STRUCTURAL DYNAMIC OF FLS2 LRR D122N BAK1 LRR CRYSTAL COMPLEX

By
Anika Tasfia Rodoshi
15136013

A thesis submitted to the Department of Mathematics and Natural Sciences in partial fulfillment of the requirements for the degree of
BSc

Mathematics and Natural Science
BRAC University
[December] [2020]

Declaration

It is hereby declared that

1. The thesis submitted is my own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I/We have acknowledged all main sources of help.

AnikaTasfiaRodoshi

15136013

Approval

The thesis/project titled “Structural dynamic of FLS2 LRR D122N BAK1 LRR CRYSTAL Complex” submitted by

Anika Tasfia Rodoshi (15136013)

Of Fall, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Science in Biotechnology.

Examining Committee:	
Supervisor: (Member)	_____ M H M Mubassir Lecturer, Biotechnology Program Department of Mathematics and Natural Sciences
Program Coordinator: (Member)	_____ Iftekhar Bin Naser, PhD Assistant Professor Department of Mathematics and Natural Sciences
Departmental Head: (Chair)	_____ Prof. Dr. A F M Yusuf Haider Professor and Chairperson Department of Mathematics and Natural Sciences

Ethics Statement

No animal or living things were used in this study.

Dedication

To the two who gave me birth.

Acknowledgement

I would like to express my sincere thanks to Almighty Allah for all He has done. Throughout my life, He has rewarded me with his wisdom and determination to make this possible.

I would like to thank Professor A F M Yusuf Haider, Ph.D., Professor and Chairperson of Department of Mathematics and Natural Sciences, BRAC University. I would like to thank Zubaida Marufee Islam, Department of Mathematics and Natural Sciences, BRAC University for the encouragement and direction that she offered me as an advisor. She is a role model of mine. Also, I would like to thank Professor Dr. M. Mahboob Hossain for his innate support.

I have a debt to my supervisor, Mr. M H M Mubassir, Lecturer, Biotechnology program, Department of Mathematics and Natural Sciences, BRAC University for keeping trust in me and helping me anytime I needed. Once again, I would like to thank him for his persistent guidance and support during this project.

In specific, I would like to thank my classmates and friends, Eera Ashrafy Anonna, Munira Tanzim, Touhidul Islam Jisan. I want to express my sincere gratitude to Imtiaz Rahman, who gave me mental support to make this stressful time easier, for giving me company and supporting me unconditionally during my undergraduate studies.

Abstract

This study includes a protein subunit extraction from *Arabidopsis Thaliana*. *Arabidopsis Thaliana*, an ideal plant for any type of research leading to a new direction of genetic experiments. Its importance shows by its minimum size that restrict the need of growth facilities, self-pollination that speeds the seed production. A mutation in the third leucine repeat of the BRI1 linked kinase 1(BAK1) protein of 4mn8 in *Arabidopsis Thaliana* changes the 122nd amino acid which is aspartate to asparagine along with deleted flg22.The mutation is specifically associated with photo morphogenesis. The BAK1 enzyme, however, also plays a key role in plant immunity by producing a heterodimer with a pattern-triggered flagellin sensing 2 (FLS2) receptor, a bacterial flagellin elicitor protein that leads to BAK1 and FLS2 trans-phosphorylation and subsequent initiation of the signal transduction pathway which is involve in activation of the immune response. The impact of the D122N mutation in BAK1 on the structural integrity of the FLS2-BAK1 complex was investigated here through molecular dynamics simulation. The D122N mutation in the native FLS2-BAK1 crystallographic structure of the BAK1 protein was induced using the *in silico* process. Simulation was done for both native and mutant complexes by subsequently utilizing molecular dynamics techniques for 5ns.The mutant and native complex is compared based on some parameters like RMSD (Root Mean Square Deviation), RMSF (Root Mean Square Fluctuation), Rg (Radius of Gyration), Hydrogen bond etc. The mutation is found to result in lower deviation of atoms from reference structure for the whole complex, FLS2-BAK1, compared to counterpart from the non-mutated complex. RMSF analysis revealed that mutation caused lower fluctuation of amino acid in both the N terminal and C terminal of FLS2-BAK1 complex that hinder stable interactions.FLS2-BAK1 had a lower radius of gyration at the complex containing mutation compared to the non-mutated complex. The mutation hindered the

formation of hydrogen bonds among all the peptides, often terminating all the hydrogen bonds between peptide.

Table of content

List of Tables and Figures

List of Figures

List of Tables

List of Acronymes

PTI = Pattern Triggered Immunity

ETI =Effector Triggered Immunity

PRR = Pattern Recognition Receptor

RLK =Receptor Like Kinase

PAMP =Pathogen Associated Molecular Pattern

RMSD =Root Mean Square Deviation

RMSF =Root Mean Square Fluctuation

Rg=Radius of Gyration

SASA =Solvent Accessible Surface area

Dd-a = Distance Between Donor and Acceptor

Dh-a = Distance Between Hydrogen and Acceptor

A(d-H-N) = Angle Between Donor-H-N

A(a-O=C) = Angle Between Acceptor-O=C

MO = Multiple Occupancy

Chapter 1: Introduction

1.1 PTI and ETI

The growth and reproduction of plants are hampered by a range of plant-associated microbes which acts as pathogens for the plant. Plants respond to these infections using a two-branched innate immune system. The first branch of this innate immune system is activated when interaction occurs with different molecules from different microbes, which can be non-pathogens as well. This system, since it is activated by different molecular patterns, is referred to as Pattern-Triggered Immunity (PTI). On the other hand, the second branch responds to factors associated with pathogen virulence. This activity is either direct or done through its effect on host goals. This system is known as Effector-Triggered Immunity (ETI)[1]. Wide-spectrum immunity to subsequent distal tissue pathogen attacks, a mechanism called Systemic Acquired Resistance(SAR), is often triggered by local PTI and ETI activation [2].

Incredible insights into molecular characterization, cell biology and evolution through biological kingdoms are presented by these plant immune systems, and the pathogen molecules to which they react. A thorough knowledge of plant immune function would promote increase in crop production for fruit, fiber and biofuels [1].

