EFFECT OF D122N MUTATION IN BAK1 ON STRUCTURAL INTEGRITY OF PROTEIN COMPLEX CONSISTING HAESA, IDA AND BAK1 ECTODOMAIN— AN *IN-SILICO* STUDY

By

Kuhu Kinnori
ID: 17136012

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

Department of Mathematics and Natural Sciences
BRAC University
August 2021

Declaration

It is hereby declared that

1. The thesis submitted is my/our 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 have acknowledged all main sources of help.

Name of Student and Signature:

Kuhu Kinnori

ID: 17136012

Approval

The thesis titled "EFFECT OF D122N MUTATION IN BAK1 ON STRUCTURAL INTEGRITY OF PROTEIN COMPLEX CONSISTING HAESA, IDA AND BAK1 ECTODOMAIN—AN *IN-SILICO* STUDY" submitted by Kuhu Kinnori (ID: 17136012) of Spring semester, 2017 has been accepted as satisfactory in partial fulfillment of requirement for the degree of Bachelors 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
Biotechnology Program
Department of Mathematics and Natural Sciences

Departmental Head:
(Chair)

_____

A F M Yusuf Haidar
Professor and Chairperson
Biotechnology Program
Department of Mathematics and Natural Sciences

Ethics Statement:

No living organism was used or harmed in this study.

Abstract

Plants display immunity towards invading pathogens through layers of immunity. The pattern triggered immunity in *Arabidopsis thaliana* is the first layer of immunity against pathogens. Pattern Recognition Receptor Kinase, HAESA binds to Damage Associated Molecular Pattern, IDA, with the coreceptor SERK1. SERK1 is a member of the greater SERK family of receptor-like kinases. BAK1, alternatively known as SERK3, is also a representative of the SERK protein family. Therefore, it can be hypothesized that BAK1 could function alternatively to SERK1 protein in the HAESA-IDA complex. Besides, the elongated (elg) phenotype of BAK1 shows a mutation in the 122nd residue from Aspartate (Asp) to Asparagine (Asn). The viability of this mutated BAK1 as an alternative to SERK1 protein in the complex of HAESA LRR and IDA DAMP in terms of structural and residual alterations is checked in this research. The HAESA-IDA complex is docked and run through molecular dynamics simulation. The external molecular structure and internal residual arrangements are studied before and after the molecular dynamics simulation and checked for similarities and dissimilarities. Structural comparisons were made using HAESA-IDA-SERK1 (PDB ID: 5IYX) complex, and the FLS2-flg22-BAK1 complex (PDB ID: 4MN8) were used as references for the constructed experimental complex of HAESA-IDA-BAK1. Parameters like Root Mean Square Deviation (RMSD), Root Mean Square Fluctuations (RMSF), Hydrogen bonding, Solvent Accessible Surface Area (SASA), and Radius of Gyration (Rg) of the constructed tri-protein complex showed whether the complex is viable or not. High fluctuations in respective graphical data and incoherence of bond stability showed that mutated BAK1 in HAESA-IDA-BAK1 cannot form as a stable complex, hence the mutated BAK1 protein cannot be substituted for SERK1 in plant pattern triggered immunity.

Dedication

To the people who transcended beyond the biological definition of ‘family’.

Acknowledgement

Foremost, I would like to owe a debt of gratitude to my supervisor M H M Mubassir, Lecturer, Department of Mathematics and Natural Sciences, for his cordial support and expert guidance throughout the process of work. He has patiently taught me the necessary knowledge and techniques to complete my thesis work successfully.

I acknowledge my esteem to Professor A F M Yusuf Haider, the Department of Mathematics and Natural Sciences Chairperson, BRAC University, whose noble effort for the department has improved the department and inspired students like myself to work hard for the best results. I want to express my heartfelt gratitude for Professor Vincent Chang, the Vice-Chancellor, whose inspirational speeches inspired me to transform for the better and keep moving forward in the darkest of times.

My profound respect goes towards the course coordinator and lecturer, Iftekhar Bin Naser, Ph.D., a phenomenal teacher throughout my undergraduate period.

Furthermore, I want to thank respected seniors Raghib Ishraq Alvy, Araf Bin Satter and Rad Shahmat, who have guided me and helped troubleshoot through my academic queries.

I thank my cousin Atiab Jobayer for teaching me the ABCs of computers and my best friend Tanjeel Sultan Khan to provide me moral support through crises.

Lastly, I thank my two fathers, my mother, and my ‘Nanu’, who has raised me and tended to my mental, emotional and spiritual needs with their constant support, without whom I could not have come this far.

Contents

List of Figures

List of Tables

List of Acronyms

PTI- Pattern Triggered Immunity

ETI- Effector Triggered Immunity

PRR- Pathogen Recognition Receptor

LR- Leucine-rich Receptor

RLK- Receptor-like Kinase

PAMP- Pathogen Associated Molecular Pattern

MAMP- Microbe Associated Molecular Pattern

DAMP- Damage Associated Molecular Pattern

IDA- Inflorescence Deficient in Abscission

BAK1- Brassionsteroid Insensitive Kinase 1

NVT- Number of particles, volume and temperature step

NPT- Number of particles, pressure and temperature step

MD- Molecular Dynamics

RMSD- Root Mean Square Deviation

RMSF- Root Mean Square Fluctuation

SASA- Solvent Accessible Surface Area

PIC- Protein Interaction Calculator

List of Symbols

K - Kelvin

nm - Nano meter

ns - Nanosecond

ps – Picosecond

Chapter 1: Introduction

1.1 Background of Study

In the plant *Arabidopsis thaliana*, the Pattern-Triggered Immunity (PTI) is imparted by the action of pattern recognition receptor (PRR) called HAESA, along with the involvement of signalling peptide Damage Associated Molecular Patterns (DAMP) called IDA. Usually in natural abundance, HAESA and IDA work with Somatic Embryogenesis Receptor-like Kinase protein, abbreviated as SERK1. In other instances, it is found that BAK1 forms a complex with two other proteins FLS2 and flg22, to impart pattern triggered immunity against bacterial flagellin. BAK1, archaically known as SERK3, is a member of the greater SERK protein family of proteins and has some structural and functional similarities to the SERK1 protein. Therefore, this dissertation investigates the possibility of BAK1 potentially being used as a replacement for SERK1 in future studies to interact with HAESA and IDA proteins successfully.

1.2 Significance of Study

This research allows us to observe the changes in structural properties and associated energy coefficients that arise due to the mutation of one protein (BAK1) from the virtually constructed tri-protein complex, HAESA-IDA-BAK1.

1.3 Research Aims and Objectives

This research aims to use biotechnological expertise to investigate how the Pattern Triggered Immunity of the model plant *Arabidopsis thaliana* connected in relation to HAESA- LRR ectodomain is related to the BAK1 protein complex. Specific objectives are fulfilled to reach the final result of the research. Those are overviewed as follows:

1. Examining the relation between PRR HAESA, peptide IDA and co-receptor BAK1 by the use of Molecular Dynamics Simulation
2. Significance of mutation in BAK1 coreceptor in HAESA-IDA-BAK1 complex
3. Examining HAESA-IDA-BAK1 complex in terms of its structural dynamics

1.4 Literature Review

This chapter presents the overview of the binding of protein structures HAESA, IDA and BAK1 which are related in the pattern triggered immune responses and the lateral root development of the plant species *Arabidopsis thaliana*. Reference complexes for the experimental complex of the study, that is, FLS2-flg22-BAK1 complex and HAESA-IDA-SERK1 complex have been studied.

1.5 Plant Immune System

Living organisms display defensive responses towards invasive infectious agents, which are triggered by the invasion of infectious agents to inactivate, neutralize, and kill the infectious agents to impart immunity within the host organism. The combination of immune responses builds the overall broad-spectrum immunity that protects the plant from external and internal harm. [1]

The overall broad-spectrum immunity of plants, otherwise called Systemic Acquired Resistance (SAR), is administered by two types of immunological responses combined, namely – Pattern Trigger Immunity and Effector Triggered Immunity. [2] This defense mechanism was adapted through years of evolution of the selective pressure to distinguish pathogenic substances within the host organisms. Multicellular host organisms can detect non-self compounds from supposedly pathogenic sources through germline-encoded receptors, thus imparting the "innate immunity" in the hosts as a primary layer of defense. This mechanism is referred to as Pattern Triggered Immunity, abbreviated as PTI. [1,6,8]

Eventual signaling events, for example, the activation of the mitogen-activated protein kinase (MAPK) kinase cascades, an influx of Ca^{2+} ion into the cytosol, and the generation of reactive oxygen species trigger the subsequent activation of defense genes that secrete antimicrobial compounds. Subsequently, the secondary layer of PTI defense comes to play, where nucleotide-binding (NB) and leucine-rich-repeat (LRR)-containing receptors (NLRs) detect isolate-specific pathogen effectors. A resultant conformational change triggers the hypersensitive response (HR) or other defense responses that ultimately impart innate immunity as Effector Triggered Immunity, abbreviated as ETI. Contrary to multicellular animals, plants lack specialized immune cells and an adaptive immune system; therefore, they must rely exclusively on innate immune responses that result in local and systemic responses. [5,8] The Pattern Triggered Immunity was formerly

considered to play a minor role in Effector Triggered Immunity. On the other hand, recent investigations have demonstrated the natural occurrences of crosstalk and collaboration between the two types of immune responses. [8]

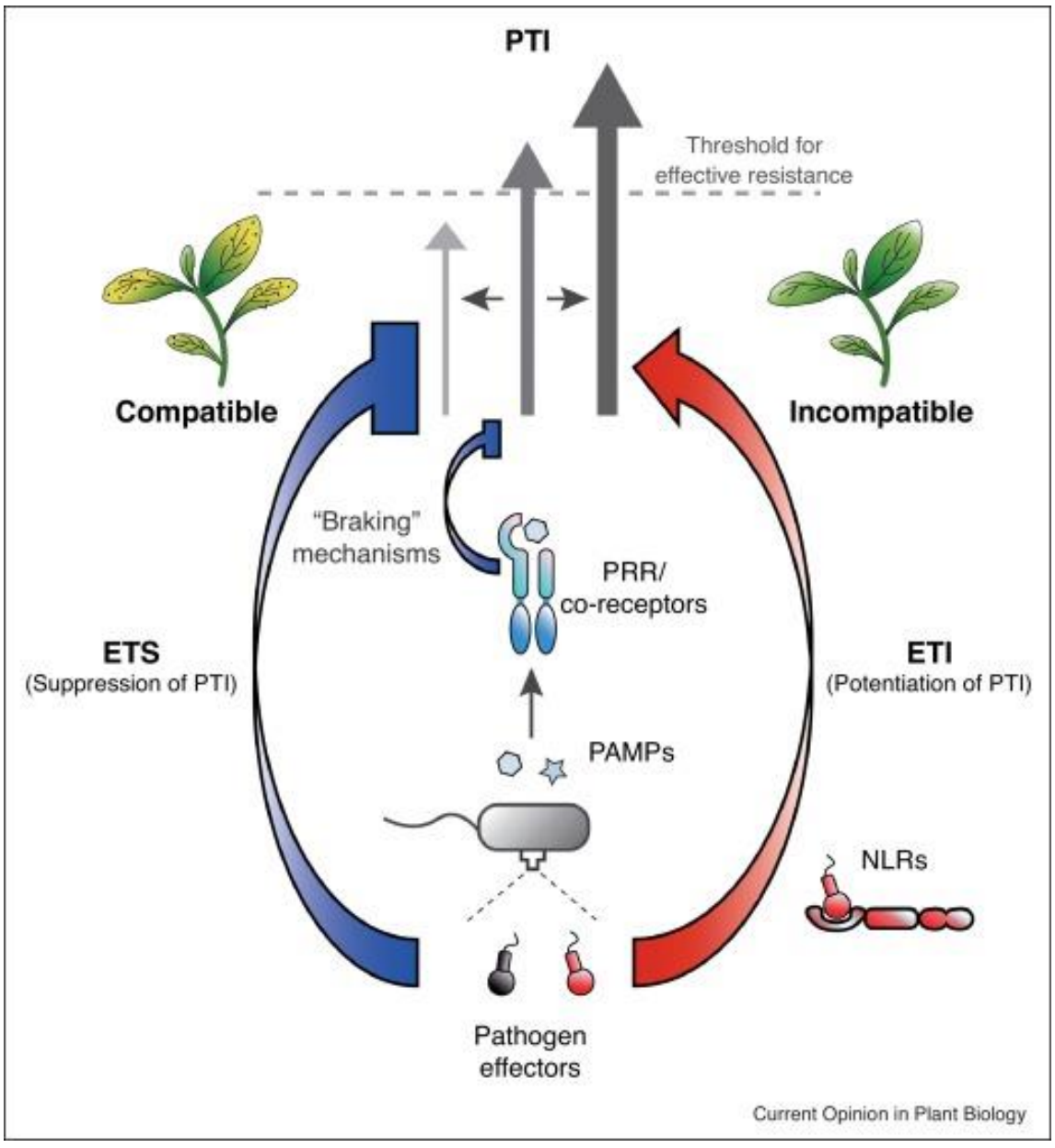

Figure : A brief overview of the plant immune system. [44]

