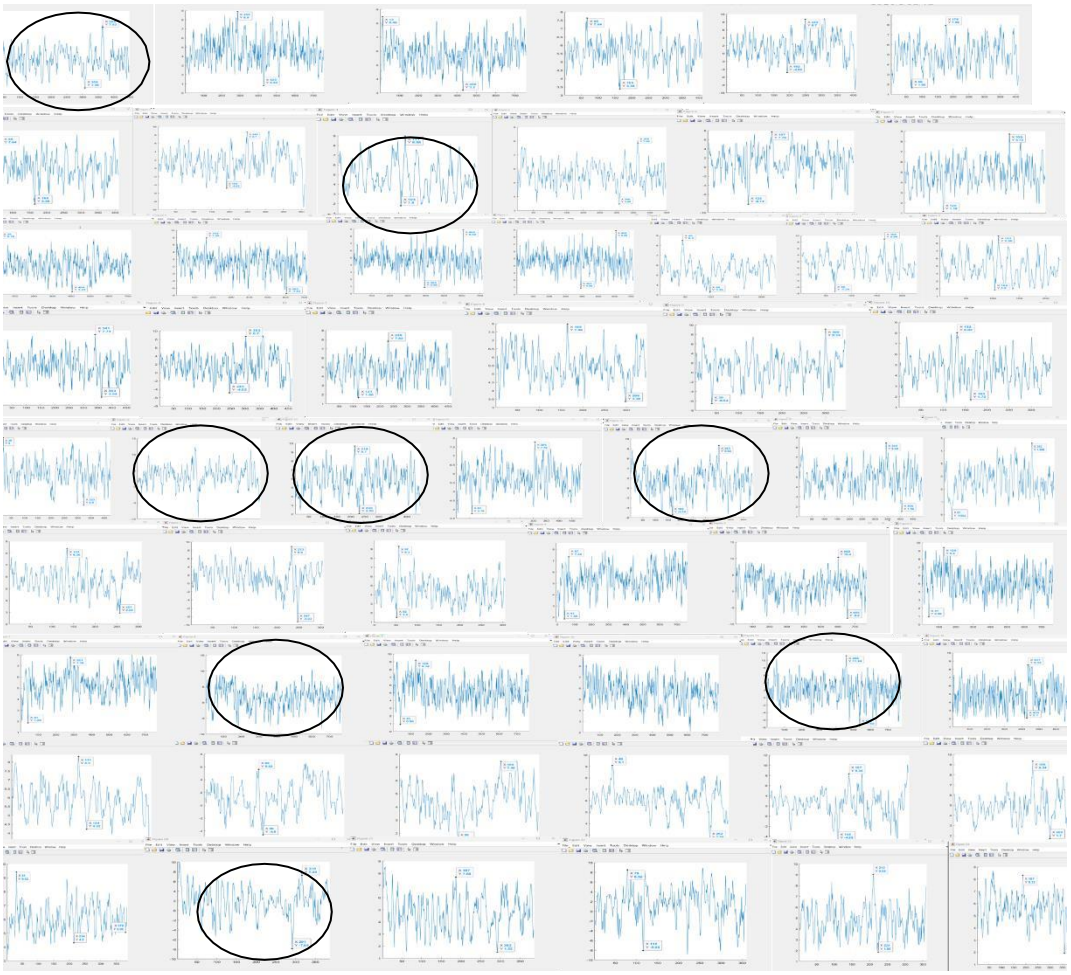
Functionality of protein: find the function of each protein and how they play a role in **T3SS, PAMP, PTI** related pathway also in host signaling defense pathway.

Source – research papers, and **NCBI** data where we the protein also functional annotation

Stability of protein: the stability of protein tells how stabile they are when they bind to nearby protein or ligand or in surrounding environment

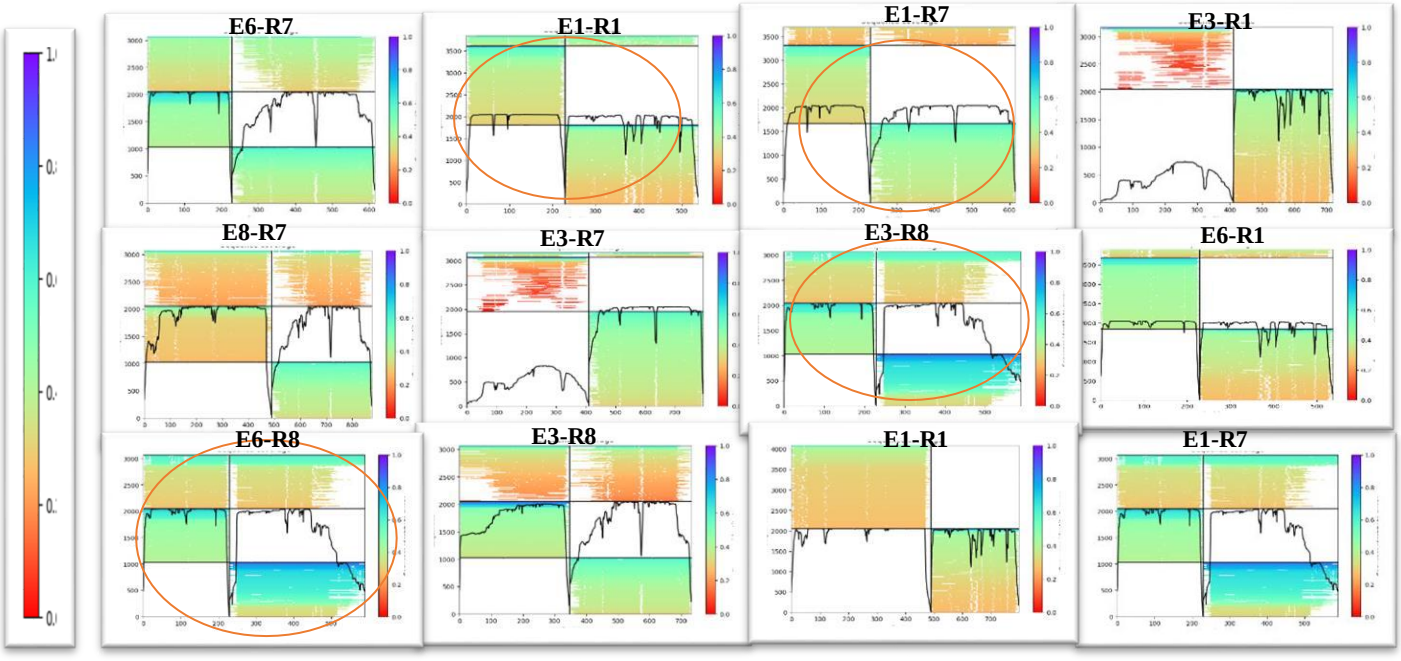
Source: MATLAB software ((%accessible residues, HPLC retention PH 7.4(Meek), Hydrophobicity (Aboderin)

We generate the 60 graph that tell the stability of protein of those three parameters.



ALPHAFOLD-2

In AF2 PPI interaction we got result based on three aim output on which we get the main conclusion that which protein play a major role in plant pathogen interaction in signaling pathway as AF2 is the AI based tools tells the output based on ERROR PLOT(an error plot typically shows the difference between predicted distances (or inter-residue distances) and experimental distances for a given protein),binding of two protein based on C-Terminal and N-terminal and residues(The error values calculated for each residue pair are then plotted on a graph.) also by PLDDT("Per-residue Log Normalized Distance Difference Test") plot help in predicted value as These interactions correspond to regions of the protein where the predicted structure closely matches the experimental data. Higher error values indicate regions where the predicted structure deviates more from the experimental data.



PLDDT: "Per-residue Local Distance Difference Test (1-100) >70 TELL ACCURATE RESULT OF PPI

PLDDT SCORE VALUE OF ALPHAFOLD 2

PPI	RP1	RP2	RP3	RP4	RP5	RP6	RP7	RP8	RP9	RP10
EP1	80	67	69	76	63	65	75	71	68	78
EP2	54	62	61.7	65	60	56	61	58	60.3	59
EP3	81.6	63.1	64.1	76.9	65	68	81	82.5	75	79
EP4	75.2	61	65.4	68	59	63	76	52	68.1	50
EP5	60.2	54	53.2	62	38	42	51	56	57	84
EP6	83	67	70	78	61	65	87.9	84	69	54
EP7	78.1	62	64	71	60	76	81.2	69	69.4	64
EP8	78	64	60	68	61	64	81	79	71	78
EP9	79.8	61	67	74	60	64.5	79.6	72.1	70.7	79
EP10	83	62	74	73	52	84	82.8	75	74	50