1.2 PAMPs

The first line defense mechanism in plants is Pattern-triggered immunity. For any kind of non-self microbial signature, pathogen-associated molecular patterns (PAMPs) is canonically triggered through the detection. Pattern recognition receptor (PRRs) can easily recognize PAMP as they are tightly stored molecules with reflective properties of a whole microbe community. Some PRRs, such as FLAGELLIN SENSING 2 (FLS2) and EF-Tu RECEPTOR, are trans-membrane receptor-like kinases (RLKs) or receptor-like proteins (RLPs) (EFR).For providing a downstream immune signaling, PRR's interference or noninterference with its partner protein depends on PAMP perception. For example, it may include the fast explosion of Ca2+ and accelerated release of Reactive Oxygen species (ROS), Ca2+-dependent protein kinase (CDPK) induction along with miltogen-activated protein kinase (MAPK) activation, phytohormone synthesis and comprehensive transcriptional and metabolic reprogramming [3].

In order to achieve disease tolerance after identification of PAMPs during the early phases of pathogen penetration, immune signal amplification is very essential to plants.

1.3 PRR

The first and most critical step in removing the bacteria from the host's body is the recognition of a bacterial attack. To do this, pattern recognition receptors (PRRs) for instance, Toll-like receptors (TLRs), nucleotide-binding leucine-rich repeat-containing receptors (NLRs) and alsoscavenger receptors (SRs) are mostly present in innate immune cells in the host innate

immune system. The presence of bacteria is detected by these receptors and helps to disperse the signal to the host, resulting in the recruitment of other immune cells that expel the bacteria from the system[10]. Plants have a large number of PRRs with remarkable structural and functional similarities to drosophila TOLL and mammalian TLRs. The first PRR detected in plants or animals was the protein Xa21, which confers resistance to *Xanthomonasoryzae*, a gram-negative bacterial pathogen. Since then, two other PRR from plants have been isolated: Arabidopsis FLS2 (flagellin) and EFR (Tu receptor elongation factor). It has defined the corresponding PAMPs for FLS2 and EFR. The plant PRRs transduce PAMP-triggered immunity(PTI) upon the identification of ligands. Resistance proteins that mimic NOD-like receptors are also encoded by plant immune systems which feature NBS and LRR domains and can also bear other retained interaction domains such as the cytoplasmic TIR domain found in Toll and Interleukin receptors[11]. For effector-triggered immunity (ETI), the NBS-LRR proteins are required.

1.4 *Arabidopsis thaliana*: An ideal plant for studying plant models

In this research, the protein subunit used here is derived from *A. thaliana*. Arabidopsis thaliana has strengthened its position as an optimal plant for scientific studies over the past 8 to 10 years. It grows, reproduces and responds in much the same way to stress and disease, as do many crop plants. In contrast to other plants, *Arabidopsis* has a small genome that simplifies and encourages genetic research. Unlike other plants, repetitive DNA sequences that hinder the study of genomes are missing [13].

A. thaliana, similar to animal model species, is easy to look after. It grows rapidly, produces many very tiny seeds, has a tiny genome of ~114.5 Mb and is genetically well defined because of the amount of work concentrated on this plant [13]. According to some other research, the genome size was found to be about ~135 Mb [14].

By studying their *Arabidopsis* homologues, the role of several genes isolated from crop plants can be understood better. For example, information gained from *Arabidopsis* about defense mechanisms against pathogens can be used directly in the production of disease-resistant plants in other species.

As seen by the vast number of Arabidopsis publications cited in 1993, which also included studies of crop plants such as soybeans, corn, maize, wheat, barley, rye, pepper, tomatoes, potatoes, cotton or sorghum, genetic similarities between Arabidopsis and crop species are increasing [16].While of no agricultural importance on its own, A. Thaliana is part of the *Brassicaceae* family which includes species of crops such as cabbage, cauliflower, sprouts of brussels, bokchoy, radish, and Chinese cabbage [15]. There is sufficient reason to believe that Arabidopsis will increasingly serve as a resource base for crop plant breeders in the coming years and as a model plant species that promotes the awareness of plant scientists worldwide.

1.5 4mn8: Crystal structure of flg22 in complex with FLS2 and BAK1

1.5.1: FLS2

FLS2 (Flagellin sensing-2) is associated with flagellin receptor and basically leucine rich kinase protein (LRR-RLK) repeat. This receptor has a specific focus in plant defense mechanism. It is activated by bacterial elicitor.Mutation in FLS2 gene can cause lack of response to flg22.

FLS2 is made of three major parts: an extracellular, a trans-membrane, and an intracellular region. The extracellular domain is known as the Leucine-rich repeat (LRR) domain. It is in this region, the amino-terminus, where the response of FLS2 to flg22 is said to have direct interaction with flagelin. Trans-membrane domain is where proteins migrate from extracellular to intracellular and is very much thermodynamically stable.This occurs only in phospholipid membranes of cells.

The other one is intracellular domain which consists of serine/threonine kinase. A protein kinase cascade that leads to a response is catalyzed by phosphorylation in this domain. This response expresses the changes of plant growth and defense in FLS2 [21].

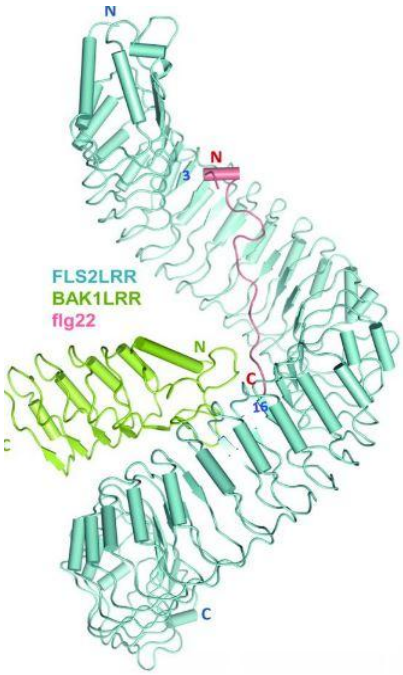

Figure 1 : flg22 in complex with FLS2LRR D122N BAK1 LRR crystal structure [19]

1.5.2 BAK1 co-receptor

BRI1 (Brassinosteroid Insensitive 1)-associated receptor kinase 1 is known as BAK1. This is also known in another name SERK 3 (Somatic Embryogenesis Receptor Kinase 3).Since researchers discovered in 2002 that SERK3 serves as a signaling partner of another LRR receptor kinase, BRI11, SERK3 was renamed BAK1.