1.6 Plant Pattern Triggered Immunity

The plant Pattern Triggered Immunity, otherwise known as PTI, is triggered when non-self, highly conserved microbial signature molecules called the Pathogen Associated Molecular Patterns (or, PAMPs) come in contact with the plant body. Specialized receptors trigger the surveillance of 'non-self' agents through plant innate immune responses to distinguish patterns called the Pattern Recognition Receptors (PRRs), which recognize conserved patterns called the pathogen-associated molecular patterns PAMPs. Alternatively, the resistance proteins in the plant adaptive immune system perceive isolate-specific pathogen effectors. For the sake of innate immunity, the detection of PAMP is critical to warn the host of the presence of pathogens and the activation of plant defense mechanisms. [3, 4, 6] The PAMPs that are specific to microbes are called Microbe Associated Molecular Pattern or MAMPs. These PAMPs and MAMPs a have evolved in the plant to detect the attacking pathogens or microbes from the harmless non-attacking substances inside or outside the plant body. [3, 4] Besides DAMPs, resulting in the activation of MAPK cascade and eventually the expression of defense genes against the attacking pathogens. [3]

1.7 Plant Effector Triggered Immunity

In plants, although pattern recognition receptors (PRRs) can recognize Pathogen Associated Molecular Patterns (or PAMPs) or Damage Associated Molecular Patterns (or DAMPs), some pathogens might sneak past the PTI system. [3,4] Therefore, as a second line of defense in enforced which uses the R-gene based or the vertical disease transfer-based system, which in turn administers the Effector Triggered Immunity (ETI) for more intense and qualitative responses. [7,10] The intracellular receptor proteins called the NLR receptors are employed by the host plant that mediates polymorphic effectors that are not conserved among microbes for more specific detection to bind Leucine Rich Repeat (NB-LRR). [7,8] These receptors make way for consequent

activation of calcium ion channels, formation of reactive oxygen species, and accumulation of pathogenic proteins, altering plant hormone levels. [10] This process most often includes a programmed cell death (PCD) phenomenon termed the hypersensitive response (HR), the synthesis of salicylic acid (SA) defense hormone, and the induction of defense genes. [8,9]

1.8 *Arabidopsis thaliana*

Arabidopsis thaliana is a small flowering weed of the Brassicaceae family that includes cultivated species such as mustard, cabbage, and radish. It has multiple stages in its life cycle, including seed germination, maturation of rosette, main stem bolting, and flowering. [11, 16] The history of *Arabidopsis thaliana* as a genetic model plant starts with Friedrich Laibach in the late 1800s. He proposed using the mentioned plant since it had a common genetic lineage to most eukaryotic organisms and utility in genetic research, which was finally unanimously agreed upon in the 1980s, in contrast to the other potential model organisms. [12, 13, 14, 15] Characteristics like small generation period, volumetric size, and nuclear genome compared to a more significant offspring birth number made *Arabidopsis thaliana* suitable for the choice. [16]

Figure : Mature *Arabidopsis thaliana* plant. [45]

1.9 Pattern Recognition Receptor

Within the Arabidopsis thaliana, about 600 Receptor Kinases (RKs) are composed of three main components— one transmembrane helix, one cytoplasmic kinase domain, and one ligand-binding domain that resides outside the cellular structure. A plethora of functions are carried out by several RLKs. Notably, BRI1 regulates brassinosteroid signaling, CLV1 regulates meristem growth, FLS2 regulates flagellin perception, Crinkly4 governs leaf development, HAESA regulates abscission, SERKs modulate self-incompatibility, and Xa21 promotes bacterial resistance. [17, 18, 24]

Besides, the organism consists of over 1000 genes that code for 'putative' secreted peptides found for receptor kinases, namely SRK, CLV1, BRI1, and FLS2. These components facilitate proper peptide ligand-receptor interactions, which are pivotal for inter and intra-cell communication. Additionally, RLK kinase domains communicate with multiple proteins. [17, 18, 21, 24]

An important family of receptor kinases termed Leucine-rich repeat RLK (LR-RLK) is the largest and functionally significant. They consist of the Extracellular domain (ECD) containing several LRR repeats that sense small molecules, total protein or peptides; and Kinase domain (KD) containing 12 conserved and three-dimensionally folded subdomains that play crucial roles in enzyme function. [25, 26]

1.10 DAMPs

Pathogen-associated molecular patterns (PAMP) are conserved motifs that work as ligands for host pattern recognition molecules, for example, PRRs (Pattern Recognition Receptors), to activate innate immune responses. [27] Similar kinds of proteins are called DAMPs, which stand for Damage Associated Molecular Patterns. Upon distress or trauma, living organisms release these signaling “danger” peptides. These are endogenous molecules that are excreted from damaged or dying cells, which allows the activation of the innate immune system through association with pattern recognition receptors (PRRs). [28] DAMPs aid PAMPs by enabling them to amplify immune signals in order for PAMPs to detect in the beginning phases of the pathogen intrusion. Examples of DAMPs include Homogalacturonans-derived DAMPs in plant primary cell walls, Plant Elicitor Peptides (PEPs), and PAMP-induced Peptide 1 (PIP1) in *Arabidopsis thaliana*, and Rapid Alkalinization Factors (RALFs) in monocot plants. [7, 24, 29]

1.11 RLK HAESA and DAMP IDA

HAESA is present on the plasma membrane of cells at the base of the petioles and pedicels of plants and in its abscission zones of the floral organs. HAESA was previously known as RLK5 as a receptor-like kinase (RLK) type in the protein receptor family. [17, 30, 31]

On the other hand, the protein Inflorescence Deficient in Abscission, otherwise known as IDA, is an n-terminal signal peptide secreted by an ethylene sensitive *ida* gene that works as ligands to specific plant receptor kinases. [17, 32, 33]

After pollination of the flowers in Arabidopsis thaliana, they are shed by breaking up the Abscission Zone (AZ) in order to release fruits and disperse seeds to facilitate the production of plant progeny. [30, 35] The organs are shed by a mechanism where the IDA hormone binds directly and instructs HAESA to trigger the shedding process. After subsequent morphogenetic changes caused by hormonal changes, cell wall remodeling (CWR) and so forth, SERK1 comes into action. HAESA LRR senses IDA hormone through the co-receptor SERK1 protein and permits the flower shedding process. [22, 30] A good presence of IDA mutant and the HAESA/HLS2 double mutant proteins allows the breakdown of the abscission zone cell layers of the middle lamella to shed or abscise their floral organs from the base. In contrast, the absence of such mutants prevents abscission and therefore keeps the organ attached indefinitely. [19, 22, 23] Besides, an ida mutant gene forms an IDA protein that develops an Abscission Zone, but the non-mutant or wild ida gene that secretes IDA protein lacks Floral Abscission. [34]

Simultaneous to the role in floral abscission, HAESA and IDA also has an important role in the lateral root emergence of Arabidopsis thaliana. By the influence of the plant hormone Auxin, IDA peptide is induced to regulate CWR genes along with HAESA. [19, 31, 32] Alternative to its role

in the floral abscission pathway, the wild *ida* allows lateral root emergence, while the mutant *ida* restricts the activity. [34]

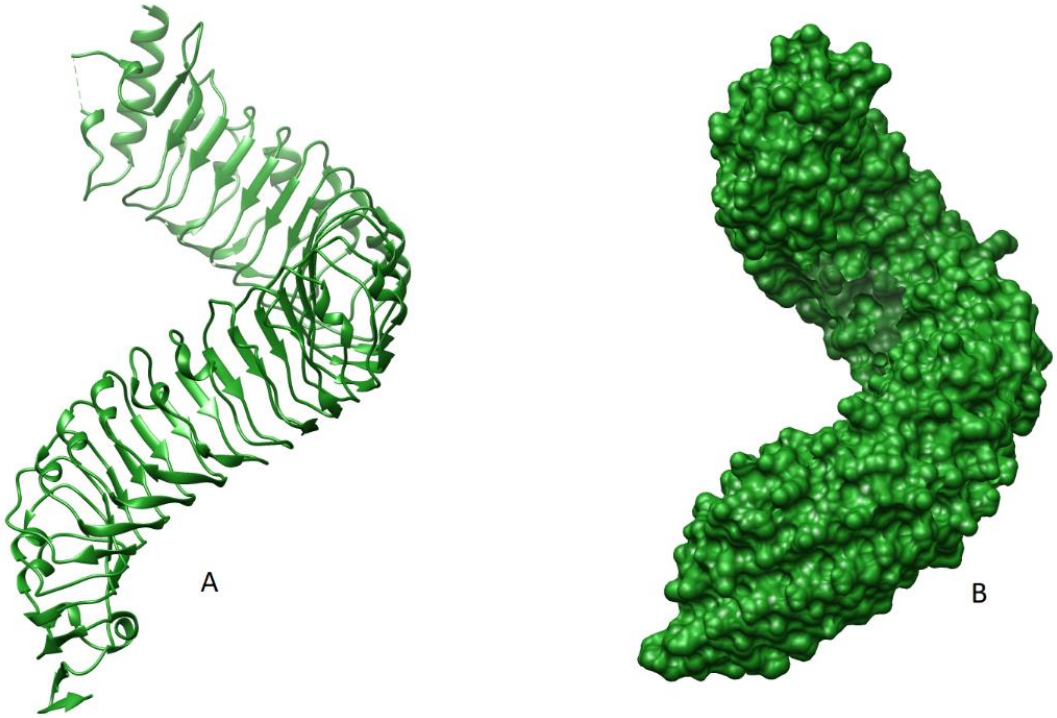

Figure : Ribbon structure (A) and surface view (B) of HAESA ectodomain

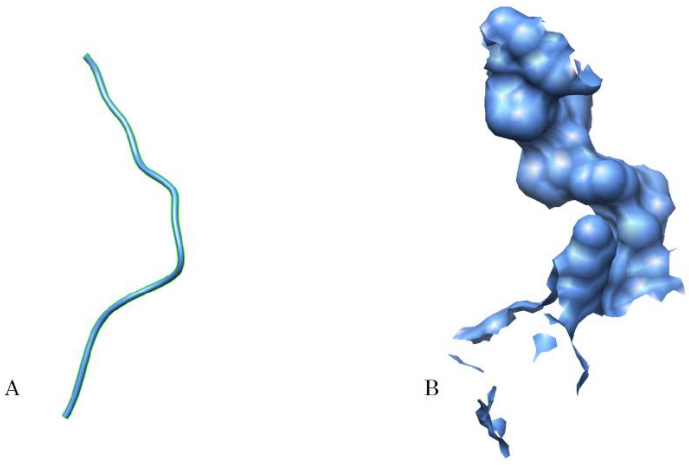

Figure : Ribbon structure (A) and surface view (B) of IDA ectodomain

1.12 Coreceptor BAK1

BRI1-associated kinase 1, otherwise known as *BAK1*, is a leucine-rich repeat receptor-like kinase (LRR-RLK) with five repeats containing the extracellular LRR domain. *BAK1* can modulate PRR responses for downstream PAMP perception events to impart immunity against a vast legion of invading pathogens [36, 37]. It was initially termed SERK3, which stands for Somatic Embryogenesis Receptor-Like Kinase-3, being one of the five members of the SERK (Somatic Embryogenesis Receptor-Like Kinase) family of Leucine-rich RLKs. The SERK LR-RLKs are classified based on their expression during somatic embryogenesis [17, 36, 37]. Besides, SERK3 was renamed as BAK1 also because researchers discovered that SERK3 function as a signaling partner of BRI1, which is the acronym for Brassinosteroid Insensitive1 [40].