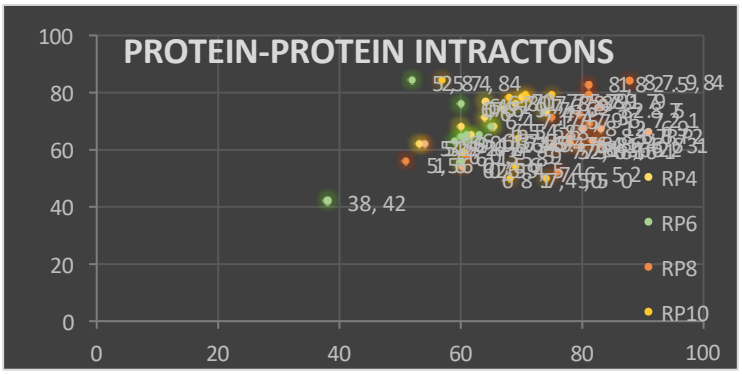
PPI interaction of data by AI generated (alphafold2) by deep mind (google) Interaction of one effector and one receptor at one time and by per-residue confidence score (pLDDT) between 0 and 100 The structural module of AF2 also creates the model confidence predictions, reported as the predicted local distance difference test (pLDDT) scores. The pLDDT scores are in the range of [0, 100]. High pLDDT scores (e.g., > 80) indicate high confidence of the residue structure, and low pLDDT scores (e.g., < 50)

81-90		Highest coverage value
75-81		Mid highest coverage
65-75		Average coverage value
55-65		MID lower coverage value
30-55		Lowest coverage value

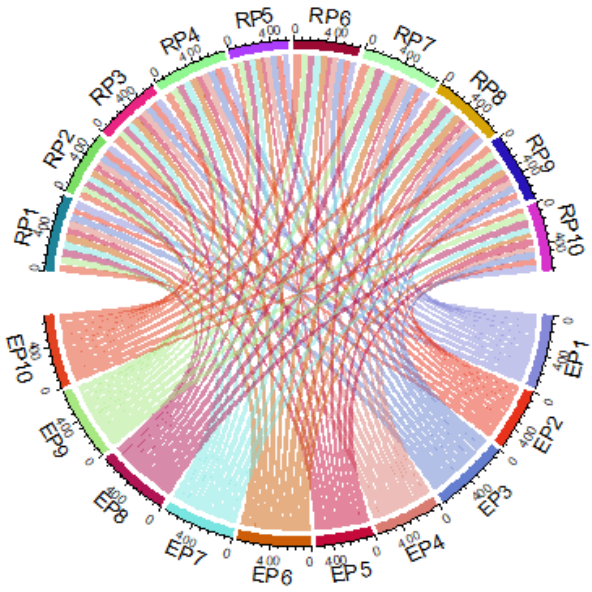
ONE RECEPTOR – ONE EFFECTOR	RP7-EP6 = 87.9, RP8-EP6 = 84, RP1, EP6 = 83, RP1-EP10 = 83, RP10-EP5 = 84	BASED ON COVERAGE TOP COVRAGE CHOOSEN RP7-EP6 = 87.9 RP8-EP6 = 84
ONE RECEPTOR – MANY EFFECTOR	RP7, RP1, RP3, RP8, RP10, RP9	CHOOSE TOP 2 MODEL RP7, RP1
MANY EFFECTOR – ONE RECEPTOR	EP3, EP7, EP6, EP8, EP9	CHOOSE TOP 2 MODEL EP3, EP7

TABLE DATA USED FOR FIGURE

PPI	RF1	RF2	RF3	RF4	RF5	RF6	RF7	RF8	RF9
XOoE1	82	67	69	75.2	58	65	84	73	71
XOoE2	54	62	61.7	65	60	56	61	58	60.3
XOoE3	81.6	63.1	64.1	76.9	65	68	81	82.5	75
XOoE4	75.2	61	65.4	68	59	63	76	52	68.1
XOoE5	60.2	54	53.2	62	38	42	51	56	57
XOoE6	83	67	70	78	61	65	87.9	84	69
XOoE7	78.1	62	64	71	60	76	81.2	69	69.4
XooE8	78	64	60	68	61	64	81	79	71
XOoE9	79.8	61	67	74	60	64.5	79.6	72.1	70.7
XOoE10	83	65	69	73	72	67	82.8	75	74



We generated circose plot that tell the how their protein is interaction by the help of color combination of PLDDT score value also tell that the interaction of two protein as this is the easy way of observation of result or interpretation of result of protein -protein interaction and also tell the interaction between two protein complexes.



Neural network tell the interaction of two different protein and its rate Generated by cirsos software
Table data fig: use to tell the data taken for developing this network

As we are having 3 different objectives on which we have chosen top 1st protein-protein interaction to understand how these interactions play a role in molecular mechanism in rice

TOTAL NUMBER OF INTRACTION:
ONE RECEPTOR-ONE EFFECTOR: 100 / **39**
ONE RECEPTOR-MANY EFFECTORS: 10 / 4
MANY EFFECTORS ONE RECEPTOR: 10 / 6

1. ONE RECEPTOR – ONE EFFECTOR - **RP7-EP6**

(ABC transporter ATP-binding protein - subtilisin-like protease Pr1B, partial

2. ONE RECEPTOR – MANY EFFECTOR - **RP7**

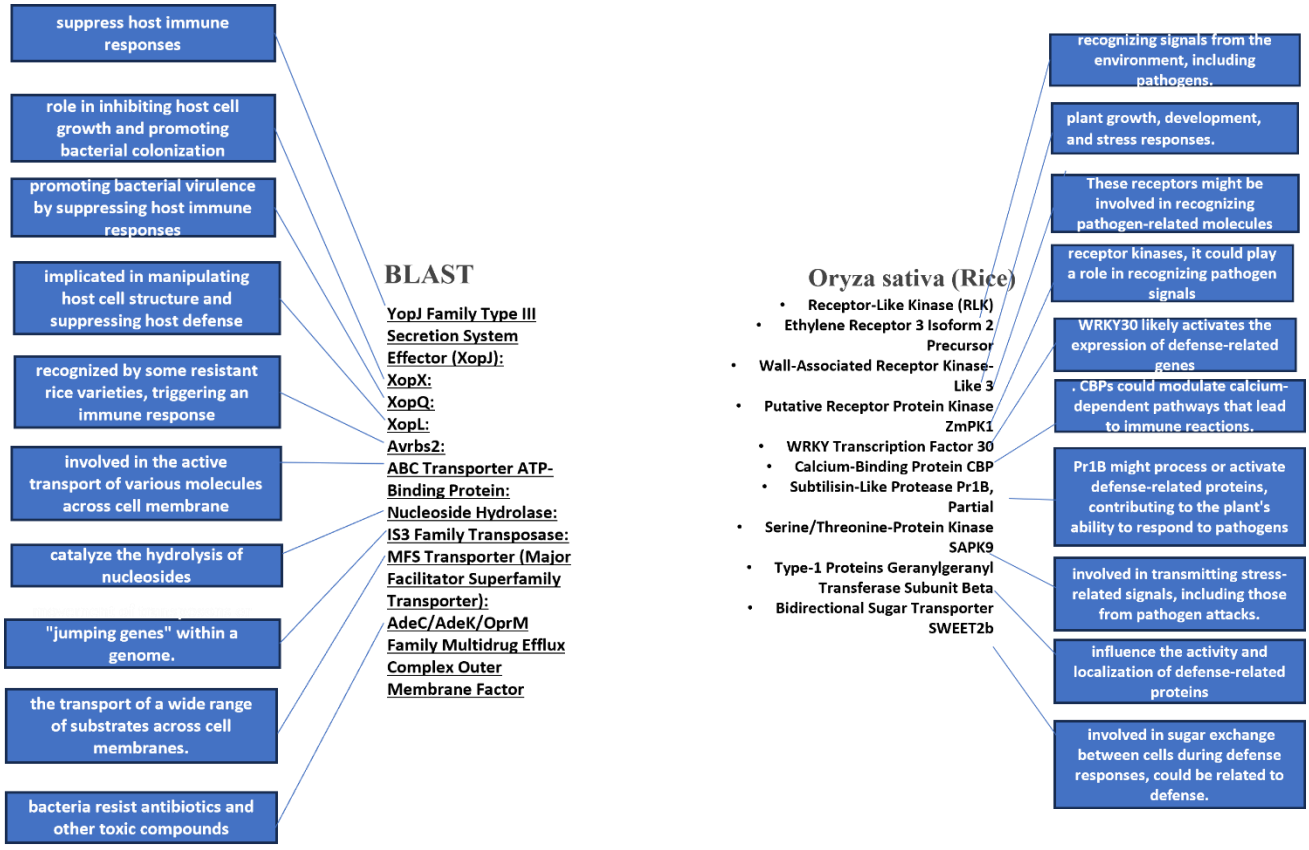
Receptor like kinase, partial

3. MANY EFFECTOR – ONE RECEPTOR - **EP3**

type III secretion system effector XopQ, partial

FUNCTIONAL ANNOTATION AND PATHWAY ANALYSIS

Know we know what are the protein which are majorly responsible for pathogen resistanet and immune response by PPI we got the result as for better understand how these proteins play a role in molecular mechanism inside plant cell we need to know function of each protein and these function of effector play a role in function of receptor in signal pathway for that we can think like which are the major protein responsible for these pathway and also how it is generating immune response.