It is a leucine rich repeat (LRR)receptor-like kinase(LRK). This LRR domain is followed by serine and leucine rich domain.It also includes of a short C domain and cytoplasmic kinase domain.

BAK1 was discovered to interact with receptor FLS2 which has a function in plant innate immunity. These researches resulted in the formulation of the idea that BAK1 has a crucial role in the regulation of several LRR-RLK by interacting with them in a stimulus dependent manner.

1.6 Importance of studying this protein

Regulatory leucine-rich repeat (LRR) RLK proteins such as SERK or BIR proteins are required by plant cells in terms of plant defense mechanism. Molecular Dynamics of this protein shows how FLS2 connects signaling complexes with BAK1 (SERK3). Moreover, by this study, FLS2-BAK1 Interactions with the domain may take place independently of intracellular domain interactions. As this is computational scenario, depending on the study many in-vivo experiments may take place with a proper acknowledgement of the resources.

1.6.1 SASA

SASA(Solvent accessible surface area) calculates a solvent molecules role over the surface of the molecule of interest. By this calculation protein folding problem can be solved. But it has some limitations in terms of measuring quantitative area.

1.6.2 RMSD

RMSD (Root Mean Square Deviation) is used to measure the average distance between the atoms of superimposed (between the 3D folds of two structure) proteins.

An RMSD of 1Å = 0.1nm means identical structures. Higher RMSD value means loss of interactions with other atoms.

1.6.3 RMSF

RMSF (Root Mean Square Fluctuation) of each residue in simulation period can be obtained only after Protein-ligand simulation studies. RMSF is in general less evolutionary then net structure. The function of a protein depends not only on its structure but also on its dynamics. Because dynamics is less conserved, RMSF is not as conserved as structure.

1.6.4 Rg

Radius of Gyration (Rg) is used to determine the compactness of protein. Protein recoiling, uncoiling in the mean of time is detected by Rg value.

1.6.5 H-bond

H-bond is very low than covalesive bond but massive H-bond gives rise to protein folding hydrophobic interaction .Maximum H-bond is favorable and it determines side chain-side chain interaction and side chain-main chain interaction.

In any complex if RMSD, RMSF values are comparatively low, it means it is favorable for the complex. If not then we remove it by mutating and again calculate RMSD,RMSF.

1.7 PIC (Protein Interaction Calculator)

Protein structure interactions and protein interactions in an assembly are important factors in understanding the molecular basis of protein stability and functions and their complexes. Many strong and weak interactions are present in any structure to represent the stability of a protein structure. The Protein Interactions Calculator (PIC) is a server which shows a view of the coordinate set of a protein or illustrates the 3D structure of the assembly and measures various 3D structures such as disulphide bonds, hydrogen bonds, hydrophobic residue interactions, ionic interactions, aromatic-aromatic interactions, aromatic-sulphur interactions and cation-π interactions within a protein or in a complex between proteins.

1.8 Purpose of the research

The purpose of the research is to find out the change in crystal interaction in between FLS2 LRR D122N BAK1 LRR complex and to find out its stability.

Chapter 2: Materials and methods

The data calculation was done on open source programs like GROMACS, Xmgrace, Chimera. These softwares were used to analyze the interaction of 4mn8 protein in complex with FLS2 and BAK1 ectodomain.

Table 1:Tools used in this study

Tool	Function
RCSB protein data bank	Retrieving data from database for the three-dimensional structural data of large biological molecules
GROMACS 5.0	Dynamic simulations of biomolecules
Xmgrance	Interactively modify plots for biomolecules
Chimera	Interactive visualization of molecular structures
Pymol	Animating 3D structures of biomolecules
PIC (Protein interaction calculator)	Recognizing bonds and interactions of protein molecules; Determining surface area and distance of a residue

2.1 Retrieving protein sequence

At first, PDB file of native 4mn8 was collected from Protein data bank[19]. PDB file is required for running files in GROMACS as it includes data obtained by X-ray crystallography and nuclear magnetic resonance (NMR) spectrometry by biologists and biochemists from all over the world.

2.2 GROMACS

It was run on GROMACS to find out its molecular interaction and stability. The three dimensional visualization was analyzed in PyMOLl. Using the Linux operating systemwhich is open for numerous softwares, this study has been a lot easier. The native version was run in GROMACs in 5ns to check its stability in terms of RMSD, RMSF,Rg,H-bond etc.

2.3 Modification of the protein

To make it more stable protein and to compare with the native version, 4mn8 protein was modified a little. Using GROMACS the PDB file was read. When BAK1, FLS2 and flg22 were only remaining, the flg22 portion was deleted from the whole sub-unit as flg22-FLS2-BAK1 mutated complex. In BAK1 (which is the longest) 122 number residue was mutated. It was changed with asparagine from aspartate. Then the remaining portion of mutated BAK1-FLS2 complex was run at 5 ns which would ultimately help to analyze graphs of RMSD, RMSF and Rg. The model was run at 5 ns.

2.4 Computation approach for Molecular Dynamic (MD) simulation

It is technically known as Molecular dynamics to model the movements of a system of particular particles. The mechanism is normally small like an atom and a diatomic molecule undergoes a greater chemical reaction like a galaxy [23]. Understanding of the interaction potential for the particles is important for performing a molecular dynamics simulation [24].

Model simulations are used to obtain conclusive descriptions of the passage of a single particle as an act of time and to comprehend the properties of a model. In the field of macromolecular

science, three distinct types of simulation methods exist. The first works by means of sampling configuration space, which helps to identify configurations with knowledge obtained from real-life experiments. The second approach is used to explain the equilibrium mechanism with the assistance of structural and motional characteristics and thermodynamic parameter values. The real dynamics was studied by the third methodology in which the motion and growth of over time was calculated. [25]

Program availability and computing power must be needed to perform practical studies on biological macromolecules using molecular dynamic (MD) simulation. In older days, conducted in less than 10 ps in duration, but in the current scenario, 1000 times longer simulations of the same sizes are mostly performed in half the time needed than before. Apart from time saving computing capacity from which several numbers of simulations can be done is another significant fact of simulation. The most frequently used applications are CHARMM20, GROMOS22 along with AMBER21. In order not to have intense and regulated temperature and pressure where numerical simulations solve this problem, it is often difficult to get the desired outcome in the laboratory [25].