BAK1 has small ectodomains composed of five LRRs with a serine and proline rich domain follow the LRR domain, which actually defines the SERK protein family. [17, 33, 40] Additionally, it contains a membrane-spanning domain, a cytoplasmic kinase domain, and a short C-terminal tail. In *Arabidopsis thaliana*, BAK1 (AtSERK3) generates an elicitor-dependent complex with the FLAGELLIN SENSING 2 (FLS2) receptor for the bacterial PAMP flagellin and its peptide derivative flg22. In one example for *Arabidopsis thaliana*, BAK1 (AtSERK3) generates an elicitor-dependent complex with the FLAGELLIN SENSING 2 (FLS2) receptor for the bacterial PAMP flagellin and its peptide derivative flg22 in Arabidopsis thaliana. [17, 40, 42]

Additionally, BAK1 is engaged in identifying the bacterial PAMPs elongation factor Tu (EF-Tu), the cold shock protein PGN, lipopolysaccharides (LPS), DAMP AtPep1, and so forth [39, 40, 41, 42, 43]. Besides signaling in Pattern Triggered Immunity, it also contributes to the modulation and control of brassinosteriod (BR) responses, light signaling, and cell death. [39]

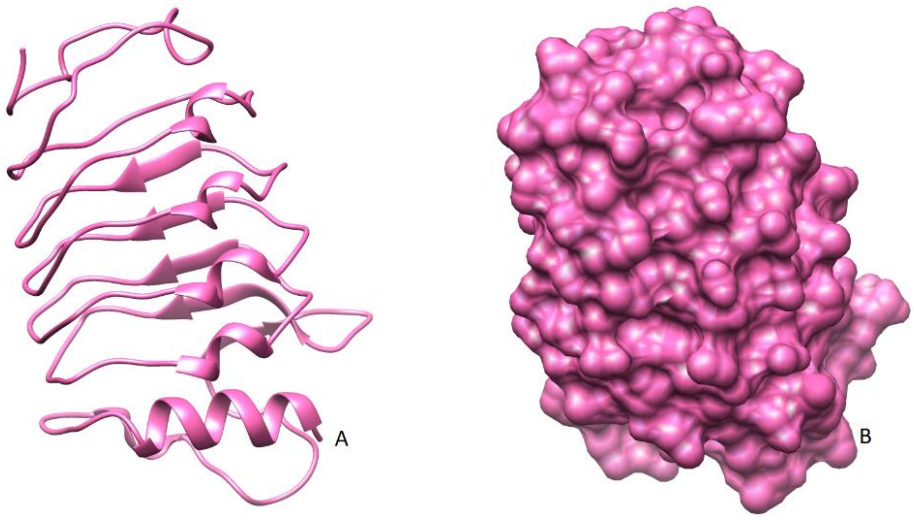

Figure : Ribbon structure (A) and surface view (B) of BAK1 enzyme

1.13 Significance of mutation in BAK1

A point mutation (GGA to GAA) which is precisely 1,370 bp downstream from the transcriptional start site in the third Leucine-rich repeat, was discovered when the BAK1/SERK3 (At4g33430) locus of *elg* (elongated) mutant plants was sequenced. The mutation occurs at the position D122N that changes Aspartate (D-122) to an Asparagine (N-122). [42] In vitro experiments proved the inability of wild-type BAK1 to interact with FLS2, where BRI1 interacts with FLS2. Therefore, an introduction of the D122N mutation in BAK1 is predicted to induce a salt bridge formation within expected amino acids, which is structurally and functionally similar to the D128N mutation in *Oryza sativa* rice OsSERK2 Asp128-Arg152 and Asn128-Glu174. [43] However, the effect of the alteration in Asp122 to Asn122 in this case within BAK1 is yet to be discovered.

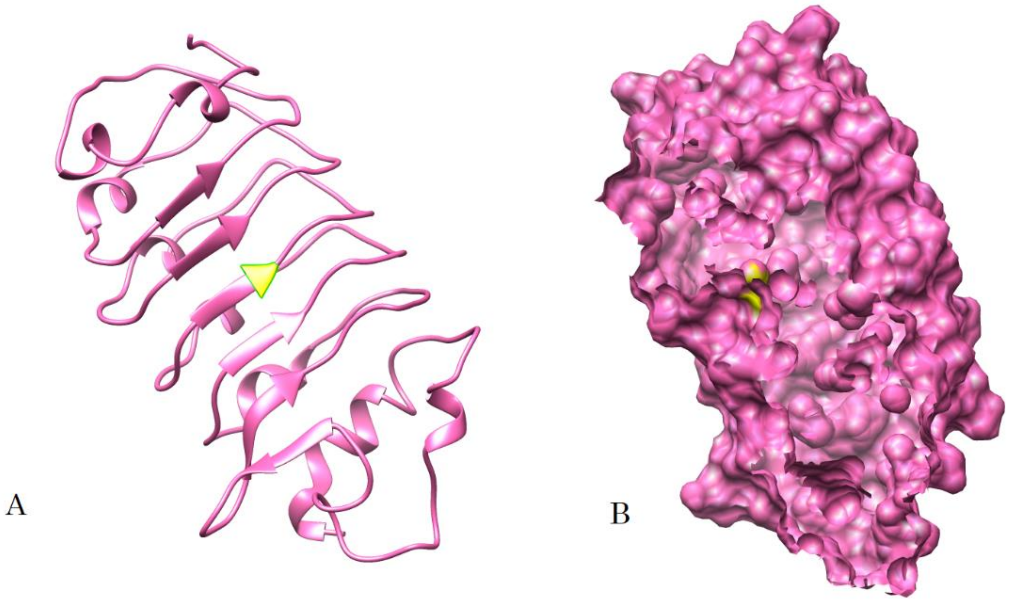

Figure : Ribbon structure (A) and surface view (B) of BAK1 with D122N mutation (highlighted with yellow color)

1.14 Reference Model Structures

For comparison, the structure of the LRR-RLK HAESA, DAMP IDA, and coreceptor SERK1 complex (PDB ID: 5IYX), and the LRR-RLK FLS2, PAMP BAK1 and flagellin peptide flg22 (PDB ID: 4MN8) are taken into consideration. HAESA-IDA-SERK1 functions in the pattern triggered immunity in *Arabidopsis thaliana*. FLS2-flg22-BAK1 functions similarly in the immune response to detect bacterial flagellin and its derivatives. The structural comparison is necessary to observe the docking association of the components of the experimental model structure of HAESA-IDA-BAK1.

1.15 Bioinformatic Methods for Study

Computer simulations are efficient tools that supplement traditional, in-laboratory "wet lab" experiments to thoroughly understand the characteristics of molecular assemblies in terms of structural and microscopic interactions. [46] Molecular dynamics (MD) simulations function based on a general model of physics that governs the inter- or intra-molecular movement of all or some atoms in a protein, or other molecular complex system monitored within a specified time frame. [47] Such simulations capture a broad range of critical biomolecular events, such as conformational change, ligand binding, and protein folding. They are showing their locations at a temporal resolution of time units smaller than seconds. They also demonstrate the interactions of the molecules to systemic shifts, for example, the addition or subtraction of a ligand, mutation, protonation, and so forth. [48]

In order to execute dynamics simulation for a given molecular complex, appropriate programs with suitable computational power are necessary. Simulations run with greater temporal resolutions show the best results by canceling out as much noise as possible. Potential energy functions to study molecular systems are carried out by programs, for instance, CHARMM (Chemistry at

Harvard Macromolecular Mechanics), AMBER (Assisted Model Building with Energy Refinement), GROMOS (GROningen Machines for Chemical Simulations), etc. Similarities measured using Normal Node Refinement methods, such as RMSD (Root Mean Square Deviation) and RMSF (Root Mean Square Fluctuation), are used to refine the structures of extracted molecular motion characteristics. SASA (Solvent Accessible Surface Area) is used to understand the influence of solvation on complex biomolecular associations.

A couple of computational software and tools were used for the purpose of this experiment. The list of such tools is provided below:

1.15.1 GROMACS

"GROningen MAchine for Chemical Simulation", or in short, GROMACS is a molecular dynamics simulation suite. It can simulate molecular interaction of a protein or other complexes of interest within a liquid or membrane system. To generate the required outputs, it employs force fields for bonded, non-bonded or special interactions of the complex of interest for an experiment, calculate and modulate free energy and other complex algorithms. [51, 52]

1.15.2 UCSF Chimera

UCSF Chimera is a graphical representation software that uses Python, C++, and other programming languages to generate three-dimensional graphics of molecular structures and sequences. The structures can be visualized as ribbons, balls and sticks, or surface structures with interacting bonds. [54, 55]

1.15.3 XMGRACE

XMGRACE or Grace is a two-dimensional graphical plotting software to visualize input data points that is based on Python programming language. [56]

1.15.4 ClusPro, PatchDock

ClusPro and PatchDock are online molecular docking servers that allow for protein-ligand and protein-protein docking. The algorithm of this processing software can detect highly populated clusters of low energy conformations, which helps to generate the best possible results of docking in a quick and efficient manner. PatchDock results were furnished by a part of the server called FireDock, which showed the best results with low energy coefficients. The software shows cluster sizes, energy levels, and comparative structural similarities to references, making it easier to choose the perfect docked result [58, 59, 60, 61].

1.15.5 PIC

Protein Interaction Calculator (PIC) server shows the interactions of proteins input into its system. It shows results for various parameters, for example, hydrophobic interactions, hydrogen bonding, main-chain interactions with other main chain/s or side chain/s, side chains interactions, etc. [62]

1.15.6 RMSD

The RMSD (Root Mean Square Deviation) is the mean displacement of the atoms compared to the initial frame, crystallographic structure, or any structure considered a reference at a particular point in the simulation at a given time. This parameter allows to observe the binding affinity of the components of a protein-protein or protein-ligand complex. [57]

1.15.7 RMSF

The RMSF is a scale of the displacement of an individual atom or a group of atoms from the structure of reference whose mean average value is calculated over the number of atoms. This parameter tells whether the protein is folded or unfolded, therefore indicating the strength of interaction of the residues within the complex. [57]

1.15.8 SASA

Solvent-Accessible Surface Area, abbreviated as SASA, is a measure of protein stability that measures the amount of surface of a given biomolecule by a solvent or solvation system. It denotes the stability of the protein complex in the presence of a solvation layer surrounding the complex [63].

1.15.9 Rg

The radius of gyration or Rg is characterized as the root-mean-square average of the distance of all dispersed particles measured from the center of mass of the molecule-of-interest, indicating the compactness of the system. Along with SASA, it allows to measure protein stability in relationship to LRR modules. [63, 64]

Chapter 2: Materials and Methods

2.1 Preparation of the HAESA-IDA protein complex

The crystal structure of *Arabidopsis thaliana* HAESA LRR-RK in complex with the peptide hormone IDA (PDB ID: 5IXQ) [68] is taken and its PDB file is downloaded. The crystal waters, heteroatoms, for example, NAG, SO2, BMA, MAN, MG, EDO, HOH, and other unnecessary information are erased form the file and only the amino acid residue information are kept. The missing atoms of the updated protein structure file are checked with the software "Swiss PdbViewer" [70].

Using the GROMACS version 2020.1 [71], "pdb2gmx", which prompts a list of optional force fields. The force field "OPLS-AA/L: all atom force field" is chosen. Next, an aqueous box is defined and filled with solvent ("SOL" ions) to acquire solvation with SPC water model. Next Energy minimization is run for 50000 number of steps. Next, equilibration is conducted in two phases. First, the NVT equilibration is run for 1000000 for 2000 ps (2 ns). Next the NPT equilibration is run for 1500000 steps for 3000 ps (3 ns).

2.2 Preparation of Mutated BAK1 coreceptor

The crystal structure of flg22 with FLS2 and BAK1 ectodomains (PDB ID: 4MN8) [69] was taken, which was modified to remove heteroatoms along with FLS2 and flg22 chains, keeping only the BAK1 chain. UCSF Chimera software is used to locate the 122nd Amino Acid, Asparatate (Asp) and changed to Asparagine (Asn).

2.3 Docking of Mutated BAK1 to HAESA-IDA protein complex

The formed “npt.gro” file during NPT Equilibration stage is converted to “.pdb” file for interpretation as a receptor file. The mutated BAK1 and the receptor file are input in the following websites:

1. ClusPro
2. PatchDock

In both of these servers, the HAESA-IDA complex is interpreted as the receptor while the mutated BAK1 is interpreted as a ligand to the HAESA-IDA complex. After a few hours, the servers send the resulting docked results through email. Finally, using UCSF Chimera software [54], each docked results are compared with reference crystal structures of FLS2-flg22-BAK1 complex (PDB ID: 4MN8) and HAESA-IDA-SERK1 complex (PDB ID: 5IYX). Besides, the largest cluster size and the lowest energy of the complexes are also considered. Finally, the chosen docked file is prepared for molecular dynamics simulation by correcting the file format with correct parameters readable by GROMACS, namely, adding spacing, name of chains, terminal sequence and atom numbers sequence.

2.4 Molecular Dynamics Simulation of docked complex

Similar to the first step, “pdb2gmx” command is run which prompts a list of optional force fields by using the GROMACS version 2020.1. The force field “OPLS-AA/L: all atom force field” is chosen. Next, an aqueous box is defined and filled with solvent (“SOL” ions) to acquire solvation with SPC water model. Next Energy minimization is run for 50000 number of steps. Next, equilibration is conducted in two phases. First, the NVT equilibration is run for 1000000 for 2000

ps (2 ns). Next the NPT equilibration is run for 1500000 steps for 30 ns. Finally, the Molecular Dynamics simulation is run for 10000000 steps for 30 ns.

Chapter 3: Result and Discussion

3.1 Results:

This section assesses the data and graphical analysis of the interactions between HAESA LRR with signalling peptide IDA and co-receptor BAK1. It also discusses the pre-and post- Molecular Dynamics analyses of the molecular interactions among each component of the tri-protein complex, the molecular interactions between HAESA and IDA, and the molecular interaction between HAESA and SERK1 MD simulation, which are calculated using several tools, for instance: RMSD, RMSF, H-bond, PIC, SASA and Rg.

3.1.1 RMSD

The Root Mean Square Deviation, or in short RMSD values, were generated from the molecular dynamics trajectory for 30 ns. The graph shows the uneven fluctuations that exist for the HAESA-IDA-BAK1 complex. From time intervals 0 ns to 10 ns, the graph fluctuates less. Moving forward to 20 ns shows increasing fluctuations from 0.2 nm to up to 0.5 nm, ranging for about 5 ns interval. This graph has a propensity towards an upward trend, meaning that the components become more unstable with time.