As we get all the functionality of each protein, we can predict they which protein function is related to which protein fir generating immune response

BLAST 2 GO

FUNCTIONAL ANNOTAION OF ALL THE PROTIN

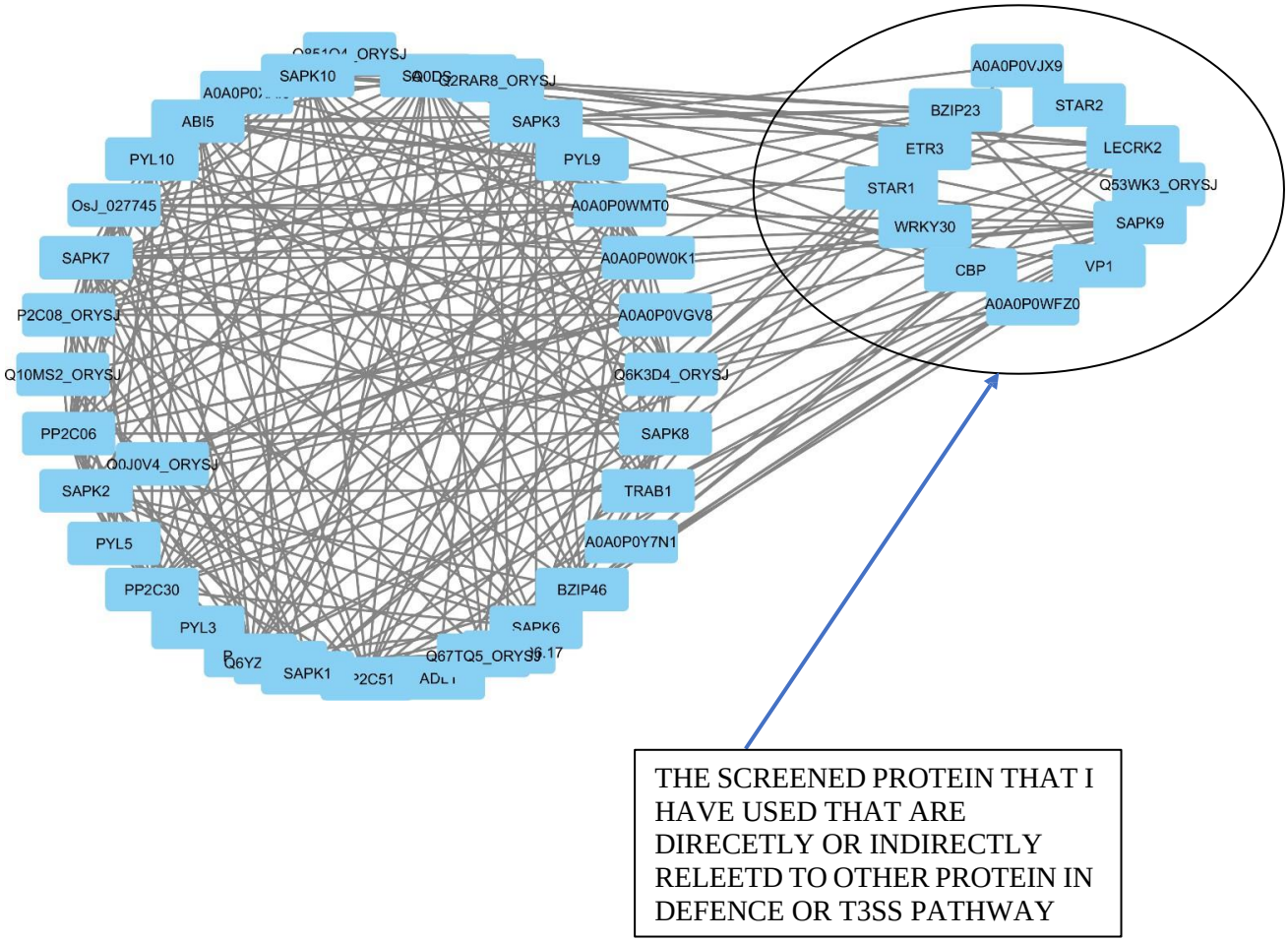
NO. 1	GENE	PROTEIN	FUNCTION	BEHAVIOUR
1	**Xanthomonas oryzae pv. oryzicola strain GX01 chromosome, complete genome**	**XopX family type III secretion system effector [Xanthomonas oryzae]**	Interspecies interaction between organisms Extracellular region Cellular anatomical entity	
2	**Xanthomonas oryzae pv. oryzicola strain GX01 chromosome, complete genome**	**type III effector protein XopL [Xanthomonas oryzae pv. oryzicola]**	Pyrimidine nucleobase metabolic process Cellular aromatic compound metabolic process Nitrogen compound metabolic process Metabolic process Nucleobase metabolic process Cellular process Cellular nitrogen compound metabolic process Cellular metabolic process Primary metabolic process Small molecule metabolic process Heterocycle metabolic process Nucleobase-containing small molecule metabolic process Organic substance metabolic process Pyrimidine-containing compound metabolic process Organic cyclic compound metabolic process Organonitrogen compound metabolic process Catalytic activity Hydrolase activity Hydrolase activity, acting on glycosyl bonds Hydrolase activity, hydrolyzing N-glycosyl compounds Uridine nucleosidase activity Ribosylpyrimidine nucleosidase activity Intracellular Cytoplasm Cytosol Cellular anatomical entity Mostly uncharacterized, incl. triphosphoric monoester hydrolase activity, and uridine nucleosidase activity Mixed, incl. uridine nucleosidase activity, and glycolate metabolic process Purine metabolism Nicotinate and nicotinamide metabolism Metabolic pathways	
3	**Xanthomonas oryzae pv. oryzicola strain GX01 chromosome, complete genome**	**AvrBs1/Avra family type III secretion system effector [Xanthomonas]**	Extracellular region Cell periphery Cellular anatomical entity Mixed, incl. Plant-pathogen interaction, and Coiled coil Mixed, incl. Plant-pathogen interaction, and tyrosine biosynthetic process Mixed, incl. host cell nucleus, and cysteine-type endopeptidase activity Mixed, incl. calcium ion binding, and cysteine-type endopeptidase activity	
4	**Xanthomonas oryzae pv. oryzicola strain GX01 chromosome, complete genome**	**ABC transporter ATP-binding protein [Xanthomonas oryzae]**		
5	**Xanthomonas oryzae pv. oryzicola strain GX01 chromosome, complete genome**	**type III secretion system effector XopQ [Xanthomonas oryzae]**	Transport Response to toxic substance Cellular process Response to chemical Xenobiotic transport Response to antibiotic Response to stimulus Localization Establishment of localization Transmembrane transport Detoxification Export across plasma membrane Export from cell Xenobiotic detoxification by transmembrane export across the plasma membrane Nucleotide binding Catalytic activity Transporter activity Binding ATP binding ATPase-coupled xenobiotic transmembrane transporter activity Primary active transmembrane transporter activity Efflux transmembrane transporter activity Pyrophosphatase activity Hydrolase activity Hydrolase activity, acting on acid anhydrides Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides ATPase activity Purine nucleotide binding Nucleoside-triphosphatase activity	