Chapter 3: RESULT

In this study, interactions between mutated BAK1 and FLS2 complex were analyzed and also illustrated with graphs and figures. Molecular differences of the wild 4mn8 were compared with mutated complex before and after MD simulations and molecular interactions of BAK1-FLS2 complex were calculated at the end of the chapter.

3.1 Molecular visualization of mutated FLS2 LRR D122N BAK1 LRR

FLS2 has total 784 residues and BAK1 has 200 residues in total after deleting everything in the protein structure. Our targeted mutation was done on the 122thresidue of BAK1. It was changed to asparagine from aspartate. The colored portion is the changed amino acid of BAK1 at residue number 122.

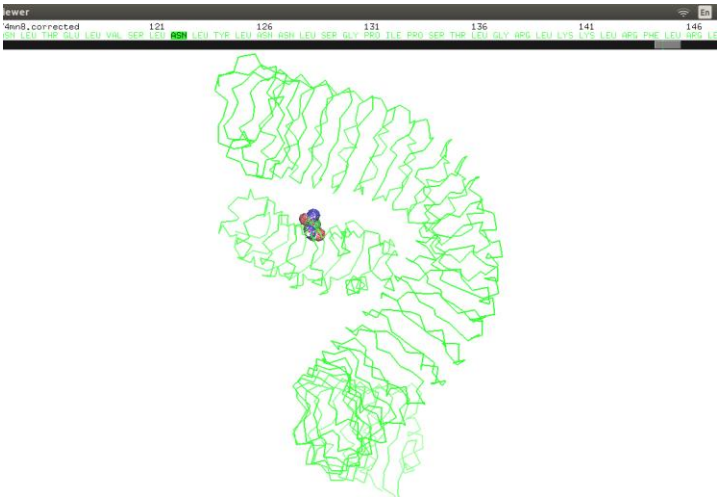

Figure 2: Molecular visualization of FLS2 LRR D122N BAK1 LRR CRYSTAL Complex after mutation.

3.2 Pressure, Temperature and Potential

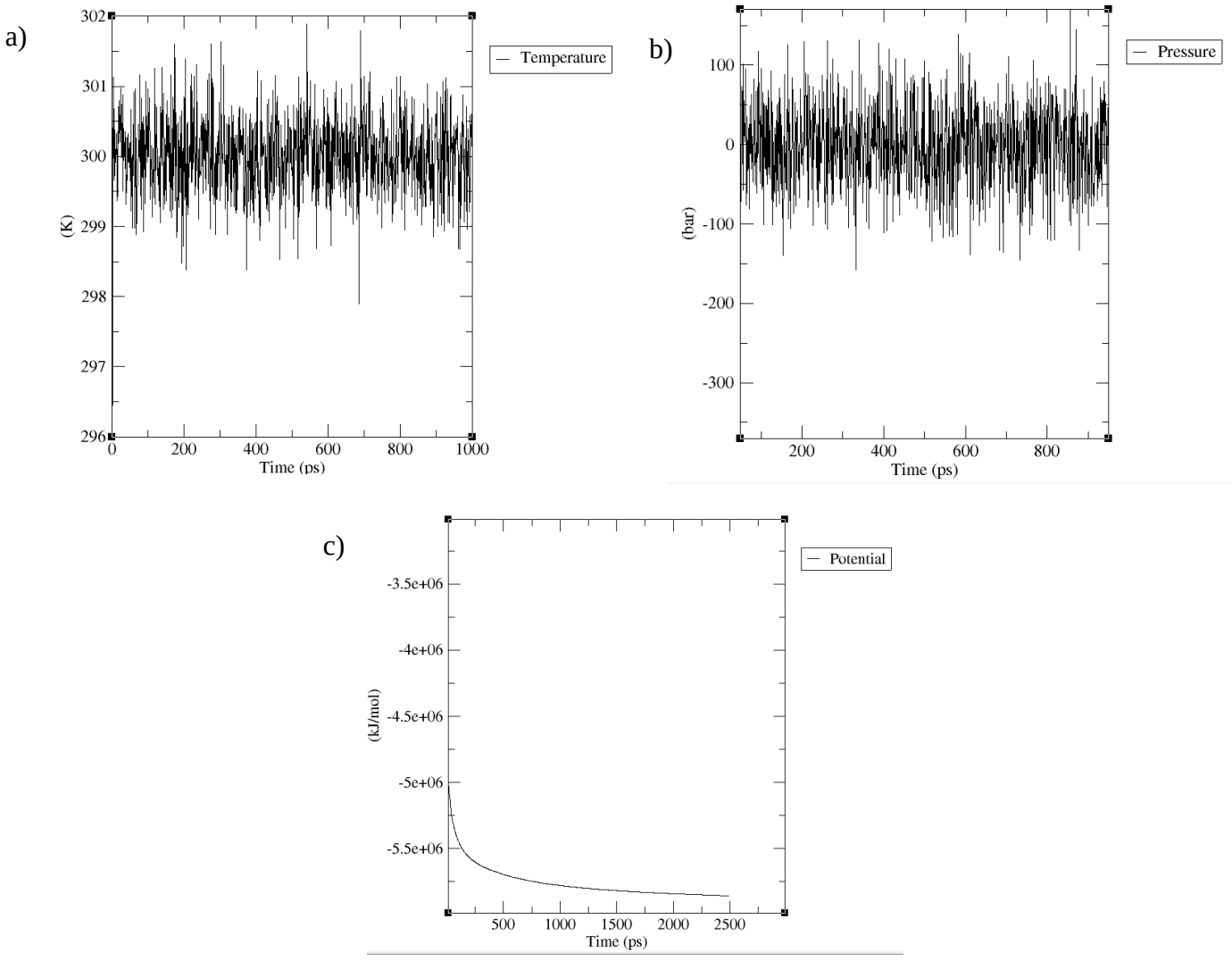

Figure 3 : Different parameter of pre-molecular dynamics processing of LRR D122N BAK1 LRR CRYSTAL protein complex. (a) Temperature equilibration of mutated complex, (b) Pressure equilibration of mutated complex, (c) Energy minimization steps of mutated complex.