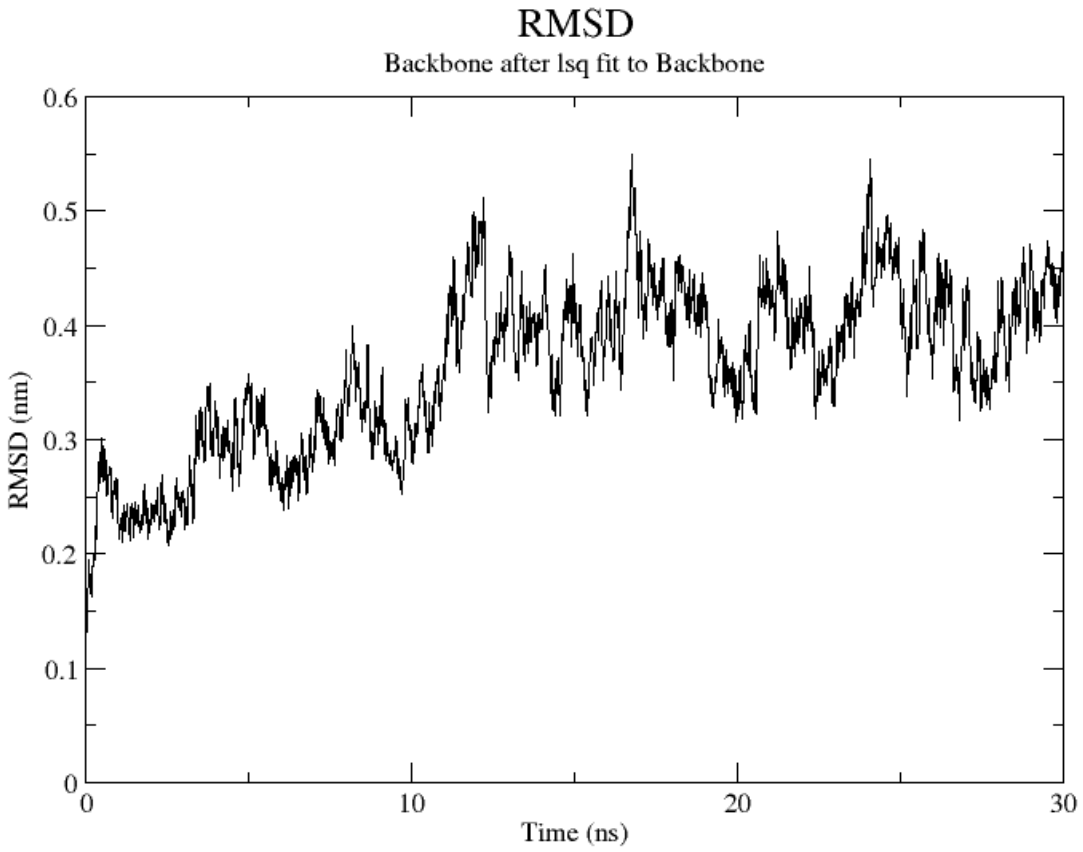

Figure : RMSD graph of HAESA-IDA-BAK1 complex. Black graph shows fluctuations.

3.1.2 RMSF

Root Mean Square Fluctuation, or in short RMSF values, were generated from the molecular dynamics trajectory for 30 ns. Three graphs were generated for HAESA, IDA, and BAK1 separately.

The generated graph for HAESA shows that HAESA within the tri-protein complex has a low fluctuation rate of less than 0.1 nm at the N terminal residues (from residue 1 to about the first 50 residues). Gradual progression towards the C terminal residues shows increasing fluctuations without regular frequency intervals. The RMSF graph for IDA shows a similar high fluctuation ranging from 0.05 nm to 0.3 nm. Finally, the RMSF graph for BAK1 shows that BAK1 within the

HAESA-IDA-BAK1 complex has increased but uniform fluctuations that spike near the C terminal.

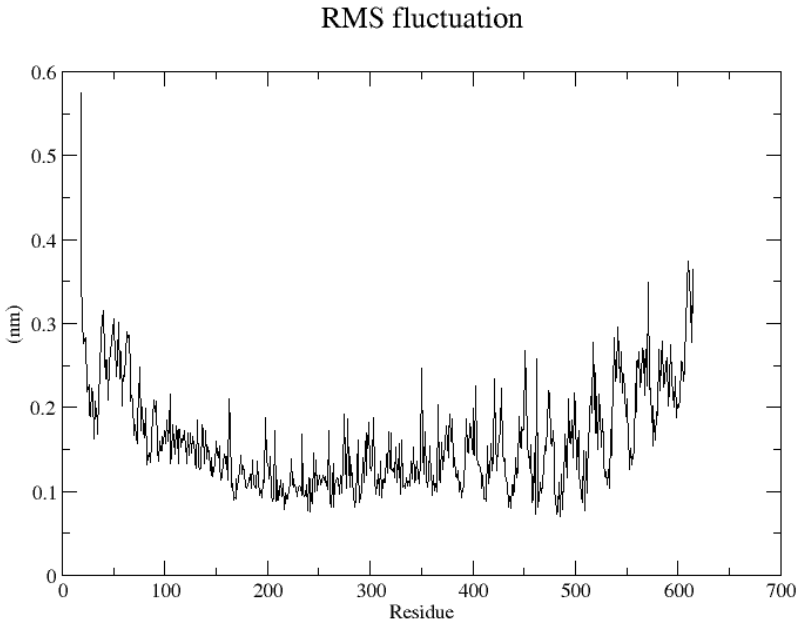

Figure : RMSF value of HAESA from 30 ns Molecular Dynamics trajectories.

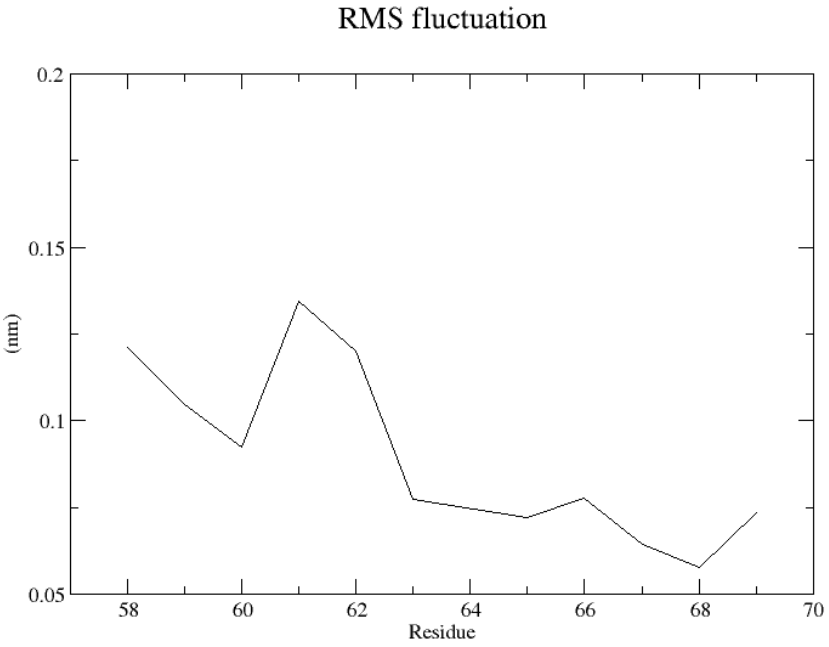

Figure : RMSF value of IDA from 30 ns Molecular Dynamics trajectories.

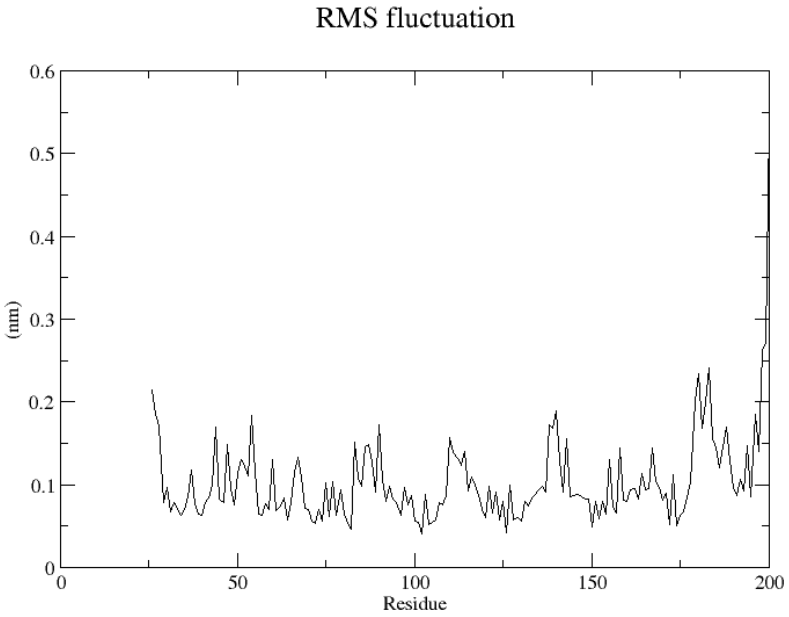

Figure : RMSF value of BAK1 from 30 ns Molecular Dynamics trajectories.

3.1.3 H-bond

Graphical analysis using xmgrace program shows the number of hydrogen bonds (h-bond) created or broken off during the course of the molecular dynamics simulation (30 ns). Here, the hydrogen bond value for the combination of proteins is calculated, viz- HAESE and IDA (Figure 11 A), HAESA-BAK1 (Figure 11 B), and IDA-BAK1 (Figure 11 C) are taken. The black lines denote the hydrogen bonds.

From the graph of the HAESA-IDA complex, it is noticed that the fluctuation is somewhat stable and consistent throughout the entire run time. Alternatively, the graphs for HAESA-BAK1 and IDA-BAK1 show extreme volatility which makes the graph almost unreadable. This means that hydrogen bonds are unstable and subject to frequent alterations.

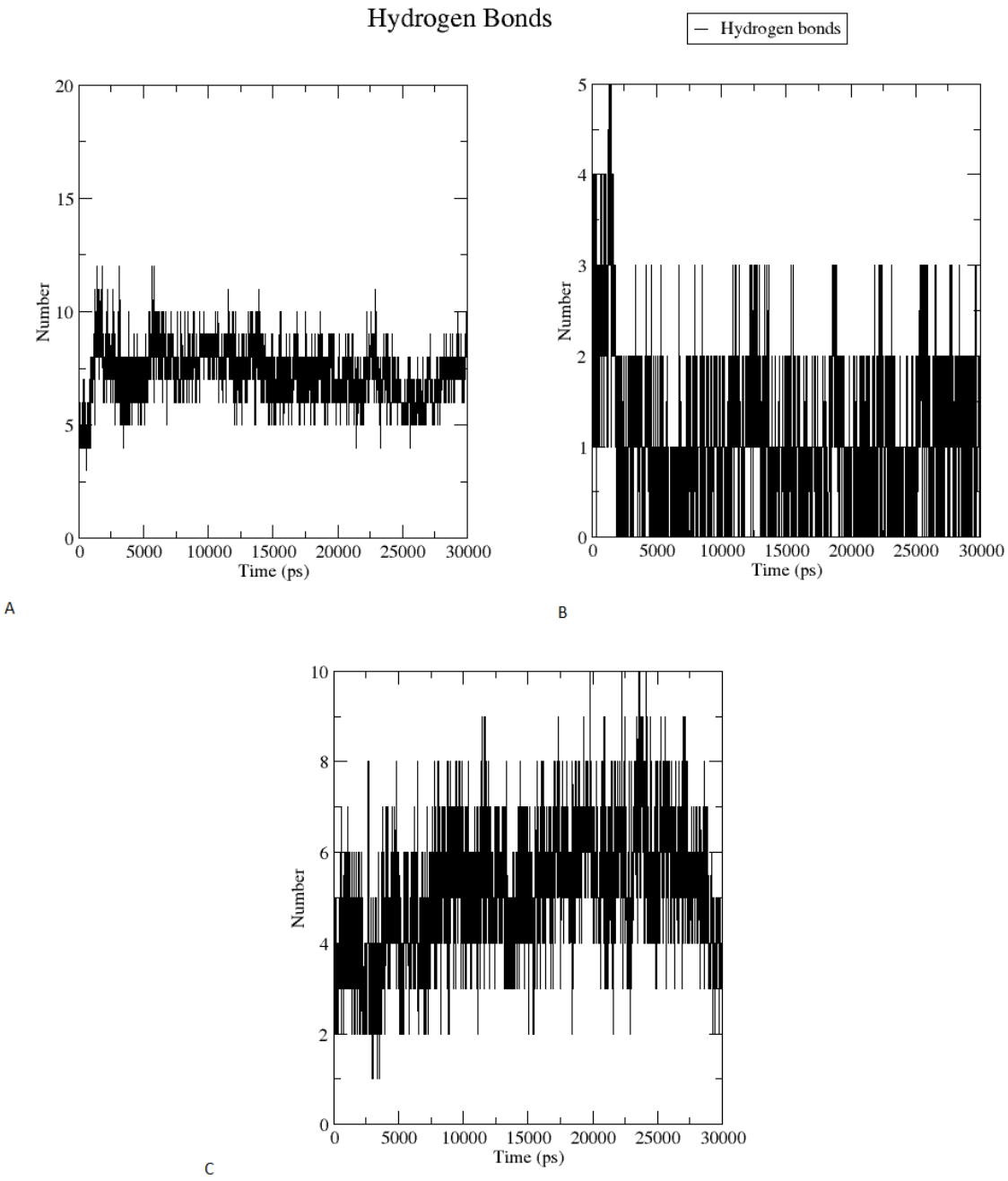

Figure : Hydrogen bond analysis of HAESA-IDA-BAK1 complex. (A) H-bond value of HAESA and IDA from 30 ns trajectory. (B) H-bond value of IDA and BAK1 from 30 ns trajectory. (C) H-bond value of HAESA and BAK1 from 30 ns trajectory.