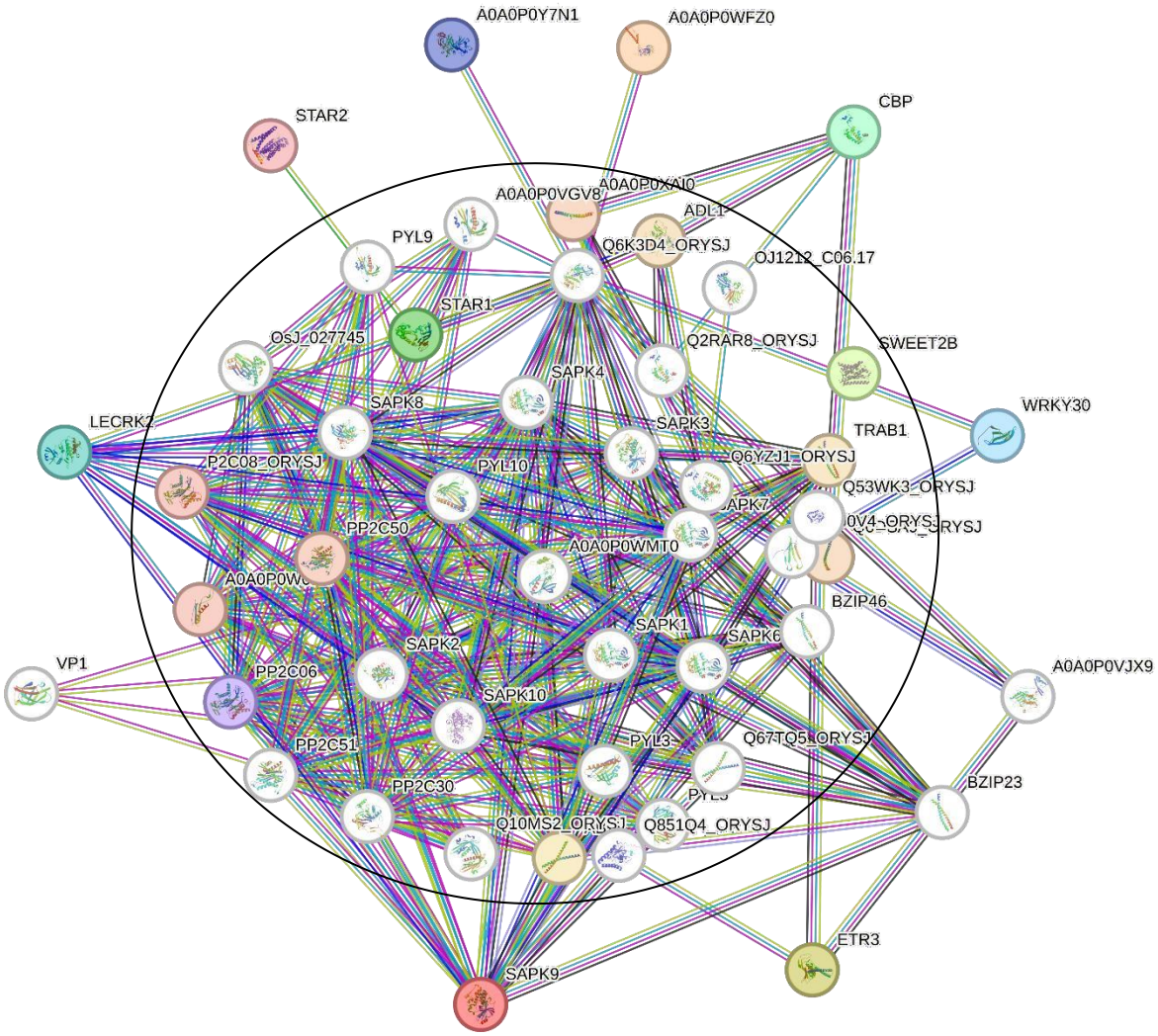
			Active transmembrane transporter activity Transmembrane transporter activity Adenyl nucleotide binding Ribonucleotide binding Purine ribonucleotide binding Adenyl ribonucleotide binding Purine ribonucleoside triphosphate binding Small molecule binding ATPase-coupled transmembrane transporter activity Xenobiotic transmembrane transporter activity Ion binding Anion binding	
6	**Xanthomonas oryzae pv. oryzicola strain GX01 chromosome, complete genome**	**IS3 family transposase [Xanthomonas oryzae]**	Nucleobase-containing compound metabolic process Pyrimidine nucleobase metabolic process Cellular aromatic compound metabolic process Nitrogen compound metabolic process Metabolic process Nucleobase metabolic process Cellular process Cellular nitrogen compound metabolic process Cellular metabolic process Primary metabolic process Small molecule metabolic process Heterocycle metabolic process Nucleobase-containing small molecule metabolic process Organic substance metabolic process Pyrimidine-containing compound metabolic process Organic cyclic compound metabolic process Organonitrogen compound metabolic process Catalytic activity Hydrolase activity Hydrolase activity, acting on glycosyl bonds Hydrolase activity, hydrolyzing N-glycosyl compounds Uridine nucleosidase activity Ribosylpyrimidine nucleosidase activity Intracellular Cytoplasm Cytosol Cellular anatomical entity Mostly uncharacterized, incl. triphosphoric monoester hydrolase activity, and uridine nucleosidase activity	
7	**Xanthomonas oryzae pv. oryzicola strain GX01 chromosome, complete genome**	**MFS transporter [Xanthomonas oryzae]**	Nucleobase-containing compound metabolic process DNA metabolic process Cellular aromatic compound metabolic process Nitrogen compound metabolic process Metabolic process Cellular process DNA integration Cellular nitrogen compound metabolic process Macromolecule metabolic process Cellular metabolic process Primary metabolic process Cellular macromolecule metabolic process Heterocycle metabolic process Organic substance metabolic process Nucleic acid metabolic process Organic cyclic compound metabolic process Nucleic acid binding Binding Organic cyclic compound binding Heterocyclic compound binding Membrane Cell periphery Cellular anatomical entity DNA integration, and Transposition DNA integration	
8	**Xanthomonas oryzae pv. oryzicola strain GX01 chromosome, complete genome**	**AdeC/AdeK/OprM family multidrug efflux complex outer membrane factor [Xanthomonas oryzae]**	Organic acid metabolic process Citrate metabolic process Transport Metabolic process Cellular process Carboxylic acid metabolic process Oxoacid metabolic process Cellular metabolic process Small molecule metabolic process Localization Establishment of localization Transmembrane transport Organic substance metabolic process Tricarboxylic acid metabolic process Transporter activity Secondary active transmembrane transporter activity Symporter activity Active transmembrane transporter activity Transmembrane transporter activity Plasma membrane Membrane Integral component of membrane Intrinsic component of membrane Cell periphery	

NETWORK ANALYSIS

Tell what are the other related protein responsible for protein- protein interaction and role of defense in molecular level of host defense signaling pathway.

Nural network plot





By this we can figure it out which are the related protein play a role in pathway of defense and how the molecular interaction is developing inside the rice pathogen interaction to stimulating the host defense mechanism.

FUNCTION OF PROTEIN

What is the function of these gene and how they are stimulation immune response we talk about two proteins here and how they play a role in developing immune response. By getting function of these protein

Function of ABC Transporter gene

The ABC transporter ATP-binding protein is a component of the ABC (ATP-binding cassette) transporter system involve in transport of wide range of substrate across the cell membrane they a basically found in various organisms, including bacteria like BLAST is a bacterium that causes bacterial blight, a devastating disease in rice plants. The ABC transporter system, including the ATP-binding protein, plays a significant role in its pathogenicity and survival.

Function of ABC Transporter ATP-Binding Protein in BLAST Oryza:

Transport of Substrates-responsible for the energy-dependent binding and hydrolysis of ATP.

Pathogenicity - the ABC transporter system can be involved in exporting virulence factors, toxins, or other molecules that contribute to the bacterium's ability to infect rice plants.

Resistance to Toxic Compounds - ABC transporters can also be involved in exporting toxic compounds or antibiotics out of bacterial cells.

Nutrient Uptake- ABC transporters can also facilitate the uptake of nutrients essential for bacterial growth and survival.

Efflux Pump - ABC transporters can function as efflux pumps, actively removing unwanted compounds from the bacterial cell. This can contribute to bacterial defense against toxic compounds or antibiotics.

Regulation of Gene Expression-The presence of certain ABC transporters can influence the

expression of genes related to bacterial physiology, stress response, and virulence.

ATP binding Domain - The ATP-binding protein contains the ATP-binding domain, which is crucial for the energy-dependent transport process. This domain binds and hydrolyzes ATP, providing the energy needed to move substrates across the membrane.

Function of Receptor-Like Kinase in Oryza sativa Indica:

Signal Transduction: Receptor-like kinases act as receptors for extracellular ligands, such as hormones, peptides, and other signaling molecules. Upon ligand binding, they trigger intracellular signaling cascades that lead to changes in gene expression and cellular responses.

Growth and Development: RLKs are involved in regulating plant growth and development processes, including root development, shoot growth, flower formation, and seed development. They help coordinate growth and adaptation to changing environmental conditions.

Abiotic Stress Response: Receptor-like kinases are crucial for sensing and responding to abiotic stressors such as drought, salinity, temperature fluctuations, and nutrient deficiencies. They activate signaling pathways that enhance the plant's ability to tolerate and adapt to stress.

Biotic Interactions: RLKs play a role in the plant's defense against pathogens and pests. They can recognize pathogen-associated molecular patterns (PAMPs) or damage-associated molecular patterns (DAMPs), triggering immune responses to fend off infections

Hormone Signaling: Some RLKs are involved in hormone signaling pathways, such as brassino steroids, auxins, cytokinin’s, and abscisic acid. These pathways regulate various physiological processes, including growth, development, and stress responses.

Cell Wall Integrity Sensing: Certain RLKs are involved in monitoring cell wall integrity. They

detect changes in the physical properties of the cell wall, triggering cellular responses to maintain structural integrity.

Reproductive Processes: Receptor-like kinases are implicated in reproductive processes such as pollen development, fertilization, and seed formation. They play a role in ensuring successful reproduction under varying conditions.