After incorporating water solvent, the protein complex was first energy-minimized and the system's charge was neutralized by adding ion. The system was stabilized first under the NVT ensemble for 1000 ps after energy minimization, where the whole system stabilizes at an average temperature of 300 K. Under the NPT ensemble, the second step of equilibration was carried out where the system stabilizes the average pressure. The system's density stabilizes at an average density. In a pair of desktop computers consisting of an intel i5 7500 processor and 16 gigabytes of RAM, the molecular dynamic simulations were then carried out for 5 ns or 5000 ps. Using all 4 threads with an average of 2.841 ns per day, the simulations took about 12 days to complete.

3.3 RMSD, RMFS, Rg analysis and H-bond

Several key parameters were analyzed to discern the difference between native and mutated protein complexes as molecular dynamic simulations were completed. Compared with the native form, RMSF analysis, Rg analysis and amount of H-bond analysis over time between the mutated FLS2-BAK1 complex and the native one, these analyses involve RMSD over time of the mutated complex.

3.3.1 RMSD of the complex

RMSD (Root Mean Square Deviation) is used to measure the average distance between the atoms of superimposed (between the 3D folds of two structure) proteins. The stability of the MD simulation was measured in terms of deviations by analyzing Root meansquare deviation (RMSD). The time evolution of the BAK1 and FLS2 (residues only) are monitored as a function of time. The RMSDs ofBAK1 and FLS2 in the four simulated systems are shown in Figure 4.

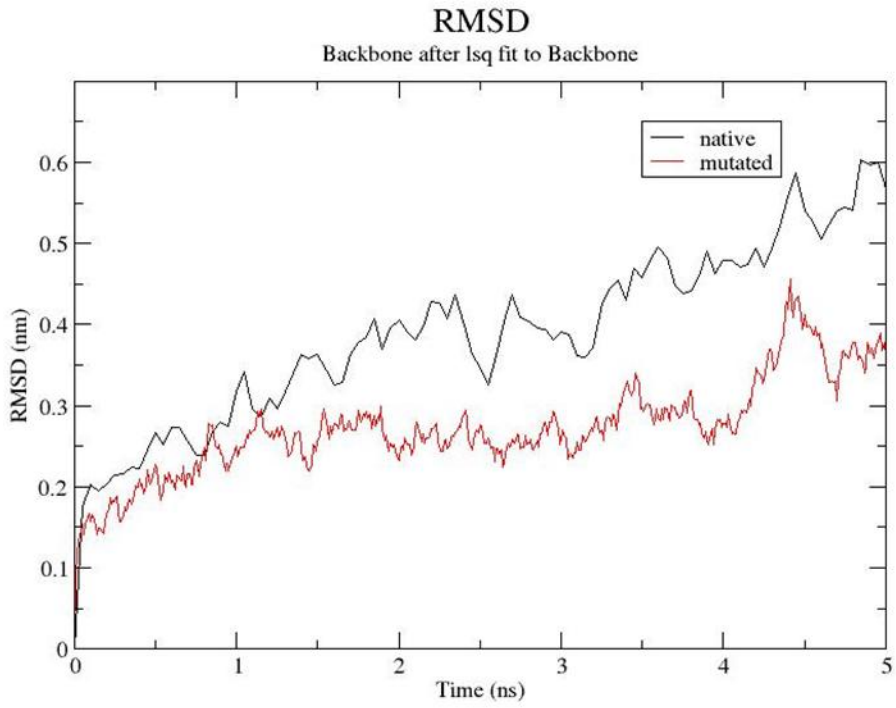

Figure 4: RMSD graph of mutated and native complex at 5ns or 5000 ps

3.3.2 RMSF

The RMSFs of the residues of BAK1 and FLS2 in all the simulated systems were calculated from the MD trajectories. 5ns were used to calculate the RMSF values.

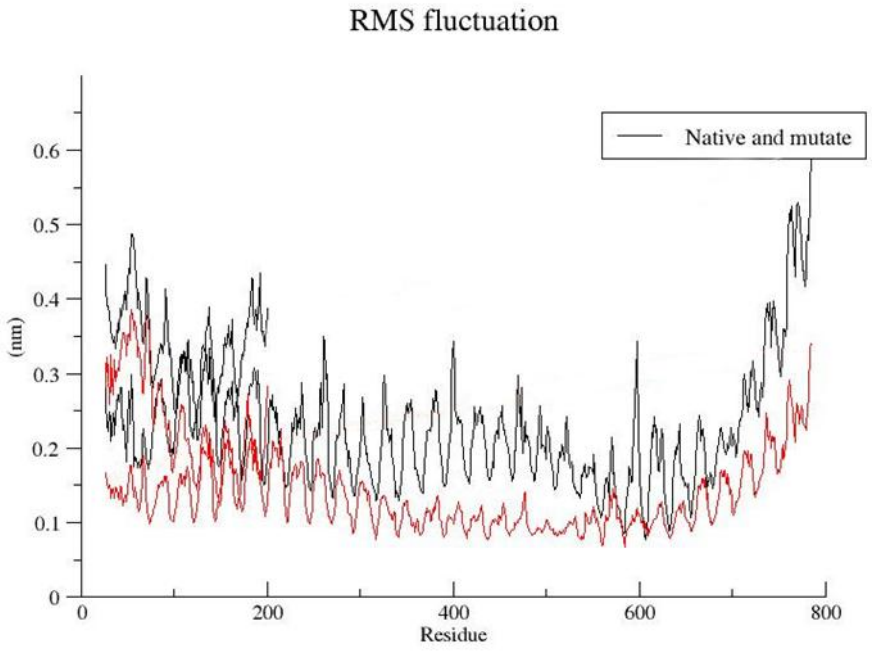

Figure 5: RMSF calculation of residues of native and mutated complex

3.3.3 Radius of Gyration

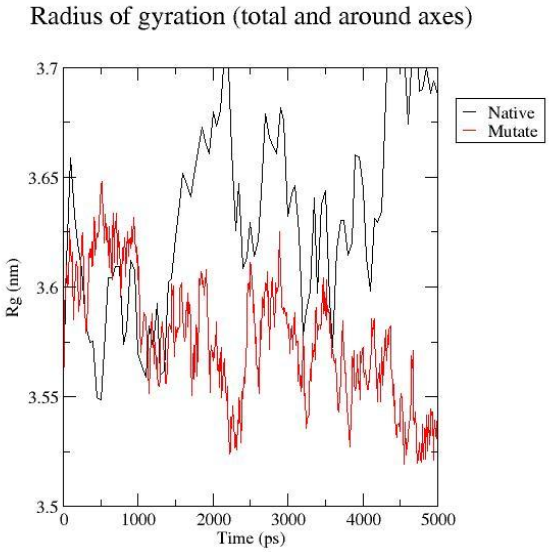

Figure 6 : Radius of gyration of both Native and Mutated complex at 5ns or 5000ps

The radius of gyration or Rg for mutated complex stayed less than the native complexuntil 0.3ns. After that, the Rg for mutated complex maintained a higher value than the native complex until 1ns. Then it stayed lower than the native complex most of the time till 5ns.