3.1.4 PIC

PIC allows to perform calculations of the exact bonds within any protein structure. Here, the PIC values of the HAESA-IDA-BAK1 complex were calculated once before the molecular dynamics (MD) simulation was run, and later after MD simulation.

The Protein Interaction calculator yielded several types of bonds of interacting amino acid residues for each component of the complex. Here, the inter-protein interactions in the main chain-main, main chain-side chain, and side chain-side chain are the most important in order to give a clear idea about the interacting hydrogen bond within the experimental protein complex, which will, later on, clarify the stability of the entire complex. Besides the hydrogen bonds, the hydrophobic interactions, ionic interactions, aromatic interactions are also considered. Furthermore, protein-protein disulfide bridges and aromatic sulfur interactions were absent for the complex both before and after the molecular dynamics simulation.

The common bonds between the pre- and post-molecular dynamics simulations are highlighted in the following tables. Additionally, a summary table of all the interacting bonds pre- and post-simulation is given at the end.

3.1.4.1 Hydrophobic Interactions within 5 Angstroms

Table : Hydrophobic Interactions within 5 Angstroms, before and after molecular Dynamics Simulations.

Before Simulation					
Position	**Residue**	**Chain**	**Position**	**Residue**	**Chain**
78	VAL	A	168	VAL	C
125	LEU	A	144	PHE	C
125	LEU	A	168	VAL	C
127	VAL	A	192	ILE	C
171	ALA	A	59	ILE	B
174	PHE	A	100	TYR	C
174	PHE	A	124	TYR	C
174	PHE	A	125	LEU	C
196	TYR	A	59	ILE	B
196	TYR	A	60	PRO	B
218	TRP	A	59	ILE	B
245	PHE	A	45	VAL	C
245	PHE	A	60	PHE	C
268	PHE	A	60	PHE	C
383	TYR	A	54	VAL	C
After Simulation					
Position	**Residue**	**Chain**	**Position**	**Residue**	**Chain**
60	PRO	B	100	TYR	C
61	PRO	B	60	PHE	C
78	VAL	A	188	LEU	C
80	PRO	A	188	LEU	C
125	LEU	A	144	PHE	C
125	LEU	A	168	VAL	C
127	VAL	A	192	ILE	C
171	ALA	A	59	ILE	B
174	PHE	A	124	TYR	C
174	PHE	A	125	LEU	C
196	TYR	A	59	ILE	B

196	TYR	A	60	PRO	B
198	LEU	A	125	LEU	C
218	TRP	A	59	ILE	B
245	PHE	A	45	VAL	C
245	PHE	A	60	PHE	C
245	PHE	A	78	ALA	C
268	PHE	A	60	PHE	C

3.1.4.2 Protein-Protein Main Chain-Main Chain Hydrogen Bonds

Table : Protein-Protein Main Chain-Main Chain Hydrogen Bonds, before and after Molecular Dynamics Simulations.

After Simulation								
Donor				**Acceptor**				**Parameter**
Position	**Chain**	**Residue**	**Atom**	**Position**	**Chain**	**Residue**	**Atom**	**Dd-a**
67	B	ARG	N	58	C	THR	O	3.29

Here,

Dd-a = Distance Between Donor and Acceptor

This parameter is considered for the interacting residues of the experimental complex.

3.1.4.3 Protein-Protein Main Chain-Side Chain Hydrogen Bonds

Table : Protein-Protein Main Chain-Side Chain Hydrogen Bonds, before and after Molecular Dynamics Simulations.

Donor				**Acceptor**				**Parameters**	
Position	**Chain**	**Residue**	**Atom**	**Position**	**Chain**	**Residue**	**Atom**	**Mo**	**Dd-A**
264	A	GLN	NE2	62	B	SER	O	1	2.95
264	A	GLN	NE2	62	B	SER	O	2	2.95
269	A	ASN	ND2	44	C	LYS	O	1	3.03
269	A	ASN	ND2	44	C	LYS	O	2	3.03
407	A	ARG	NE	69	B	ASN	O	-	3.30
407	A	ARG	NE	69	B	ASN	OXT	-	2.89
407	A	ARG	NH2	69	B	ASN	O	1	2.71
407	A	ARG	NH2	69	B	ASN	O	2	2.71
407	A	ARG	NH2	52	C	THR	O	1	2.56
407	A	ARG	NH2	52	C	THR	O	2	2.56
407	A	ARG	NH1	53	C	LEU	O	1	2.75
407	A	ARG	NH1	53	C	LEU	O	2	2.75
407	A	ARG	NH2	53	C	LEU	O	1	2.66
407	A	ARG	NH2	53	C	LEU	O	2	2.66
409	A	ARG	NH1	69	B	ASN	O	1	2.72
409	A	ARG	NH1	69	B	ASN	O	2	2.72
66	B	LYS	NZ	57	C	CYS	O	-	2.51
66	B	LYS	NZ	59	C	TRP	O	-	2.62
66	B	LYS	NZ	62	C	VAL	O	-	2.54
67	B	ARG	NH1	47	C	GLN	O	1	2.71
67	B	ARG	NH1	47	C	GLN	O	2	2.71
69	B	ASN	N	361	A	ASP	OD2	-	3.20
77	C	ASN	OD1	173	A	ASN	O	1	3.37
77	C	ASN	OD1	173	A	ASN	O	2	3.37
77	C	ASN	ND2	173	A	ASN	O	1	2.80

77	C	ASN	ND2	173	A	ASN	O	2	2.80
77	C	ASN	ND2	196	A	TYR	O	1	2.85
77	C	ASN	ND2	196	A	TYR	O	2	2.85
100	C	TYR	OH	172	A	GLY	O	-	2.63
167	C	GLN	NE2	76	A	MET	O	1	2.92
167	C	GLN	NE2	76	A	MET	O	2	2.92
After Simulation									
Donor				**Acceptor**				**Parameters**	
Position	**Chain**	**Residue**	**Atom**	**Position**	**Chain**	**Residue**	**Atom**	**Mo**	**Dd-A**
102	A	ASN	ND2	188	C	LEU	O	1	2.80
102	A	ASN	ND2	188	C	LEU	O	2	2.80
269	A	ASN	ND2	43	C	ASN	O	1	2.71
269	A	ASN	ND2	43	C	ASN	O	2	2.71
269	A	ASN	ND2	44	C	LYS	O	1	3.40
269	A	ASN	ND2	44	C	LYS	O	2	3.40
337	A	LYS	NZ	67	B	ARG	O	-	2.86
66	B	LYS	NZ	57	C	CYS	O	-	3.02
66	B	LYS	NZ	62	C	VAL	O	-	2.83
69	B	ASN	N	361	A	ASP	OD2	-	2.88
44	C	LYS	NZ	270	A	ASN	O	-	2.77
44	C	LYS	NZ	293	A	MET	O	-	2.66
77	C	ASN	OD1	196	A	TYR	O	1	3.23
77	C	ASN	OD1	196	A	TYR	O	2	3.23
190	C	THR	N	102	A	ASN	OD1	-	2.94

3.1.4.4 Protein-Protein Side Chain-Side Chain Hydrogen Bonds

Table : Protein-Protein Side Chain-Side Chain Hydrogen Bonds, before and after Molecular Dynamics Simulations.

Donor				**Acceptor**				**Parameters**	
Position	**Chain**	**Residue**	**Atom**	**Position**	**Chain**	**Residue**	**Atom**	**Mo**	**Dd-A**
150	A	ASN	OD1	124	C	TYR	OH	1	2.69
150	A	ASN	OD1	124	C	TYR	OH	2	2.69
150	A	ASN	ND2	124	C	TYR	OH	1	2.96
150	A	ASN	ND2	124	C	TYR	OH	2	2.96
62	B	SER	OG	242	A	ASP	OD2	-	2.86
67	B	ARG	NE	293	A	MET	SD	-	3.54
67	B	ARG	NH1	293	A	MET	SD	1	3.42
67	B	ARG	NH1	293	A	MET	SD	2	3.42
67	B	ARG	NH2	293	A	MET	SD	1	3.98
67	B	ARG	NH2	293	A	MET	SD	2	3.98
67	B	ARG	NE	316	A	GLU	OE1	-	2.87
67	B	ARG	NH2	316	A	GLU	OE1	1	2.73
67	B	ARG	NH2	316	A	GLU	OE1	2	2.73
67	B	ARG	NH2	316	A	GLU	OE2	1	2.67
67	B	ARG	NH2	316	A	GLU	OE2	2	2.67
68	B	HIS	NE2	58	C	THR	OG1	-	3.06
69	B	ASN	ND2	50	C	ASP	OD2	1	2.83
69	B	ASN	ND2	50	C	ASP	OD2	2	2.83
69	B	ASN	ND2	52	C	THR	OG1	1	3.38
69	B	ASN	ND2	52	C	THR	OG1	2	3.38
50	C	ASP	OD2	69	B	ASN	ND2	1	2.83
50	C	ASP	OD2	69	B	ASN	ND2	2	2.83
52	C	THR	OG1	69	B	ASN	ND2	-	3.38
58	C	THR	OG1	68	B	HIS	NE2	-	3.06
124	C	TYR	OH	150	A	ASN	OD1	-	2.69
124	C	TYR	OH	150	A	ASN	ND2	-	2.96

143	C	ARG	NH1	76	A	MET	SD	1	3.40
143	C	ARG	NH1	76	A	MET	SD	2	3.40

After Simulation

Donor				**Acceptor**				**Parameters**	
POS	**CHAIN**	**RES**	**ATOM**	**POS**	**CHAIN**	**RES**	**ATOM**	**MO**	**Dd-a**
100	A	SER	OG	167	C	GLN	OE1	-	3.47
223	A	ASN	OD1	79	C	ASN	ND2	1	3.45
223	A	ASN	OD1	79	C	ASN	ND2	2	3.45
361	A	ASP	OD1	69	B	ASN	ND2	1	3.40
361	A	ASP	OD1	69	B	ASN	ND2	2	3.40
361	A	ASP	OD2	69	B	ASN	ND2	1	2.75
361	A	ASP	OD2	69	B	ASN	ND2	2	2.75
67	B	ARG	NH1	316	A	GLU	OE1	1	3.17
67	B	ARG	NH1	316	A	GLU	OE1	2	3.17
67	B	ARG	NH2	316	A	GLU	OE1	1	2.90
67	B	ARG	NH2	316	A	GLU	OE1	2	2.90
69	B	ASN	ND2	361	A	ASP	OD1	1	3.40
69	B	ASN	ND2	361	A	ASP	OD1	2	3.40
69	B	ASN	ND2	361	A	ASP	OD2	1	2.75
69	B	ASN	ND2	361	A	ASP	OD2	2	2.75
79	C	ASN	ND2	223	A	ASN	OD1	1	3.45
79	C	ASN	ND2	223	A	ASN	OD1	2	3.45
167	C	GLN	OE1	100	A	SER	OG	1	3.47
167	C	GLN	OE1	100	A	SER	OG	2	3.47

3.1.4.5 Protein-Protein Ionic Interactions

Table : Protein-Protein Ionic Interactions, before and after Molecular Dynamics Simulation

Before Simulation					
Position	**Residue**	**Chain**	**Position**	**Residue**	**Chain**
123	GLU	A	72	ARG	C
316	GLU	A	67	ARG	B
After Simulation					
Position	**Residue**	**Chain**	**Position**	**Residue**	**Chain**
67	ARG	B	50	ASP	C
316	GLU	A	67	ARG	B

3.1.4.6 Protein-Protein Aromatic-Aromatic Interactions

Table : Protein-Protein Aromatic-Aromatic Interactions, before and after Molecular Dynamics Simulation.