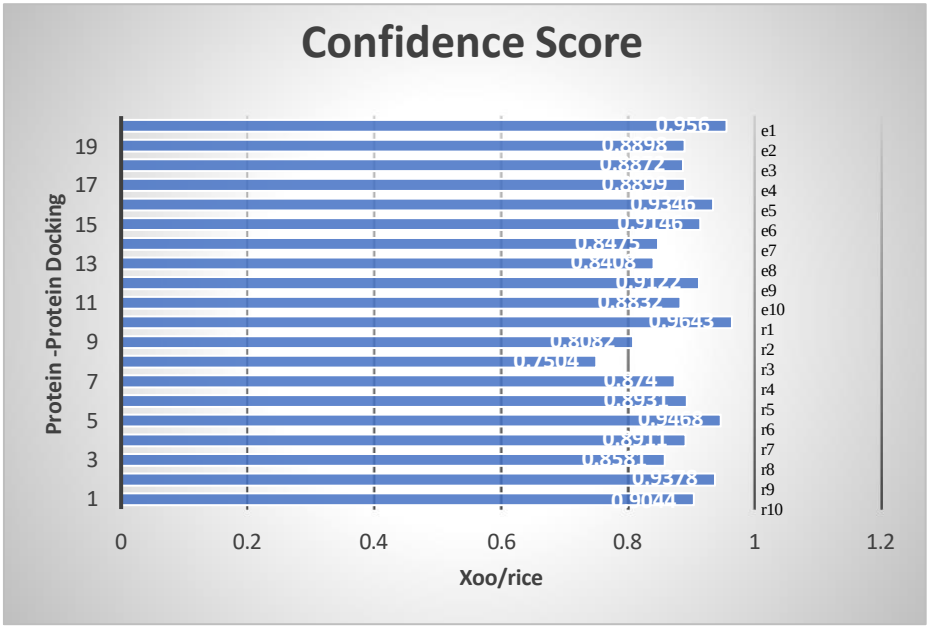
Nutrient Uptake and Homeostasis: Some RLKs participate in nutrient sensing and uptake. They help regulate nutrient homeostasis by responding to changes in nutrient availability and distribution.

Secondary Metabolism: RLKs are linked to the regulation of secondary metabolites, including flavonoids, terpenoids, and alkaloids. These compounds play roles in defense, attraction of pollinators, and adaptation to environmental challenges.

Symbiotic Interactions: In addition to pathogenic interactions, RLKs are involved in forming symbiotic relationships, such as mycorrhizal associations and nodulation in legumes.

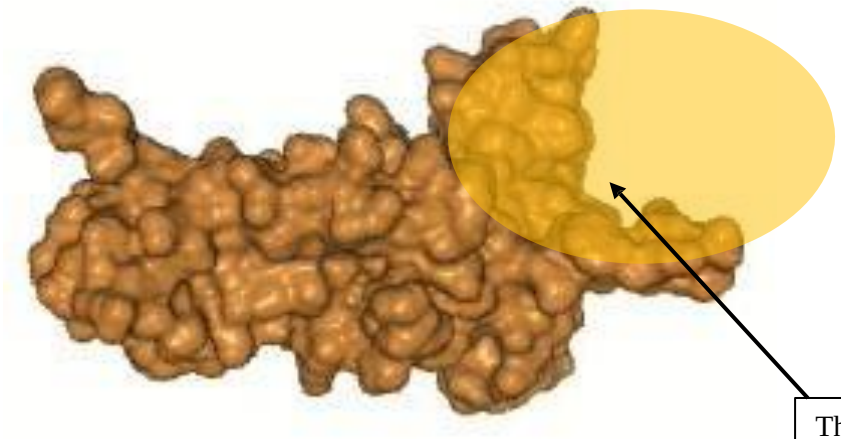
PROTEIN-PROTEIN DOCKING

Protein-protein docking is a computational technique that serves as a powerful tool that help to find the 3D complex structure of protein complex docking lies in its capacity to predict potential binding sites on proteins. It serves as a molecular detective, uncovering regions where interactions between proteins are likely to occur. These binding sites hold the key to understanding the molecular "handshake" between proteins, shedding light on the critical junctures where they engage in their intricate biochemical as further, we need to find the complex signaling pathways, which orchestrate the cellular responses that underpin life protein-protein interactions are the threads that weave the intricate tapestry of signal transduction. Protein-protein docking emerges as a multidimensional tool that bridges the gap between structure and function in the intricate world of molecular interactions. By deciphering binding modes, predicting complex structures, identifying binding sites, and estimating binding affinities, it unveils the molecular intricacies that govern cellular processes.



BINDING SITES

I find the binding sides of top three highest score protein that play a multiple receptor binding as this region of protein cause higher transmission of pathogenesis and also this may cause suppressed an immune response of rice by T3SS pathway where PAMPs bind to PRR by using PTI the rice secrete immune response but as T3SS has the capability to manipulate the defense mechanism of rice to cause infection so I find the protein that is majorly involve in this process and cause the infection of rice as that region is having multiple binding sites so if we bloc one side of receptor it can bind with another receptor to cause infection, in rice. So for that we can find the binding region which is most responsible for this infection and cause disease and inhibit that side by use of ligand in further process.



Effector

The most of the protein binding occur in this portion of protein as it is the most stable site of the protein that can also help further analysis for modification of protein to inhibition of pathogenesis

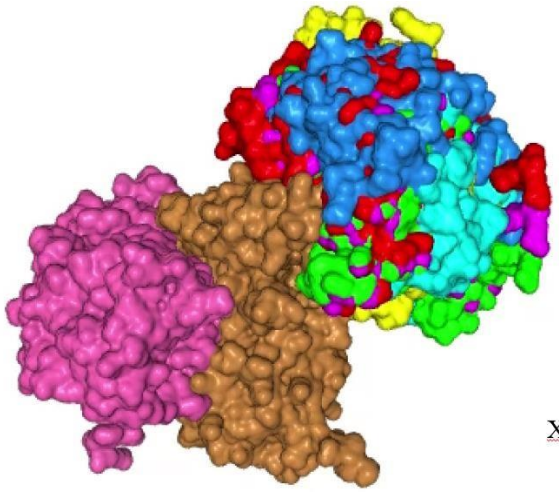
Query: chain: xopqchimera.pdb #1/A, database: pdb, cutoff: 0.001, maxSeqs: 100, matrix: BLOSUM62, version: 4

Hit #	Name	E-Value	Score	Title	Resolution	Ligand Symbols	Chain Names	# Residues
1	7JLU_B	0.0	785	Structure of the...	3.8	CA	A: Disease ...	1805
2	4P5F_A	0.0	764	The crystal ...	2.1	CA, AR6	A,B: Inosine-...	766
3	4P5F_B	0.0	764	The crystal ...	2.1	CA, AR6	A,B: Inosine-...	766
4	4KL0_A	0.0	755	Crystal structur...	1.598	CA	A: Putative ...	379

Bp2 [1] [ID: bp2 [1]]

```
#1/A RSQRAQLAKGVFDRLALPDVRVARGQDYPMTSTQAREHSKFLAEGAALRA
7JLU_B RSQRAQLAKGVFDRLALPEVRVARGQDYPMTSTQAREHSKFLAEGAALRA

          151       161       171       181       191
Conservation
#1/A APDAVHTDGVRAMCERLATSPHKLGMVVIAGMTDASALLAEAGDLVREKV
7JLU_B APDAVHTDGVRAMRERLATSPHKLGMVVIAGMTDASALLAEAGDLVREKL

          201       211       221       231       241
Conservation
#1/A ASITIMGGIDPARDADGLVQPDTRAYNNATDIHAARALYRRAQQLGIPLR
7JLU_B ASITIMGGIDPARDADGLVQPDTRAYNNATDIHAARALYRRAQQLGIPLR

          251       261       271       281       291
Conservation
#1/A ILTKEAAYKAAVPPAFYEGIARNGHPVGEYLRDVQKNALKGLWEGIQANL
7JLU_B ILSKEAAYRAAVPPAFYEGIARNGHPVGEYLRDVQKNALKGLWEGIQANL

          301       311       321       331       341
Conservation
#1/A IPGLDTAWFFRTFVAAQPQDPVAADQQGALSFDAIWPQVTKLNLYDPLTL
7JLU_B IPGLDTAWFFRTFVAAQPQDPAAADQQGAMSFDAIWPQVTKLNLYDPLTL

          351       361       371       381       391
Conservation
#1/A LAALPGTARLLFQPTPMHREGASPVEHVGHAEVVRPEKARLLLSALAKAA
7JLU_B LAALPGAARLLFQPTPMHREGASPVEHVGHAEVVRPEKARLLLSALAKAA
```

Receptor



XopQ

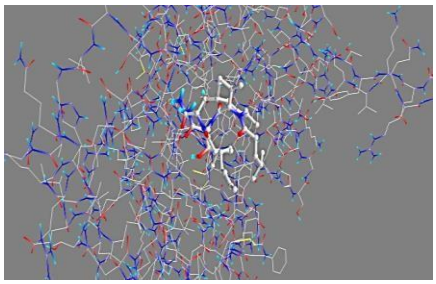
PLDDT SCORE: - 81.6
Confidence score: 0.9402
Docking score: -287.74

PROTEIN LIGAND DOCKING

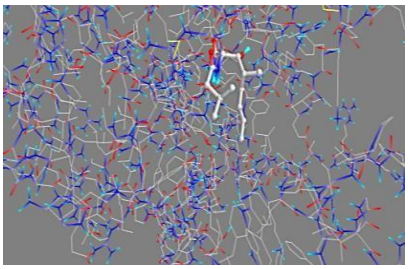
We got the effector protein and we got the function by that we get the conformation predation that the most pathogenic protein and that cause disease in rice by these three protein so here we are using inhibitor ligands which are nontoxic to plant also can be use for inhibition of protein as these protein majorly play a role in pathogen as they are having multi protein binding properties by there binding sites also they can cause infection by binding with any other protein so inhibition of these protein is important for survival of plant.