3.3.4 H-bond

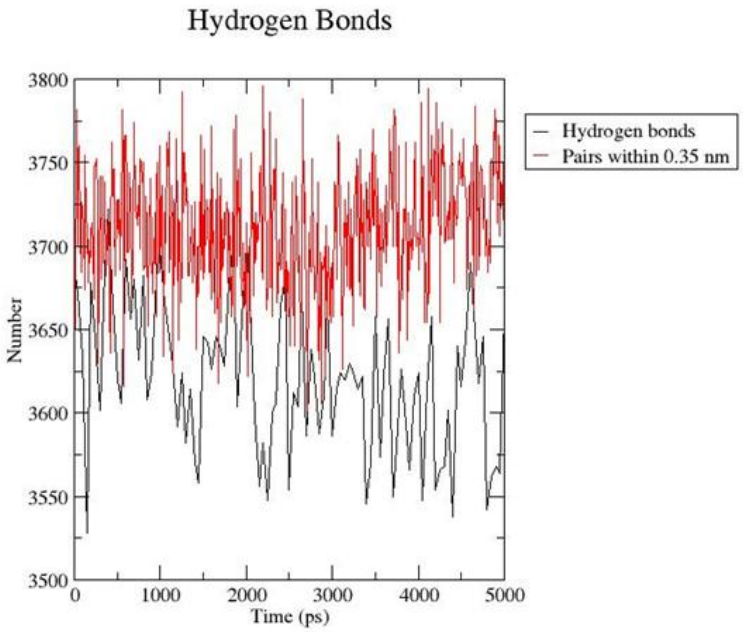

Figure 7: H-bond graph of both native and mutated complex

The cumulative number of hydrogen bonds between the FLS2 and BAK1 peptides suggests that, compared to the non-mutated BAK1-FLS2 relationship, the mutated BAK1 forms a smaller number of hydrogen bond.

Table 2: Hydrophobic Interactions within 5 angstroms

POS	**Residue**	**Chain**	**Position**	**Residue**	**Chain**
369	PHE	A	54	VAL	B
507	ILE	A	60	PHE	B
509	TYR	A	60	PHE	B

555	VAL	A	60	PHE	B
650	MET	A	144	PHE	B

Table 3: Protein-protein main chain-side chain interaction

POS	CHAIN	RES	ATOM	POS	CHAIN	RES	ATOM	MO	Dd-a	Dh-a	A(d-H-N)	A(a-O=C)
530	A	GLN	NE2	44	B	LYS	O	1	2.74	1.98	126.43	131.29
530	A	GLN	NE2	44	B	LYS	O	2	2.74	3.12	59.40	131.29

Dd-a = Distance Between Donor and Acceptor

Dh-a = Distance Between Hydrogen and Acceptor

A(d-H-N) = Angle Between Donor-H-N

A(a-O=C) = Angle Between Acceptor-O=C

MO = Multiple Occupancy

Table 4: **Protein peotein side chain-side chain Hydrozen bond**

POS	CHAIN	RES	ATOM	POS	CHAIN	RES	ATOM	MO	Dd-a	Dh-a	A9d-H-N)	A9a-O=C)
625	A	ASN	ND2	100	B	TYR	OH	1	3.01	3.35	62.26	999.99

625	A	ANS	ND2	100	B	TYR	OH	2	3.01	2.99	81.51	999.99
625	A	ANS	ND2	124	B	TYR	OH	1	3.34	3.38	78.58	999.99
625	A	ASN	ND2	124	B	TYR	OH	2	3.34	2.58	130.44	999.99
674	A	ASN	OD1	96	B	TYR	OH	1	3.26	2.22	162.88	999.99
674	A	ASN	OD1	96	B	TYR	OH	2	3.26	3.96	42.29	999.99
698	A	ASP	OD2	167	B	GLN	OE1	1	2.42	1.90	105.48	999.99
698	A	ASP	OD2	167	B	GLN	OE1	2	2.42	2.26	85.62	999.99
698	A	ASP	OD2	167	B	GLN	NE2	1	2.43	2.76	60.85	999.99
698	A	ASP	OD2	167	B	GLN	NE2	2	2.43	1.46	149.75	999.99
96	B	TYR	OH	674	A	ASN	OD1	-	3.26	9.99	999.99	999.99
100	B	TYR	OH	625	A	ASN	ND2	-	3.01	9.99	999.99	999.99
124	B	TYR	OH	625	A	ASN	ND2	-	3.34	9.99	999.99	999.99
143	B	ARG	NH2	674	A	ASN	OD1	1	2.97	3.57	47.89	999.99
143	B	ARG	NH2	674	A	MET	OD1	2	2.97	3.20	130.59	999.99
143	B	ARG	NH1	699	A	MET	SD	1	3.84	3.45	104.30	999.99
143	B	ARG	NH1	699	A	MET	SD	2	3.84	4.00	73.64	999.99
146	B	ARG	NH1	650	A	MET	SD	1	4.00	4.95	22.02	999.99
146	B	AR	NH1	650	A	ME	SD	2	4.00	3.28	128.80	999.99

		G				T						
146	B	ARG	NH2	650	A	MET	SD	1	3.30	3.92	47.29	999.99
146	B	ARG	NH2	650	A	MET	SD	2	3.30	2.35	153.54	999.99
167	B	GLN	OE1	698	A	ASP	OD2	1	2.42	1.59	128.93	999.99
167	B	GLN	OE1	698	A	ASP	OD2	2	2.42	3.31	26.63	999.99
167	B	GLN	NE2	698	A	ASP	OD2	1	2.43	1.62	129.93	999.99
167	B	GLN	NE2	698	A	ASP	OD2	2	2.43	3.32	24.95	999.99