Before Simulation							
Position	**Residue**	**Chain**	**Position**	**Residue**	**Chain**	**D(centroid-centroid)**	**Dihedral Angle**
174	PHE	A	100	TYR	C	4.53	57.41
174	PHE	A	124	TYR	C	5.79	40.71
245	PHE	A	60	PHE	C	4.94	116.52
268	PHE	A	60	PHE	C	5.26	10.84
After Simulation							
Position	**Residue**	**Chain**	**Position**	**Residue**	**Chain**	**D(centroid-centroid)**	**Dihedral Angle**
174	PHE	A	124	TYR	C	5.45	73.36
245	PHE	A	60	PHE	C	6.42	82.56
268	PHE	A	60	PHE	C	5.80	66.26

3.1.4.7 Protein-Protein Cation-Pi Interactions

Table : Protein-Protein Cation-Pi Interactions

Before Simulation							
Position	**Residue**	**Chain**	**Position**	**Residue**	**Chain**	**D(cation-Pi)**	**Angle**
60	PHE	C	67	ARG	B	5.94	128.83
After Simulation							
Position	**Residue**	**Chain**	**Position**	**Residue**	**Chain**	**D(cation-Pi)**	**Angle**
339	PHE	A	67	ARG	B	5.39	84.41

3.1.4.8 Summary of interactions among HAESA, IDA and BAK1 proteins before and after simulation

Table : Summary of Interactions among HAESA, IDA and BAK1, before and after simulation

Complex	**Interacting proteins**	**H-bond**		**Hydrophobic Interactions**		**Ionic Interaction**		**Cation Pi-Interaction**		**Aromatic - Aromatic Interaction**		**Aromatic - Sulphur Interaction**	
		B.m.d.	A.m.d.	B.m.d.	A.m.d.	B.m.d.	A.m.d.	B.m.d.	A.m.d.	B.m.d.	A.m.d.	B.m.d.	B.m.d.
HAE-IDA-BAK1	HAE-IDA	20	14	4	4	1	1	0	1	0	0	0	0
	HAE-BAK1	25	18	11	12	1	0	0	0	4	3	0	0
	IDA-BAK1	14	3	0	2	0	1	1	0	0	0	0	0
HAE-IDA-SERK 1[67]	HAE-IDA	36	14	5	7	1	0	0	0	0	0	0	0
	HAE-SERK 1	35	74	9	8	4	2	1	1	1	0	0	0
	IDA-SERK 1	4	3	0	0	1	0	0	0	0	0	0	0

3.1.5 Radius of Gyration (Rg)

The radius of gyration (Rg) denotes the compactness of a given molecular complex. The HAESA-IDA-BAK1 complex was processed to generate the radius of gyration at 30 ns run time. The graph below illustrates the state of the radius of gyration of the complex HAESA-IDA-BAK1. Here, the graph is seen to be fluctuating from the beginning, which gradually increases with the increasing time parameter. Although all radius of gyration graphs usually fluctuate until they reach a stable time stamp, this graph does not show any sign of stability throughout the run time.

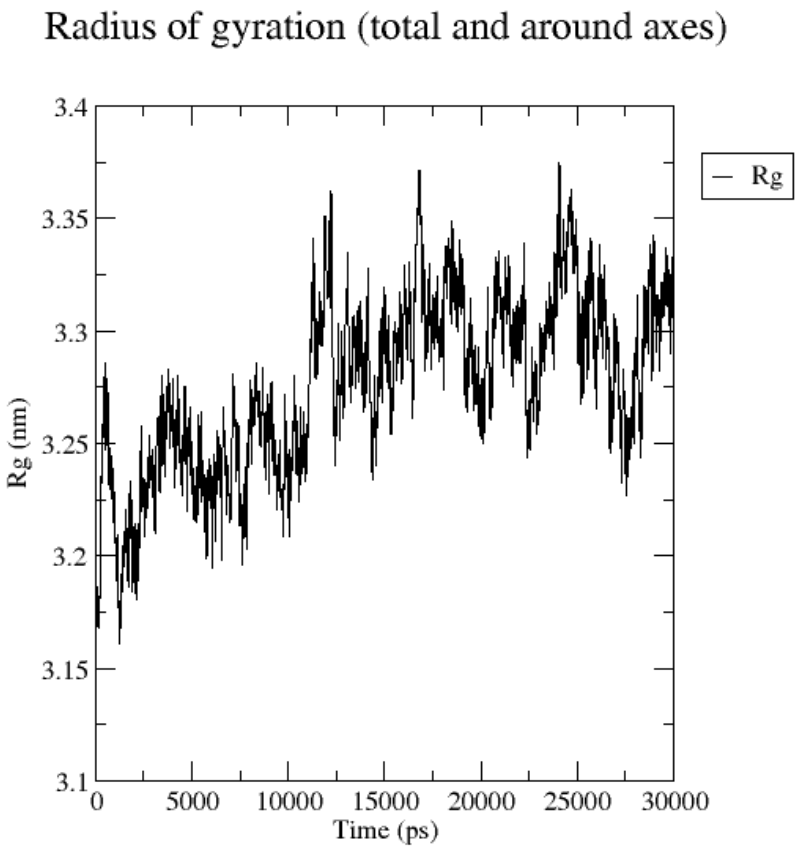

Figure : Radius of gyration (Rg) of HAESA-IDA-BAK1 complex

3.1.6 Solvent-Accessible Surface Area (SASA)

The Solvent-Accessible Surface Area (SASA) graph helps to measure the compactness of a given protein complex. For the HAESA-IDA-BAK1 complex, SASA analysis was run for 30 ns. The graph is observed to fluctuate through 5 ns intervals. The significant fluctuation in the area (nm^2) values implies that the residues are largely dispersed throughout the solvation area, which holds the HAESA-IDA-BAK1 complex.

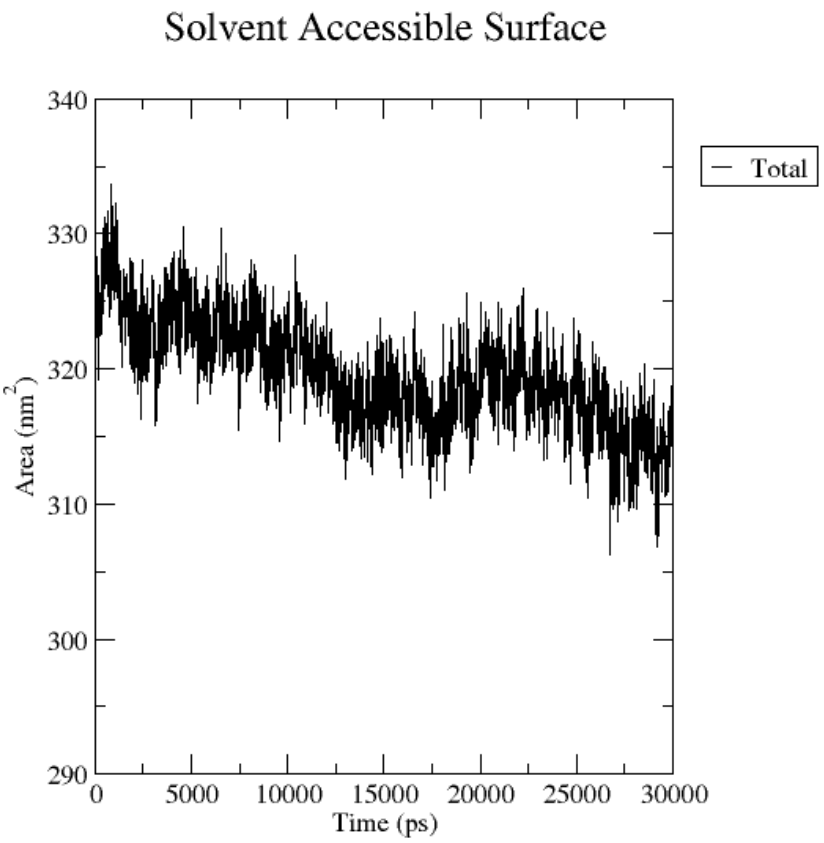

Figure : Solvent Accessible Surface Area (SASA) of HAESA-IDA-BAK1 complex

3.1.7 Structural Comparison of Docked Model with Reference Models

The docked complex of LRR HAESA, DAMP IDA, and coreceptor BAK1 is constructed from the references of two reference complexes, viz.- HAESA-IDA-SERK1 complex in *Arabidopsis thaliana* (PDB ID: 5IYX) [72] and flg22-BAK1-FLS2 complex in *Arabidopsis thaliana* (PDB ID: 4MN8) [69]. The comparative illustration shows that the binding of LRR HAESA and peptide IDA in HAE-IDA-BAK1 is visually similar to the construction of LRR HAESA and DAMP IDA from the HAE-IDA-SERK1 reference complex. On the other hand, the coreceptor BAK1 in the HAE-IDA-BAK1 complex is positioned similarly to the BAK1 present in the fls2-bak1-flg22 complex. This structural comparison can acquire not much information about the residual changes through visualizing the molecular structures.

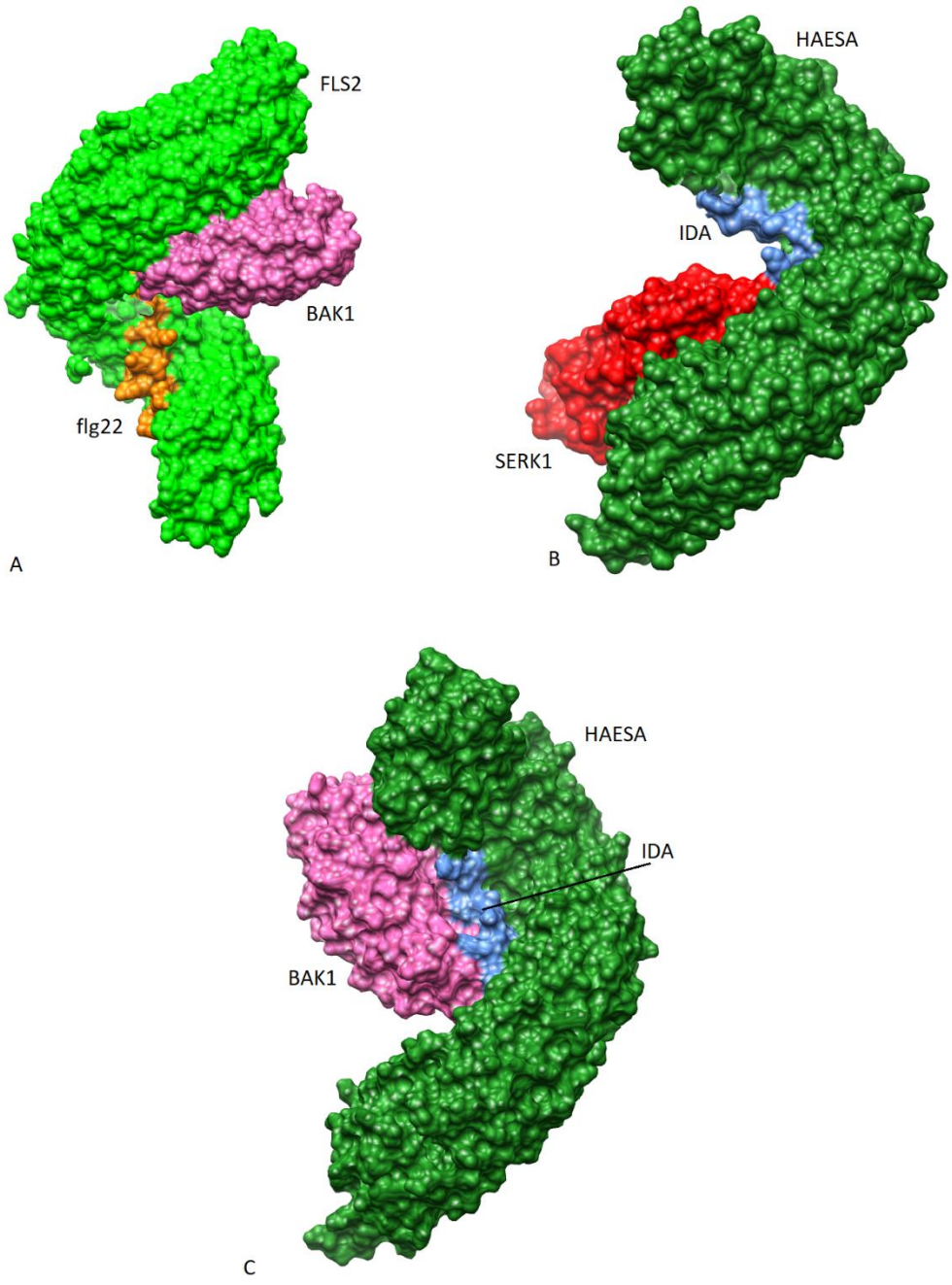

Figure : Structural Comparison (surface) between (A)FLS2-flg22-BAK1, (B)HAESA-IDA-SERK1 and (C)HAESA-IDA-BAK1

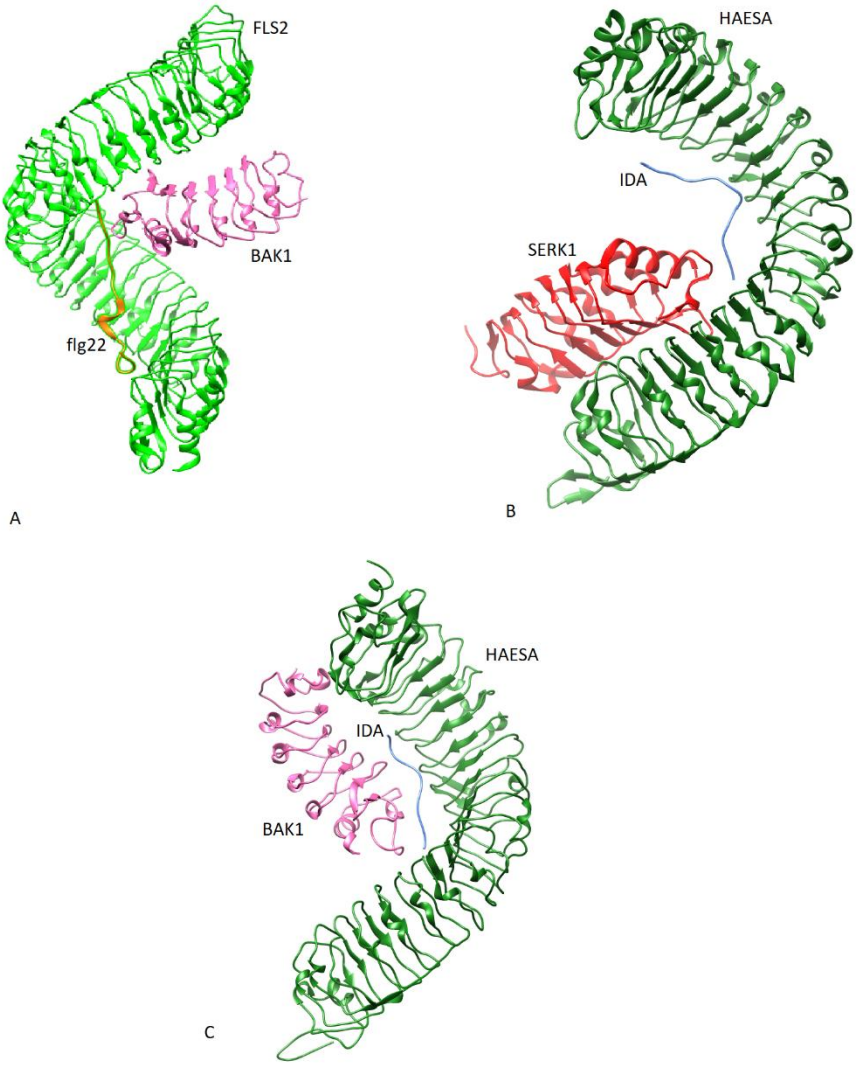

Figure : Structural Comparison (ribbon) between (A)FLS2-flg22-BAK1, (B)HAESA-IDA-SERK1 and (C)HAESA-IDA-BAK1

3.2 Discussion

This section of the dissertation describes the computational analysis of PRR-RLK HAESA, which imparts Pattern Triggered immunity in *Arabidopsis thaliana* and its interaction with DAMP IDA and Co-receptor BAK1.