INHIBITORS LIGANDS	SMILES OF THESE LIGANDS
Tyrphostin	C1=C(C=C(C(=C1O)O)O)C=C(C#N)C#N
Salicylic Acid	C1=CC=C(C(=C1)C(=O)O)O
Antagonists	CC1=CC=C(C=C1)NC(=O)CN2C(=O)C(=CC3=CC=CN3C4=CC=CC(=C4)C(=O)O)NC2=O
Abscisic acid	CC1=CC(=O)CC(C1(C=CC(=CC(=O)O)C)O)(C)C
Gibberellic acid	CC12C(C=CC3(C1C(C45C3CCC(C4)(C(=C)C5)O)C(=O)O)OC2=O)O
Ethylene	C=C
Calmodulin	CC(=O)OC1=C2C=C(CC3C4=C(C(=C(C=C4CCN3C)OC)OC)OC5=C(C=C6CCN(C(C6=C5)CC7=CC=C(O2)C=C7)C)OC)C=C1

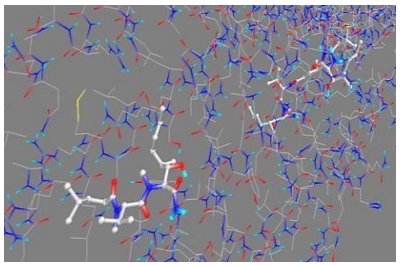
Based on Binding energy we get the result of interaction of ligands and proteins lesser the binding energy more higher the binding rate is in research paper binding energy should be less than -7.0 so the interaction should be accurate



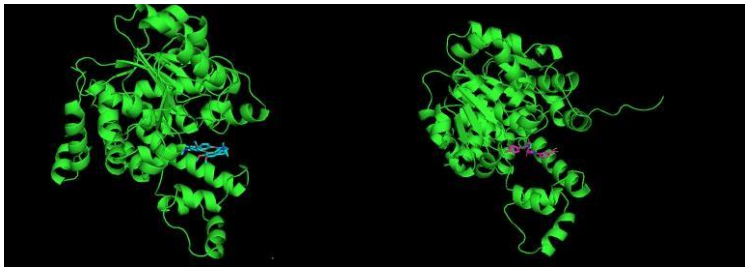
ABC Transporter ATP-Binding Protein



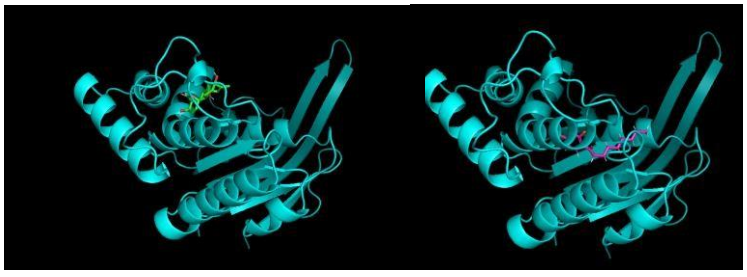
YopJ Family Type III Secretion System Effector xopq



Receptor Like Kinase



ABC Transporter ATP-Binding Protein



YopJ Family Type III Secretion System Effector xopq

Ligand	Binding Affinity	rmsd/ub	rmsd/lb
COMPND 6325	-9.5	0	0
COMPND 2061	-8.9	3.81	2.943
COMPND 5280896	-8.2	18.831	15.684
COMPND 5281166	-10.3	0	0
COMPND 6325	-10.2	6.892	2.418
COMPND 338	-10.1	9.323	2.473
COMPND 2061	-10.1	8.279	3.198

BEHAVIOR OF PROTEIN (MOLECULAR DYNAMICS (MD))

Molecular dynamics (MD) simulations of proteins, such as SAPK9, are essential for studying their real-time behavior and properties. To conduct these simulations, we created a virtual environment, mimicking the protein's surroundings, and utilized software tools like GROMACS (https://github.com/gromacs/gromacs) for simulating atomic movements. Output results included trajectory files, revealing the protein's dynamic behavior, and identification of binding sites relevant to Xoo effectors. These binding sites were analyzed using AlphaFold2 (https://github.com/deepmind/alphafold), which aided in understanding the protein's multiple binding sites. Additionally, MD simulations helped validate structural predictions and explore protein-ligand interactions, ultimately providing valuable insights into protein dynamics, stability, and molecular mechanisms.

THE TOOLS AND PACKAGES USED FOR MD

GROMACS (http://www.gromacs.org/): A widely used molecular dynamics simulation package known for its efficiency and performance in simulating large biomolecular systems.

AMBER (https://ambermd.org/): A suite of programs for performing classical molecular dynamics simulations, focusing on biomolecules.

NAMD (https://www.ks.uiuc.edu/Research/namd/): A parallel molecular dynamics simulation package designed for high-performance simulations of large biomolecular systems.

CHARMM (https://www.charmm.org/): A versatile molecular modeling and simulation package with a focus on biomolecular systems.

LAMMPS (https://lammps.sandia.gov/): A powerful and highly customizable molecular dynamics simulator capable of simulating a wide range of materials and systems.

Desmond (https://www.schrodinger.com/desmond): A molecular dynamics package developed

by Schrödinger for simulating biological systems, especially for drug discovery.

VMD (https://www.ks.uiuc.edu/Research/vmd/): Visual Molecular Dynamics, a molecular visualization and analysis program used to visualize MD simulation results.

PyMOL (https://pymol.org/): A molecular visualization system that allows users to create and visualize 3D molecular structures.

AlphaFold2 (https://github.com/deepmind/alphafold): A deep learning-based tool developed by DeepMind for protein structure prediction, which can be used to provide initial protein structures for MD simulations.

OpenMM (http://openmm.org/): A high-performance molecular dynamics library designed for a wide range of simulations.

MDAnalysis (https://www.mdanalysis.org/): A Python library for the analysis of molecular dynamics simulations.

Bio3D (http://thegrantlab.org/bio3d/): A package for the analysis of structural and functional properties of biomolecules.

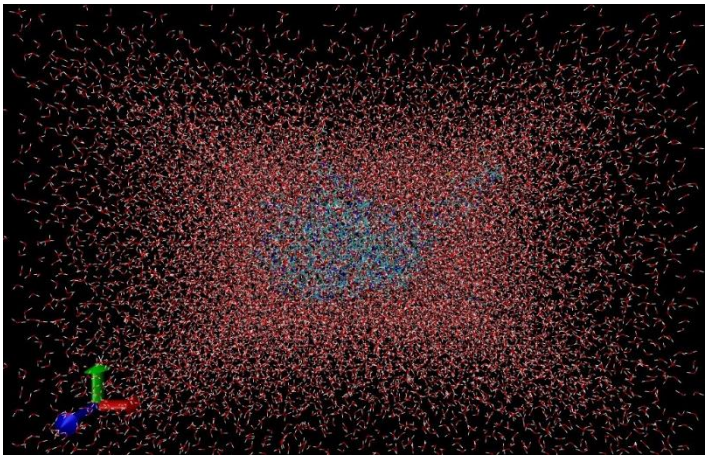
PLUMED (https://www.plumed.org/): A plugin for enhancing MD simulations with advanced analysis and biasing methods.

Schrodinger Suite (https://www.schrodinger.com/): A comprehensive suite of software tools for molecular modeling and simulations, including Desmond.

CHARMM-GUI (http://www.charmm-gui.org/): An interface for generating input files and systems for CHARMM simulations.

MOLECULAR DYNAMICS

Molecular dynamics (MD) simulations are a way to study how a protein, like the highly pathogenic SAPK9 in plants, behaves in real time. To do this, you create a virtual environment, mimicking the protein's surroundings, and then let the computer model the movements of individual atoms. During the simulation, In our analyzed, we used SAPK9 protein which is the protein that has the highest affinity of all the effectors of Xoo by Alphafold2 analysis these binding sites are caused by triggering infecting by Xoo(Effector) by this analysis understanding the nature of a protein that has multiple binding sites. You can see how its structure changes when it binds with different substances, understanding the dynamic nature of this process. This is crucial because SAPK9's interactions with multiple effectors play a role in its pathogenicity. You can also examine how flexible the protein is, which parts move the most, and what kind of energy changes occur during binding.