Dd-a = Distance Between Donor and Acceptor

Dh-a = Distance Between Hydrogen and Acceptor

A(d-H-N) = Angle Between Donor-H-N

A(a-O=C) = Angle Between Acceptor-O=C

MO = Multiple Occupancy

Table 5 **: Protein-protein Aromatic-aromatic Interactions**

Residue	**Position**	**Chain**	**Residue**	**Position**	**Chain**	**D (centroid-centroid)**	**Dihedral Angle**
509	**TYR**	**A**	**60**	**PHE**	**B**	**6.91**	**63.82**

Table 6: Protein – protein cation-Pi Interaction

Position	Residue	Chain	Position	Residue	Chain	D (cation-Pi)	Angle
676	PHE	A	143	ARG	B	5.71	124.48

3.3.5 SASA

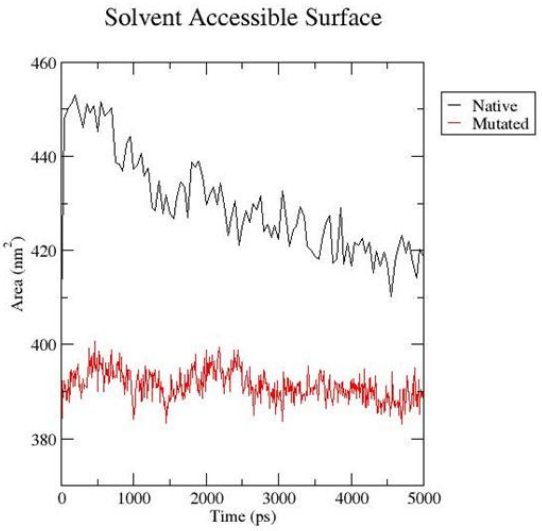

Figure 8: SASA calculation of both Native and mutated complex at 5ns or 5000 ps

Chapter 4: Discussion

Why use computational approach?

In recent years, with the progress of an expanding scope of genome sequencing programs, the number of identified protein sequences has risen exponentially.However, a 3D structure is required.As of 2019, UniProtKB/TrEMBL has 147 million protein sequences and just 140,393 RSCB PDB structures.This large distance is due to the inherently time-consuming and dynamic nature of conventional X-ray crystallography structure determination techniques, NMR spectroscopy, which restricts their usefulness.

Significance and findings of the study:

The different analytical techniques were used in this research to understand Arabidopsis thaliana Pattern Triggered Immunity (PTI) against FLS2 by using the BAK1 domain of 4mn8. Besides, the in-depth analysis to the relationships between FLS2 and BAK1 reveals the 122ndresidue (aspartate changes to asparagine) in BAK1 plays a significant role in the function of the FLS2-BAK1 interaction. This residue gives higher binding energy in the interaction between FLS2 and BAK1.

The special dynamic characteristics of proteins are derived from their fluid existence, which can be well characterized in terms of a multi-minimum energy landscape.In fact, proteins occur in a very large number of slightly different, hierarchically ordered, conformative sub-states.Each protein molecule is limited to one local minimum in the energy landscape at low temperatures, and the system behaves as a harmonic solid.Proteins undergo reversible folding/unfolding transitions when subjected to hydrostatic pressures of 2–10 kilo bars (kbar).

In the crystal structure, the mutation of aspartate to asparagine at the 122nd residue in the BAK was induced to preserve the conformation of other amino acids to the indigenous form. The proteins can cause unnecessary changes to the conformation of the complex as they mutate in the raw amino acid sequence and then modeling and docking is done.

The mutation was done in a way that guaranteed the less hysterical collision of new amino acid asparagines with neighboring molecules. Both the native and mutated protein complexes were positioned at the middle of a three-dimensional cubic box at least 1.0 nm away from the edge of the protein in order to conduct the molecular dynamics simulation.Since the distance was set to follow the minimal image convention, we used periodic boundary status in the simulation (PBC). The procedure subsequently reduced energy to relax the protein complex, since the crystallized protein form does not entirely reflect the existing state of nature due to the complex handled with the crystallization process.The machine was then adjusted in two phases, adding about 300 Kelvin to the system temperature or room temperature in the first process. The machine pressure and density were modified in the second phase to get them to a running average pressure,

RMSD is useful in evaluating the motions of the structure in the trajectory of simulation in a time-dependent way [20]. Here, we used the first simulation frame as the reference structure, as what we used to start the simulation was the balanced structure rather than the unchanged crystallographic structure.The Value of RMSD, comparison of native and mutated forms revealed that atoms of the mutated complex deviated less than the native complex initially.

The RMSD of the mutated complex stayed lower than the native complex before 0.9 ns. After 0.9 nsthe RMSD of mutated complex were same as the RMSD of native one in two peaks.From the graph it can be seen that the mutated one is more stable than the native one.Specifically,

between1.5ns to 2 ns and some region after 2.5 ns and 3.5ns mutated one was more stable whereas the native one in this 5ns did not show any stable graph.

Another parameter for calculating the variance of individual atoms or groups of atoms relative to the reference structure average over atoms is the Root Mean Square Fluctuation, or RMSF. An increase in the value of RMSF means protein is unstable at that specific point and RMSF decreasing means protein is becoming morestable.

After calculating the average for each residue, RMSF data revealed that almost all FLS2 residues and most BAK1 residues form a mutated complex fluctuate less than the native counterpart.

The radius of gyration or Rg is a protein compactness measurement that is measured by calculating the distance between the protein's center of mass and each of its terminal ends.It checks how often the structure of a protein folds or unfolds. Rg values of the whole complex show that the mutated complex evolves more than the native complex, suggesting that the loss of rigid interactions over time and mutated protein extends more than the native complex.

The radius of the mutated one showed higherand lowervalues than the native oneup to 1.5ns. After 1.5 ns it decreasedwithtime and stayed lower than the native values.After 1.5 ns till 2 ns it showed very much compactness than the native one and again after 2.5ns till end it showed morecompactness than the native one.

If we take a close look at the number of hydrogen bonds formed over the simulation period between two different complexes, we can get an idea of the pattern of intra-protein interaction taking place. The number of hydrogen bonds formed between BAK1 and FLS2 stayed higher in the mutated graph and lower in native one.

It shows that the BAK1-induced mutation had a strong effect on the formation of hydrogen bonds between these two peptides. Compared with the native complex, the mutated complex permitted less hydrogen bonds between FLS2 and BAK1.