Firstly, the Root Mean Square Deviation or, in short, RMSD was determined for the HAESA-IDA-BAK1 complex. The graph's large deviation values and erratic fluctuations imply that the association of HAESA LRR and IDA peptide with the co-receptor BAK1 is an unstable complex. The high fluctuations tell that the components of the complex have a low binding affinity; therefore, the root-mean-square deviation value cannot be efficiently generated. Besides, irregular spikes in data could occur due to the short simulation period of 30 ns used for the experiment. A simulation of 100ns would smoothen out the fluctuations and could yield a more reliable result.

Next, three separate graphs determined the Root Mean Square Fluctuation, or in brief, RMSF, for the HAESA-IDA-BAK1 complex. RMSF data for all three, viz- HAESA, IDA, and BAK1, show that they fluctuate a lot. Large amounts of residual fluctuations hint that the amino acids within the tri-protein complex are unstable, therefore, incapable of efficiently interacting with surrounding residues. The N-terminal residues of HAESA have more fluctuations because BAK1 binds to the N-terminal residues of HAESA.

Hydrogen bonding after the simulation was determined using a graphical method using xmgrace, and before and after simulations using the Protein Interaction Calculator. The reference data for HAESA-IDA-SERK1 protein complex was taken for comparison with the HAESA-IDA-BAK1 complex. [67, 72] Both kinds of data infer that the hydrogen bonds are greatly unstable within the

complex. The constant bonds before and after molecular dynamics simulation do not show any effect on the mutated BAK1 protein.

The Radius of Gyration or Rg gives an idea about the compactness of protein, calculated by taking the distance between the center of mass of the protein to both of its terminal ends. It allows calculating the degree of folding of the protein structure. The Rg value of the HAESA-IDA-BAK1 protein complex demonstrates that the complex lacks rigidity. Since no reliably rigid interactions are observed in the complex, the result tells that the Rg parameter does not support the formation of the HAESA-IDA-BAK1 protein complex similar to the previously mentioned parameters.

The Solvent Accessible Surface Area or SASA measures how much solvent is accessed or allowed upon the surface of the protein-of-interest. The graph for SASA shows massive fluctuations of residues per unit area, which implies that the residues on the surface are far apart and non-interactive. Hence the surface formed by the HAESA-IDA-BAK1 is unsustainable. The solvents from the solvation box, which was made to observe the molecular dynamics, enters through the surface of the protein complex. Strong residual interactions would allow the protein complex to occupy a condensed area in an ideal output and prevent solvents from entering inside the complex.

The structural comparison of the reference models FLS2-flg22-BAK1 and HAESA-IDA-SERK1 and consequently comparison with HAESA-IDA-BAK1 complex before and after the simulation does not show a massive change in the visual structure.

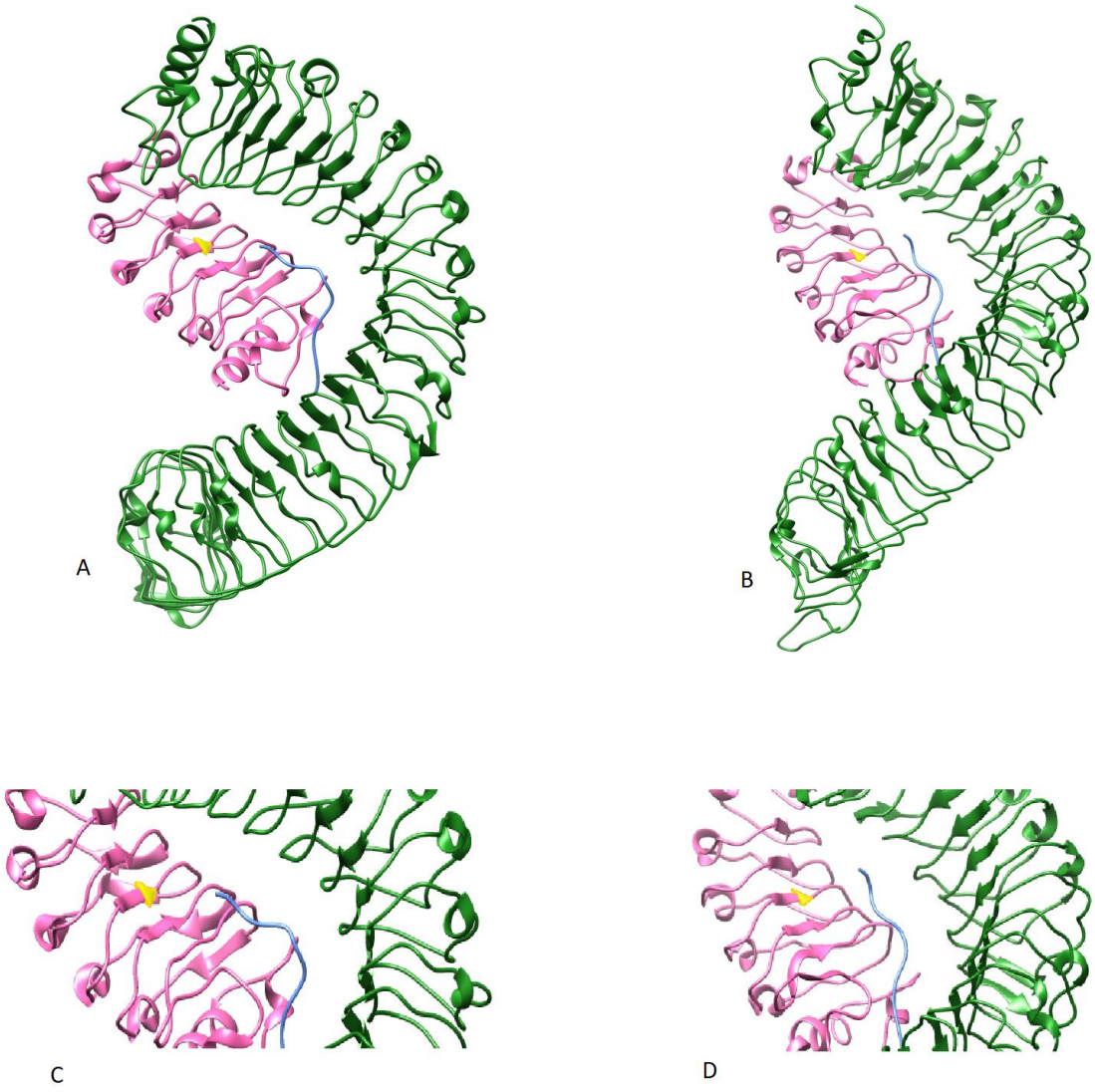

Figure : Structural Differences (ribbon) of HAESA-IDA-BAK1 complex, before (A) and after (B) Molecular Dynamics Simulation. Zoomed view of mutated BAK1, before (C) and after (D) Molecular Dynamics Simulation.

Chapter 4: Conclusion and Recommendation

4.1 Conclusion

After the molecular dynamics simulation at 30 ns and subsequent confirmatory examinations, it can be said that the mutated BAK1 cannot be associated with HAESA LRR and IDA peptide. It was hypothesized in this paper that BAK1, similar to SERK1 protein kinase, could bind to the HAESA-IDA complex to activate pattern-triggered immunity or other alternative functions to SERK1. The hypothesis is proven wrong because the tri-protein association is proven unfavorable. No literature has shown an available mutation of SERK1 protein that is naturally present for BAK1 protein. Therefore, the association of mutated BAK1 to the PRR-signal proteins cannot be transcribed as equal or valid to the association of SERK1 to the same complex.

4.2 Recommendations

The research done here can be further developed by adopting some measures.

Firstly, this study can be improved by running the molecular dynamic simulation for much longer durations, for instance, 50 ns or 100 ns. This would yield more excellent temporal resolution and allow more selective and conclusive results from the study to improve understanding of protein nature that could be constructed. This experiment run at 30 ns could not yield reliably good results.

Secondly, the study might be helpful in the interaction of HAESA LRR with other mutated PAMPs in plant *Arabidopsis thaliana*. The interaction of the HAESA PRR and IDA signalling peptide complex with the mutated co-receptor BAK1 does not work, but an equivalent mutation in the coreceptor SERK1 can show how mutant molecules can affect specific residues, ultimately leading to pattern triggered immunity (PTI).

Lastly, post-processing free energy calculation methods, for instance, MM/PBSA could be used to calculate the free energy from the interior energy (MM) of the residues and estimating its

connection with a solvent (PBSA). For solvation-free energies of proteins, DNA, and other molecules, MM/PBSA is sufficiently precise while avoiding the enormous expense that would result from using decoupling methods. [66]

References

1. Medzhitov, R., & Janeway, C. (1997). Innate Immunity: The Virtues of a Nonclonal System of Recognition. Cell, 91(3), 295-298. https://doi.org/10.1016/s0092-8674(00)80412-2

2. Nishad, R., Ahmed, T., Rahman, V., & Kareem, A. (2020). Modulation of Plant Defense System in Response to Microbial Interactions. Frontiers In Microbiology, 11. https://doi.org/10.3389/fmicb.2020.01298

3. Zhang, J., & Zhou, J. (2010). Plant Immunity Triggered by Microbial Molecular Signatures. Molecular Plant, 3(5), 783-793. https://doi.org/10.1093/mp/ssq035

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

5. Zipfel, C. (2014). Plant pattern-recognition receptors. *Trends In Immunology*, *35*(7), 345-351. https://doi.org/10.1016/j.it.2014.05.004

6. Sanabria, N., Goring, D., Nürnberger, T., & Dubery, I. (2008). Self/nonself perception and recognition mechanisms in plants: a comparison of self-incompatibility and innate immunity. New Phytologist, 178(3), 503-514. https://doi.org/10.1111/j.1469-8137.2008.02403.x

7. Boller, T., & Felix, G. (2009). A Renaissance of Elicitors: Perception of Microbe-Associated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors.

Annual Review Of Plant Biology, 60(1), 379-406. https://doi.org/10.1146/annurev.arplant.57.032905.105346

8. Nguyen, Q., Iswanto, A., Son, G., & Kim, S. (2021). Recent Advances in Effector-Triggered Immunity in Plants: New Pieces in the Puzzle Create a Different Paradigm. International Journal Of Molecular Sciences, 22(9), 4709. https://doi.org/10.3390/ijms22094709

9. Jones, J., & Dangl, J. (2006). The plant immune system. Nature, 444(7117), 323-329. https://doi.org/10.1038/nature05286

10. Tsuda, K., & Katagiri, F. (2021). Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. Retrieved 7 August 2021. https://doi.org/10.1016/j.pbi.2010.04.006

11. TAIR - About Arabidopsis. Arabidopsis.org. (2021). Retrieved 7 August 2021, from https://www.arabidopsis.org/portals/education/aboutarabidopsis.jsp.

12. Laibach, F. (1943). Arabidopsis thaliana (L.) Heynh. als Objekt für genetische und entwicklungsphysiologische Untersuchungen. Bot. Archiv, 44, 439-455.