20ns/50000
SAPK9 protein (blue) surrounded by ≈ >15,000 water molecules (oxygen atoms are red and hydrogen atoms are white). The simulation system consists of ≈50,000 atoms, including potassium and chloride ions (purple and orange spheres, respectively).

The parameters for choosing a Protein/Ligand for Analysis

Protein Structure: Start with a high-quality protein structure obtained from experimental data (e.g., X-ray crystallography or NMR) or computational modeling (e.g., AlphaFold2 predictions).

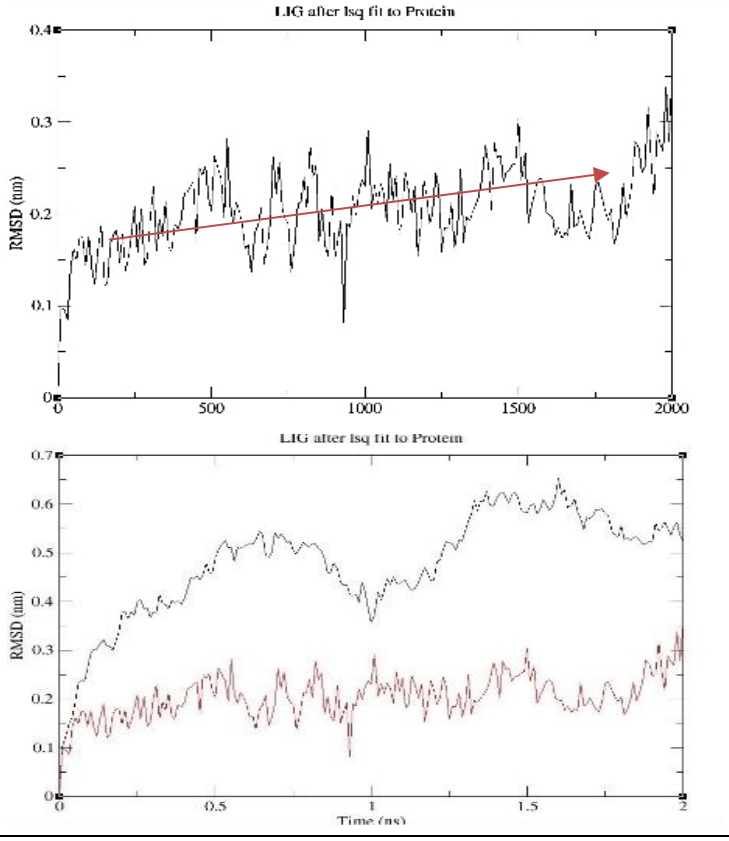
Chain and Segments: Choose the specific protein chains and segments to include in the simulation if your protein has multiple subunits or domains.

Protein Force Field Parameters: Assign appropriate force field parameters for the protein, specifying atom types, charges, and connectivity

Ligand Parameters: Assign force field parameters for the ligand, ensuring accurate representation

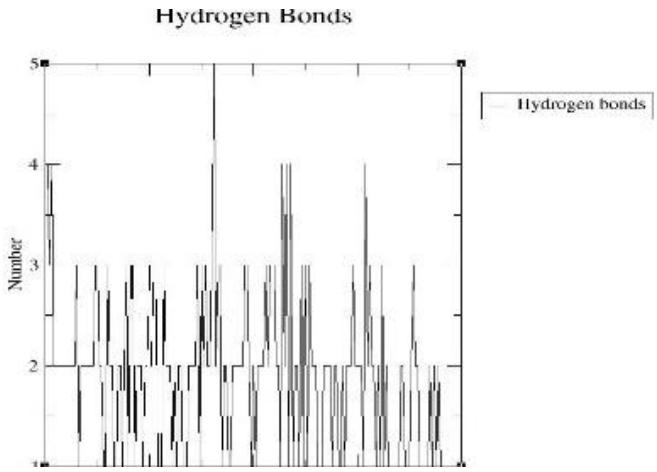
of atomic interactions.

In preparing a molecular dynamics (MD) simulation, selecting the protein, ligand, and water molecules involves choosing an experimental or computational protein structure, specifying protein chains and segments, assigning force field parameters for both the protein and ligand, and obtaining a 3D structure of the ligand. Additionally, water molecules should be modeled using a chosen water model (e.g., TIP3P) with corresponding parameters and solvent density. The box size must be determined for the simulation box, and ions, if needed, should be included to neutralize charges or achieve the desired ionic strength. Solvation tools like VMD, PyMOL, or **CHARMM-GUI** can assist in solvating the protein-ligand complex. A subsequent energy minimization and equilibration phase is essential to establish a stable configuration. Simulation parameters, such as temperature, pressure, time step, and integration algorithm, are set for controlling system dynamics. Periodic boundary conditions are applied for simulating an infinitely repeating system. Proper selection and parameterization of the system components and experimental data are crucial for generating reliable results in MD simulations.
These simulations provide a detailed, atomic-level view of SAPK9's behavior, helping researchers design strategies to inhibit its pathogenic effects or develop treatments for plant diseases caused by pathogens like Xoo.



Fig, RMSD can be used to identify large changes in protein structure as compared to the starting point. A levelling off or flattening of the RMSD curve can also indicate that the protein has equilibrated. The higher the motion of plot the higher the chain of protein will be in motion

RMSF is a calculation of individual residue flexibility, or how much a particular residue moves (fluctuates) during a simulation. RMSF per residue is typically plotted vs. residue number, and can indicate



Hydrogen bonds facilitate molecular interactions and are ubiquitous in nature. The bond between a H on a water molecule and an O on another water molecule is resposible for the cohesive property of water. H-bonds also play a big role in the interaction of ligands and proteins. *Hydrogen bonds can stabilize inhibitors in active sites*

HYPOTHESIS

In my research, I developed the AlphaFold2 model to investigate protein-protein interactions (PPI) between rice and BLAST proteins. Through this model, I identified a complex network of interactions, including scenarios where one receptor binds to one effector, one effector binds to multiple receptors, and one receptor interacts with numerous effectors. These intricate interactions have provided valuable insights into the molecular mechanisms underlying immune responses in rice.

Based on my findings, your hypothesis proposes that the immune response in rice is orchestrated through a sophisticated network of protein interactions. This network involves receptors recognizing effectors from Xanthomonas, triggering a series of events that initiate immune signaling pathways. These signaling pathways then activate various defense mechanisms within rice cells, such as the production of antimicrobial compounds, induction of pathogenesis-related proteins, and reinforcement of the cell wall.

By conducting functional annotations, I gained a deeper understanding of the roles of specific proteins within this network. These proteins are crucial in relaying the signal from the initial protein interactions to downstream components of immune signaling pathways. This orchestration ensures the coordinated and targeted activation of immune responses to combat BLAST infection

We also get the protein that help to play Role in immune response to developed the signaling pathway if that protein is inhibited by the Blast pathogen by effector it get reform himself and rebind or other effector to produce the cell signaling if one receptor is not present at the time of pathogen invasion we have identified all the kind of receptor which can bind to every protein also can stimulating immune response

List of protein that we identified for our hypothesis:

1) **Receptor-Like Kinase (RLK)**
2) **Subtilisin-Like Protease Pr1B, Partial**
3) **Wall-Associated Receptor Kinase-Like 3**
4) **Serine/Threonine-Protein Kinase SAPK9**

DISCUSSION

The model we have chosen is Alphafold4 why we have chosen this model what are the resect and advancement the AI having, as this AI-ML is basically released in 2020 November this tool is very successful in biotechnology also in genomic and proteomics field as it has the highest accuracy all the tools in up to date as this tool basically used for PPI as many research paper came out on the basis of this tool that have predicted PPI of human protein of 60,000 protein complex,

Why we have chosen rice and Blast: as the rice and blast is the most pathogenic disease as this the best model to get to understand what really happing in the molecular level of cell, pathogenies interaction

Pipeline: this pipeline we have work on it the novel as we get all the information of research paper also. We generated this pipeline on the basis of molecular level and bioinformatic understanding for the analysis of protein to use all computational method of protein analysis

Why effector and receptors. they play a key role in cell signaling pathway and all the cell to pathogen interaction only done by this these proteins involve in various pathway as well as understand how plant response in the infection phase.

Conclusion: understanding how plant response in the defense phase also what the major protein involves in pathogen interaction in rice and blast, we get the major protein which bind to multiple effectors to stimulate the immune response in host also get the function of these protein we understand the molecular machines of all these pathways and how the response occurs in mode of protein pathogen interaction.