By analyzing all the data produced through the simulation of molecular dynamics, RMSD, RMSF, Rg, hydrogen bond interaction, we can see that the mutation had more stable effect on the FLS2-BAK1 complex. The Arabidopsis thaliana elongated (elg) phenotype is caused by the D122N mutation in the BAK1 coding protein, which, among other phenotypes, causes the plant to have elongated stem and early flowering. Not only does BAK1 protein play a role in cell growth, but it also participates in plant growthand immunity by forming a heterodimer caused by flg22 with FLS2 plant PRR. But in this study flg22 was deleted at the very first step in the mutation which dramatically changes the result of mutated version compared to the native one.

Chapter 5: References

1. Jones, J. D., &Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(7117), 323–329. https://doi.org/10.1038/nature05286

2. Fu, Z. Q., and Dong, X. (2013). Systemic acquired resistance: turning local infection into global defense. *Annu. Rev. Plant Biol.* 64, 839–863. doi: 10.1146/annurev-arplant-042811-105606

3. Hou, S., Z. Liu, et al. (2019). "Damage-Associated Molecular Pattern-Triggered Immunity in Plants." Frontiers in Plant Science10(646).

4. Vega, Karina &Kalkum, Markus. (2012). Chitin, Chitinase Responses, and Invasive Fungal Infections. International journal of microbiology. 2012. 920459. 10.1155/2012/920459.

5. Lotze, M. T., Zeh, H. J., Rubartelli, A., Sparvero, L. J., Amoscato, A. A., Washburn, N. R., et al. (2007). The grateful dead: damage-associated molecular pattern molecules and reduction/oxidation regulate immunity. *Immunol. Rev.* 220, 60–81. doi: 10.1111/j.1600-065X.2007.00579.x

6. Boller, T., and Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant. Biol.* 60, 379–406. doi: 10.1146/annurev.arplant.57.032905.105346

7. Seong, S. Y., and Matzinger, P. (2004). Hydrophobicity: an ancient damage-associated molecular pattern that initiates innate immune responses. *Nat. Rev. Immunol.* 4, 469–478. doi: 10.1038/nri1372

8. Tanaka, K., Choi, J., Cao, Y., and Stacey, G. (2014). Extracellular ATP acts as a damage-associated molecular pattern (DAMP) signal in plants. *Front. Plant Sci.* 5:446. doi: 10.3389/fpls.2014.00446

9. Jones, J., Dangl, J. The plant immune system. *Nature* 444, 323–329 (2006). https://doi.org/10.1038/nature05286

10. Gulati, A., Kaur, D., Krishna Prasad, G., &Mukhopadhaya, A. (2018). PRR Function of Innate Immune Receptors in Recognition of Bacteria or Bacterial Ligands. Advances in experimental medicine and biology, 1112, 255–280. https://doi.org/10.1007/978-981-13-3065-0_18

11. Franchi, L., Warner, N., Viani, K., & Nuñez, G. (2009). *Function of Nod-like receptors in microbial recognition and host defense. Immunological Reviews, 227(1), 106–128.* doi:10.1111/j.1600-065x.2008.00734.x

12. Hou, S., Liu, Z., Shen, H., & Wu, D. (2019). *Damage-Associated Molecular Pattern-Triggered Immunity in Plants. Frontiers in Plant Science, 10.* doi:10.3389/fpls.2019.00646

13. Endogenous Chemiluminescence from Germinating Arabidopsis Thaliana Seeds, HomaSaeidfirozeh, AzizollahShafiekhani, Michal Cifra and Amir Ali Masoudi, Scientific Reports, (2018) 8:16231, DOI:10.1038/s41598-018-34485-6 (iXon)

14. Cheng, C.-Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: a complete reannotation of theArabidopsisthalianareference genome. The Plant Journal, 89(4), 789–804. doi:10.1111/tpj.13415

15. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature, 408(6814), 796–815. doi:10.1038/35048692

16. José M. Alonso, Anna N. Stepanova, Thomas J. Leisse, Christopher J. Kim, Huaming Chen, Paul Shinn, Denise K. Stevenson, Justin Zimmerman, Pascual Barajas, Rosa Cheuk, Carmelita Gadrinab, Collen Heller, Albert Jeske, Eric Koesema, Cristina C. Meyers, Holly Parker, Lance Prednis, Yasser Ansari, Nathan Choy, HashimDeen, Michael Geralt, NishaHazari, Emily Hom, Meagan Karnes, Celene Mulholland, RalNdubaku, Ian Schmidt, Plinio Guzman, Laura Aguilar-Henonin, Markus Schmid, DetlefWeigel, David E. Carter, Trudy Marchand, Eddy Risseeuw, Debra Brogden, AlbanaZeko, William L. Crosby, Charles C. Berry, and Joseph R. Ecker (2003) Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana Science 301: 653-657.

17. Barbara Buchanan, Plant Genome Information Center, National Agricultural Library, U.S.D.A.

18. Rajamuthiah, R., &Mylonakis, E. (2014). *Effector triggered immunity. Virulence, 5(7), 697–702.* doi:10.4161/viru.29091

19. Chai, J., Sun, Y., Han, Z. (2013) Crystal structure of flg22 in complex with the FLS2 and BAK1 ectodomainsDOI: 10.2210/pdb4MN8/pdb

20. Amano, Y., Tsubouchi, H., Shinohara, H., Ogawa, M., and Matsubayashi, Y. (2007). Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18333–18338. doi: 10.1073/pnas.0706403104

21. Martínez, L., 2015. Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. PLoS One 10. https://doi.org/10.1371/journal.pone.0119264

22. Sun, Y., Li, L., Macho, A. P., Han, Z., Hu, Z., Zipfel, C., … Chai, J. (2013). *Structural Basis for flg22-Induced Activation of the Arabidopsis FLS2-BAK1 Immune Complex. Science, 342(6158), 624–628.* doi:10.1126/science.1243825

23. Barnes, J.E.J.N., Evolution of compact groups and the formation of elliptical galaxies.1989. 338(6211): p. 123.

24. Karplus, M. and G.A.J.N. Petsko, Molecular dynamics simulations in biology. 1990.347(6294): p. 631.

25. Karplus, M., J.A.J.N.S. McCammon, and M. Biology, Molecular dynamics simulations ofbiomolecules. 2002. 9(9): p. 646.