13. Meyerowitz, E. (2001). Prehistory and History of Arabidopsis Research. Plant Physiology, 125(1), 15-19. https://doi.org/10.1104/pp.125.1.15

14. Nishimura, M., & Dangl, J. (2010). Arabidopsis and the plant immune system. The Plant Journal, 61(6), 1053-1066. https://doi.org/10.1111/j.1365-313x.2010.04131.x

15. Woodward, A., & Bartel, B. (2018). Biology in Bloom: A Primer on the Arabidopsis thaliana Model System. Genetics, 208(4), 1337-1349. https://doi.org/10.1534/genetics.118.300755

16. Meinke, D. (1998). Arabidopsis thaliana: A Model Plant for Genome Analysis. Science, 282(5389), 662-682. https://doi.org/10.1126/science.282.5389.662

17. Hohmann, U., Lau, K., & Hothorn, M. (2017). The Structural Basis of Ligand Perception and Signal Activation by Receptor Kinases. Annual Review Of Plant Biology, 68(1), 109-137. https://doi.org/10.1146/annurev-arplant-042916-040957

18. Shiu, S., & Bleecker, A. (2001). Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proceedings Of The National Academy Of Sciences, 98(19), 10763-10768. https://doi.org/10.1073/pnas.181141598

19. Kumpf, R., Shi, C., Larrieu, A., Sto, I., Butenko, M., & Peret, B. et al. (2013). Floral organ abscission peptide IDA and its HAE/HSL2 receptors control cell separation during lateral root emergence. Proceedings Of The National Academy Of Sciences, 110(13), 5235-5240. https://doi.org/10.1073/pnas.1210835110

20. Lease, K., & Walker, J. (2006). The Arabidopsis Unannotated Secreted Peptide Database, a Resource for Plant Peptidomics. Plant Physiology, 142(3), 831-838. https://doi.org/10.1104/pp.106.086041

21. Dunning, F., Sun, W., Jansen, K., Helft, L., & Bent, A. (2007). Identification and Mutational Analysis of Arabidopsis FLS2 Leucine-Rich Repeat Domain Residues That

Contribute to Flagellin Perception. The Plant Cell, 19(10), 3297-3313. https://doi.org/10.1105/tpc.106.048801

22. Santiago, J., Brandt, B., Wildhagen, M., Hohmann, U., Hothorn, L., Butenko, M., & Hothorn, M. (2016). Mechanistic insight into a peptide hormone signaling complex mediating floral organ abscission. Elife, 5. https://doi.org/10.7554/elife.15075

23. Niederhuth, C., Cho, S., Seitz, K., & Walker, J. (2013). Letting Go is Never Easy: Abscission and Receptor-Like Protein Kinases. Journal Of Integrative Plant Biology, 55(12), 1251-1263. https://doi.org/10.1111/jipb.12116

24. De Lorenzo, G., Brutus, A., Savatin, D., Sicilia, F., & Cervone, F. (2011). Engineering plant resistance by constructing chimeric receptors that recognize damage-associated molecular patterns (DAMPs). FEBS Letters, 585(11), 1521-1528. https://doi.org/10.1016/j.febslet.2011.04.043

25. Hanks, S.K. and T.J.T.F.j. Hunter, Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. 1995. 9(8): p. 576-596. 42.

26. Hanks, S.K., A.M. Quinn, and T.J.S. Hunter, The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. 1988. 241(4861): p. 42-52.

27. Heise, M. (2014). Viral Pathogenesis. Reference Module In Biomedical Sciences. https://doi.org/10.1016/b978-0-12-801238-3.00079-9

28. Roh, J., & Sohn, D. (2018). Damage-Associated Molecular Patterns in Inflammatory Diseases. *Immune Network*, *18*(4). https://doi.org/10.4110/in.2018.18.e27

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

30. Taylor, I., Wang, Y., Seitz, K., Baer, J., Bennewitz, S., Mooney, B., & Walker, J. (2016). Analysis of Phosphorylation of the Receptor-Like Protein Kinase HAESA during Arabidopsis Floral Abscission. PLOS ONE, 11(1), e0147203. https://doi.org/10.1371/journal.pone.0147203

31. Tsung-Luo Jinn, J. (2021). HAESA, an Arabidopsis leucine-rich repeat receptor kinase, controls floral organ abscission. PubMed Central (PMC). Retrieved 11 August 2021, from http://www.ncbi.nlm.nih.gov/pmc/articles/pmc316334/.

32. Estornell, L., Wildhagen, M., Pérez-Amador, M., Talón, M., Tadeo, F., & Butenko, M. (2015). The IDA Peptide Controls Abscission in Arabidopsis and Citrus. Frontiers In Plant Science, 6. https://doi.org/10.3389/fpls.2015.01003

33. Aalen, R., Wildhagen, M., Stø, I., & Butenko, M. (2013). IDA: a peptide ligand regulating cell separation processes in Arabidopsis. Journal Of Experimental Botany, 64(17), 5253-5261. https://doi.org/10.1093/jxb/ert338

34. Butenko, M. (2003). INFLORESCENCE DEFICIENT IN ABSCISSION Controls Floral Organ Abscission in Arabidopsis and Identifies a Novel Family of Putative Ligands in Plants. THE PLANT CELL ONLINE, 15(10), 2296-2307. https://doi.org/10.1105/tpc.014365

35. Brandt, B., & Hothorn, M. (2016). SERK co-receptor kinases. Current Biology, 26(6), R225-R226. https://doi.org/10.1016/j.cub.2015.12.014

36. Kim, B., Kim, S., & Nam, K. (2012). Assessing the diverse functions of BAK1 and its homologs in arabidopsis, beyond BR signaling and PTI responses. Molecules And Cells, 35(1), 7-16. https://doi.org/10.1007/s10059-013-2255-3

37. Heese, A., Hann, D., Gimenez-Ibanez, S., Jones, A., He, K., & Li, J. et al. (2007). The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. *Proceedings Of The National Academy Of Sciences*, *104*(29), 12217-12222. https://doi.org/10.1073/pnas.0705306104

38. Valérie Hecht, S. (2021). The Arabidopsis Somatic Embryogenesis Receptor Kinase 1 Gene Is Expressed in Developing Ovules and Embryos and Enhances Embryogenic Competence in Culture. PubMed Central (PMC). Retrieved 10 August 2021, from http://www.ncbi.nlm.nih.gov/pmc/articles/pmc129253/.

39. Li, J., Wen, J., Lease, K., Doke, J., Tax, F., & Walker, J. (2002). BAK1, an Arabidopsis LRR Receptor-like Protein Kinase, Interacts with BRI1 and Modulates Brassinosteroid Signaling. Cell, 110(2), 213-222. https://doi.org/10.1016/s0092-8674(02)00812-7

40. Shan, L., He, P., Li, J., Heese, A., Peck, S., & Nürnberger, T. et al. (2008). Bacterial Effectors Target the Common Signaling Partner BAK1 to Disrupt Multiple MAMP Receptor-Signaling Complexes and Impede Plant Immunity. Cell Host & Microbe, 4(1), 17-27. https://doi.org/10.1016/j.chom.2008.05.017

41. Krol, E., Mentzel, T., Chinchilla, D., Boller, T., Felix, G., & Kemmerling, B. et al. (2010). Perception of the Arabidopsis Danger Signal Peptide 1 Involves the Pattern Recognition Receptor AtPEPR1 and Its Close Homologue AtPEPR2. Journal Of Biological Chemistry, 285(18), 13471-13479. https://doi.org/10.1074/jbc.m109.097394

42. Whippo, C., & Hangarter, R. (2005). A Brassinosteroid-Hypersensitive Mutant of BAK1 Indicates That a Convergence of Photomorphogenic and Hormonal Signaling Modulates Phototropism. Plant Physiology, 139(1), 448-457. https://doi.org/10.1104/pp.105.064444

43. McAndrew, R., Pruitt, R., Kamita, S., Pereira, J., Majumdar, D., & Hammock, B. et al. (2014). Structure of the OsSERK2 leucine-rich repeat extracellular domain. Acta Crystallographica Section D Biological Crystallography, 70(11), 3080-3086. https://doi.org/10.1107/s1399004714021178

44. Yuan, M., Ngou, B., Ding, P., & Xin, X. (2021). PTI-ETI crosstalk: an integrative view of plant immunity. Current Opinion In Plant Biology, 62, 102030. https://doi.org/10.1016/j.pbi.2021.102030

45. Page, D., & Grossniklaus, U. (2002). The art and design of genetic screens: Arabidopsis thaliana. Nature Reviews Genetics, 3(2), 124-136. https://doi.org/10.1038/nrg730

46. Allen, M. (2004). Introduction to Molecular Dynamics Simulation (1st ed., pp. 1-28). John von Neumann Institute for Computing.

47. Karplus, M., & McCammon, J. (2002). Molecular dynamics simulations of biomolecules. Nature Structural Biology, 9(9), 646-652. https://doi.org/10.1038/nsb0902-646

48. Hollingsworth, S., & Dror, R. (2018). Molecular Dynamics Simulation for All. Neuron, 99(6), 1129-1143. https://doi.org/10.1016/j.neuron.2018.08.011

49. Becker, O.M., et al., Computational biochemistry and biophysics. 2001: Marcel Dekker New York.

50. Van Gunsteren, W.F., P.K. Weiner, and A.J. Wilkinson, Computer simulation of biomolecular systems: theoretical and experimental applications. Vol. 3. 2013: Springer Science & Business Media.

51. Van Der Spoel, D., et al., GROMACS: fast, flexible, and free. 2005. 26(16): p. 1701-1718.

52. Van der Spoel, D., P.J. van Maaren, and H.J.J.T.J.o.c.p. Berendsen, A systematic study of water models for molecular simulation: derivation of water models optimized for use with a reaction field. 1998. 108(24): p. 10220-10230.

53. Desta, I., Porter, K., Xia, B., Kozakov, D., & Vajda, S. (2020). Performance and Its Limits in Rigid Body Protein-Protein Docking. Structure, 28(9), 1071-1081.e3. https://doi.org/10.1016/j.str.2020.06.006

54. Pettersen, E., Goddard, T., Huang, C., Couch, G., Greenblatt, D., Meng, E., & Ferrin, T. (2004). UCSF Chimera? A visualization system for exploratory research and analysis. Journal Of Computational Chemistry, 25(13), 1605-1612. https://doi.org/10.1002/jcc.20084

55. Couch, G. (2006). Nucleic acid visualization with UCSF Chimera. Nucleic Acids Research, 34(4), e29-e29. https://doi.org/10.1093/nar/gnj031

56. Turner PJ: XMGRACE, Version 5.1.19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR; 2005.

57. Martínez, L. (2015). Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. PLOS ONE, 10(3), e0119264. https://doi.org/10.1371/journal.pone.0119264

58. Vajda, S., Yueh, C., Beglov, D., Bohnuud, T., Mottarella, S., & Xia, B. et al. (2017). New additions to the Clus Pro server motivated by CAPRI. Proteins: Structure, Function, And Bioinformatics, 85(3), 435-444. https://doi.org/10.1002/prot.25219

59. Kozakov, D., Hall, D., Xia, B., Porter, K., Padhorny, D., & Yueh, C. et al. (2017). The ClusPro web server for protein–protein docking. Nature Protocols, 12(2), 255-278. https://doi.org/10.1038/nprot.2016.169

60. Duhovny D, Nussinov R, Wolfson HJ. Efficient Unbound Docking of Rigid Molecules. In Gusfield et al., Ed. Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in Computer Science 2452, pp. 185-200, Springer Verlag, 2002. https://bioinfo3d.cs.tau.ac.il/PatchDock/wabi02.pdf

61. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Research, 33(Web Server), W363-W367. https://doi.org/10.1093/nar/gki481

62. K. G. Tina, R. Bhadra and N. Srinivasan, PIC: Protein Interactions Calculator, Nucleic Acids Research, 2007, Vol. 35, Web Server issue W473–W476.

63. Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S., Xia, B., Hall, D., & Vajda, S. (2013). How good is automated protein docking?. Proteins: Structure, Function, And Bioinformatics, 81(12), 2159-2166. https://doi.org/10.1002/prot.24403

64. Lee, Y., Lee, J., Kim, S., Lee, S., Han, J., & Heu, W. et al. (2014). Dissecting the Critical Factors for Thermodynamic Stability of Modular Proteins Using Molecular Modeling Approach. Plos ONE, 9(5), e98243. https://doi.org/10.1371/journal.pone.009824

65. Rhodes, G. (2006). Other Diffraction Methods. Crystallography Made Crystal Clear, 211-235. https://doi.org/10.1016/b978-012587073-3/50011-8

66. Wang, C., Greene, D., Xiao, L., Qi, R., & Luo, R. (2018). Recent Developments and Applications of the MMPBSA Method. Frontiers In Molecular Biosciences, 4. https://doi.org/10.3389/fmolb.2017.00087

67. Alvy, Raghib Ishraq. (2019). *In silico* analysis of structural complex of HAESA receptor like kinase with PAMP IDA and SERk1 co-receptor. BRAC University. http://hdl.handle.net/10361/13759

68. Santiago, J., Hothorn, M. (2016) Crystal structure of the Arabidopsis receptor kinase HAESA LRR ectdomain in complex with the peptide hormone IDA. Elife 5. doi: 10.2210/pdb5IXQ/pdb

69. Chai, J., Sun, Y., Han, Z. (2013) Crystal structure of flg22 in complex with the FLS2 and BAK1 ectodomains. Science 342: 624-628. doi: 10.2210/pdb4MN8/pdb

70. Guex, N. and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18, 2714-2723.

71. GROMACS version 2020.1, https://doi.org/10.5281/zenodo.3685919 (2020).

72. Santiago, J., Hothorn, M. (2016) Crystal structure of the Arabidopsis receptor kinase HAESA in complex with the peptide hormone IDA and the co-receptor SERK1. Elife 5. doi: 10.2210/pdb5IYX/pdb