REFRENCES

1. Liddington, R. C. Structural Basis of Protein–Protein Interactions. *Protein-Protein Interactions* 261, 3–14 https://doi.org/10.1385/1-59259-762-9:003 (2004).
2. Keskin, O., Gursoy, A., Ma, B. & Nussinov, R. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem. Rev.* 108, 1225–1244 (2008).
3. Nooren, I. M. A. NEW EMBO MEMBER'S REVIEW: diversity of protein-protein interactions. *EMBO J.* 22, 3486–3492 (2003).
4. Cong, Q., Anishchenko, I., Ovchinnikov, S. & Baker, D. Protein interaction networks revealed by proteome coevolution. *Science* 365, 185–189 (2019).
5. Zhang, Q. C. et al. Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* 490, 556–560 (2012).
6. Marshall, G. R. & Vakser, I. A. Protein-Protein Docking Methods. In *Proteomics and Protein-Protein Interactions* (ed. Waksman, G.) 115–146 (Springer, 2005).
7. Kundrotas, P. J., Zhu, Z., Janin, J. & Vakser, I. A. Templates are available to model nearly all complexes of structurally characterized proteins. *Proc. Natl Acad. Sci. USA* 109, 9438–9441 (2012).
8. Porter, K. A., Desta, I., Kozakov, D. & Vajda, S. What method to use for protein–protein docking? *Curr. Opin. Struct. Biol.* 55, 1–7 (2019).
9. Halperin, I., Ma, B., Wolfson, H. & Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* 47, 409–443 (2002).
10. Shammas, S. L. et al. Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies. *J. Biol. Chem*, 291, 6689–6695 (2016).
11. Eginton, C., Naganathan, S. & Beckett, D. Sequence-function relationships in folding upon binding. *Protein Sci.* 24, 200–211 (2015).
12. Andrusier, N., Mashiach, E., Nussinov, R. & Wolfson, H. J. Principles of flexible protein-protein docking. *Proteins* 73, 271–289 (2008).
13. Kurkcuoglu, Z. & Bonvin, A. M. J. J. Pre- and post-docking sampling of conformational changes using ClustENM and HADDOCK for protein-protein and protein-DNA systems. *Proteins* 88, 292–306 (2020).
14. Lensink, M. F. et al. Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment. *Proteins* 87, 1200–1221 (2019).
15. Vreven, T. et al. Updates to the integrated protein-protein interaction benchmarks: docking benchmark version 5 and affinity benchmark version 2. *J. Mol. Biol.* 427, 3031–3041 (2015).
16. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021).
17. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876 (2021).
18. Kandathil, S. M., Greener, J. G., Lau, A. M. & Jones, D. T. Ultrafast end-to-end protein structure prediction enables high-throughput exploration of uncharacterised proteins. *Proc. Natl Acad. Sci. USA* 119, e2113348119 (2022).

19. Chowdhury, R. et al. Single-sequence protein structure prediction using language models from deep learning. Preprint at *bioRxiv* https://doi.org/10.1101/2021.08.02.454840 (2021).
20. Procaccini, A., Lunt, B., Szurmant, H., Hwa, T. & Weigt, M. Dissecting the specificity of protein-protein interaction in bacterial two-component signaling: orphans and crosstalks. *PLoS ONE* 6, e19729 (2011).
21. Weigt, M., White, R. A., Szurmant, H., Hoch, J. A. & Hwa, T. Identification of direct residue contacts in protein-protein interaction by message passing. *Proc. Natl Acad. Sci. USA* 106, 67–72 (2009).
22. Hashemifar, S., Neyshabur, B., Khan, A. A. & Xu, J. Predicting protein–protein interactions through sequence-based deep learning. *Bioinformatics* 34, i802–i810 (2018).
23. Yang, J. et al. Improved protein structure prediction using predicted inter-residue orientations. Preprint at *bioRxiv* https://doi.org/10.1101/846279 (2019).
24. Pozzati, G. et al. Limits and potential of combined folding and docking using PconsDock. *Bioinformatics* 38, 954–961 (2021).
25. Lamb, J. & Elofsson, A. pyconsFold: a fast and easy tool for modelling and docking using distance predictions. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btab353 (2021).
26. Szurmant, H. & Weigt, M. Inter-residue, inter-protein and inter-family coevolution: bridging the scales. *Curr. Opin. Struct. Biol.* 50, 26–32 (2018)
27. Jun, S.-R., Sims, G. E., Wu, G. A. & Kim, S.-H. Whole-proteome phylogeny of prokaryotes by feature frequency profiles: an alignment-free method with optimal feature resolution. *Proc. Natl Acad. Sci. USA* 107, 133–138 (2010). An interesting whole-proteome phylogeny study of prokaryotes that shows the Xanthomonadaceae family profiles clustered with the Betaproteobacteria.
28. Parkinson, N. et al. Phylogenetic analysis of *BLAST*species by comparison of partial gyrase B gene sequences. *Int. J. Syst. Evol. Microbiol.* 57, 2881–2887 (2007). Highlights the complexity of the *BLAST*spp. and details useful genetic tools for species discrimination.
29. Dar, G. H., Anand, R. C. & Sharma, P. K. Genetically engineered microorganisms to rescue plants from frost injury. *Adv. Biochem. Eng. Biotechnol.* 50, 1–19 (1993).
30. Ryan, R. P. et al. Passing GO (gene ontology) in plant pathogen biology: a report from the *BLAST*Genomics Conference. *Cell. Microbiol.* 11, 1689–1696 (2009).
31. Comas, I., Moya, A., Azad, R. K., Lawrence, J. G. & Gonzalez-Candelas, F. The evolutionary origin of Xanthomonadales genomes and the nature of the horizontal gene transfer process. *Mol. Biol. Evol.* 23, 2049–2057 (2006).
32. Lima, W. C., Paquola, A. C. M., Varani, A. M., Van Sluys, M. A. & Menck, C. F. M. Laterally transferred genomic islands in Xanthomonadales related to pathogenicity and primary metabolism. *FEMS Microbiol. Lett.* 281, 87–97 (2008).
33. Pieretti, I. et al. The complete genome sequence of *BLASTalbilineans* provides new insights into the reductive genome evolution of the xylem-limited *Xanthomonadaceae*. *BMC Genomics* 10, 1471–1475 (2009).
34. Darrasse, A. et al. Transmission of plant-pathogenic bacteria by nonhost seeds without

induction of an associated defense reaction at emergence. *Appl. Environ. Microbiol.* 76, 6787–6796 (2010).

35. Cazalet, C. et al. Analysis of the *Legionella longbeachae* genome and transcriptome uncovers unique strategies to cause legionnaires' disease. *PLoS Genet.* 6, e1000851 (2010).
36. Lima, W. C., Van Sluys, M. A. & Menck, C. F. M. Non-gamma-proteobacteria gene islands contribute to the *BLAST*genome. *OMICS* 9, 160–172 (2005).
37. Ayres P.G. (2004). Alexis Millardet: France's forgotten mycologist. Mycologist 18: 23–26 [Google Scholar]
38. Barnhill J.C., Stokes A.J., Koblan-Huberson M., Shimoda L.M.N., Muraguchi A., Adra C.N., Turner H. (2004). RGA protein associates with a TRPV ion channel during biosynthesis and trafficking. J. Cell. Biochem. 91: 808–820 [PubMed] [Google Scholar]
39. Beaudoin J., Laliberté J., Labbé S. (2006). Functional dissection of Ctr4 and Ctr5 amino-terminal regions reveals motifs with redundant roles in copper transport. Microbiology 152: 209–222 [PubMed] [Google Scholar]
40. Bender C.L., Malvick D.K., Conway K.E., George S., Pratt P. (1990). Characterization of pXV10A, a copper resistance plasmid in *BLASTcampestris* pv. *vesicatoria*. Appl. Environ. Microbiol. 56: 170175 [PMC free article] [PubMed] [Google Scholar]
41. Borkow G., Gabbay J. (2004). Putting copper into action: Copper-impregnated products with potent biocidal activities. FASEB J. 18: 1728–1730 [PubMed] [Google Scholar]
42. Bughio N., Yamaguchi H., Nishizawa N.K., Nakanishi H., Mori S. (2002). Cloning an iron-regulated metal transporter from rice. J. Exp. Bot. 53: 1677–1682 [PubMed] [Google Scholar]
43. Adachi, T. , Izumi, H. , Yamada, T. , Tanaka, K. , Takeuchi, S. , Nakamura, R. , & Matsuda, T. (1993). Gene structure and expression of rice seed allergenic proteins belonging to the α-amylase/trypsin inhibitor family. *Plant Molecular Biology*, 21, 239–248. [PubMed] [Google Scholar]
44. Adebiyi, A. P. , Adebiyi, A. O. , Jin, D.-H. , Ogawa, T. , & Muramoto, K. (2008). Rice bran protein-based edible films. *International Journal of Food Science & Technology*, 43, 476–483. 10.1111/j.1365-2621.2006.01475.x [CrossRef] [Google Scholar]
45. Agrawal, G. K. , & Rakwal, R. (2006). Rice proteomics: A cornerstone for cereal food crop proteomes. *Mass Spectrometry Reviews*, 25, 1–53. 10.1002/mas.20056 [PubMed] [CrossRef] [Google Scholar]
46. Agrawal, G. K. , & Rakwal, R. (2011). Rice proteomics: A move toward expanded proteome coverage to comparative and functional proteomics uncovers the mysteries of rice and plant biology. *Proteomics*, 11, 1630–1649. [PubMed] [Google Scholar]