In silico modelling and structural dynamics of PRR CORE ectodomain and its interaction with PAMP csp22

A DISSERTATION SUBMITTED TO BRAC UNIVERSITY IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN BIOTECHNOLOGY

Submitted by:

Tanvir Jawad

Student ID: 17376003

November, 2018

Biotechnology Program

Department of Mathematics and Natural Sciences

BRAC University

Dhaka, Bangladesh.

DECLARATION

I hereby solemnly declare that the thesis project titled **"*In silico* modelling and structural dynamics of PRR CORE ectodomain and its interaction with PAMP csp22."** submitted by the undersigned has been carried out under the supervision of M H M Mubassir, Lecturer, Biotechnology Program, Department of Mathematics and Natural Sciences, BRAC University.

The presented dissertation is based on original research work carried out by myself and has not been submitted to any other institution for any degree or diploma. Any reference to work done by any other person or institution or any material obtained from other sources have been accordingly cited and referenced.

(Tanvir Jawad)

Candidate

Certified

(M H M Mubassir)
Supervisor
Lecturer
Biotechnology Program
Department of Mathematics and Natural Sciences
BRAC University, Dhaka.

ACKNOWLEDGEMENT

My sincere thanks also go to **Prof. A F M Yusuf Haider**, Ph.D, Chairperson, Department of Mathematics and Natural Sciences, BRAC University, **Prof. Naiyyum Choudhury,** Ph.D, Former Coordinator, Department of Mathematics and Natural Sciences, BRAC University and **Prof. Aparna Islam**, Ph.D, Former Coordinator, Department of Mathematics and Natural Sciences, BRAC University for granting me permission to continue my research work in my desired subject at BRAC University Bioinformatics Lab.

I would like to express my gratitude to my supervisor, **M H M Mubassir**, Lecturer, Biotechnology Program, Department of Mathematics and Natural Sciences for his support, supervision and cooperation. Without his guidance, I would have never been able to finish my research.

I also want to thank all the employees of the Bioinformatics lab, BRAC University for their support and enthusiasm, and even more so for their never dying recuperation in the course of my work.

Tanvir Jawad

November, 2018

ABSTRACT

Plants being sessile organisms are continuously being subjected to pathogens prevailing in their environment. Understanding the theory behind it would be a great step towards understanding the mechanisms making plants disease resistant. There are two ways in which plant defences are activated- first by structural interaction between the pathogen-associated molecular pattern (PAMP) and the pattern-recognition receptor (PRR) known as pattern-triggered immunity (PTI) and secondly through effectors known as effector triggered immunity (ETI). In PTI, to combat the pathogens, plants employ PRRs which detect PAMPs and employs co-receptor proteins. The aim of this study is to acquire a better understanding of the early stages of PTI mediated by PRR CORE and PAMP csp22 by modelling of these followed by docking and molecular dynamics (MD) simulation using GROMACS software suite. The leucine rich repeat (LRR) on the PRR is responsible for binding to the PAMP, so different *in silico* modelling approaches were used to acquire the CORE LRR 3D structure. Of which only the model generated by I-TASSER using the threading method gave the best results with verification tools such as ERRAT, Verify 3D and Ramachandran plot. The docking result also shows the PAMP csp22 binds at one lateral side of the CORE LRR with the co-receptor BAK1 attaching head on, on to the same lateral side, which is very consistent with the protein interaction observed in the reference FLS2 crystalline complex. The interactions between the three proteins were also analyzed using the protein interaction calculator (PIC) and it was seen that after the MD simulation the number of hydrogen bonds formed between them almost became half. Starting with 42 H-bonds before the simulation, whereas only 22 afterwards. These changes are significant indicators of conformational changes that take place over the simulation period and are vital in understanding the early events of PTI by the receptor protein CORE.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA	-	Amino Acid
ETI	-	Effector Triggered Immunity
EM	-	Energy Minimization
Fig	-	Figure
GROMACS	-	Groningen Machine for Chemical Simulations
H-bonds	-	Hydrogen Bonds
LRR	-	Leucine Rich Repeat Domain
MD	-	Molecular Dynamics
NMR	-	Nuclear Magnetic Resonance
PAMP	-	Pathogen Associated Molecular Pattern
PBC	-	Periodic Boundary Conditions
PDB	-	Protein Data Bank
PRR	-	Pattern Recognition Receptor
PRR-RKs	-	Pattern Recognition Receptor like kinases
PTI	-	Pattern Triggered Immunity
Rg	-	Radius of Gyration
RMSD	-	Root Means Square Deviation
RMSF	-	Root Means Square Fluctuations

LIST OF SYMBOLS

Aº	-	Angstrom
K	-	Kelvin
nm	-	Nano meter
ns	-	Nano second
ps	-	Pico second
α	-	Alpha
β	-	Beta
%		Percent

CHAPTER 1

INTRODUCTION

1.1 Background of the study

The plant pattern recognition receptor protein kinase (PRR-PK) CORE is activated when the pathogen associated molecular pattern (PAMP) csp22 interacts with its leucine rich repeat (LRR) region and recruits the co-receptor BAK1 to help with the process, which then positively regulates the pattern triggered immunity (PTI) (Stefanie, 2017). To better understand the activation mechanisms of these vital immunity regulating proteins, *in silico* approach of modelling, docking and MD simulation were conducted. Then the acquired results were compared to the established crystalline structures of plant PRR such as FLS2, which mediate a similar type of activity (Sun *et al.*, 2013). This *in silico* study on the PRR CORE is the first of its kind, looking into the structural basis of the protein activity in great details.

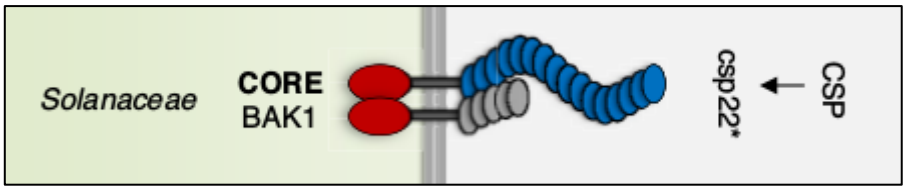


Fig 1.1: PTI mediated by CORE in tomato plant (Stefanie, 2017)

1.2 Significance of the studying

There are two ways in which plant defences are enhanced- first through antimicrobial compounds and secondly PTI. The first method invokes biosafety issues where as the second does not. This is why it is an important sector to study when considering

improvements to plant's defence mechanisms. Thus the modelling of CORE and its PAMP csp22 is imperative. Also it is very important to see the interactions between these two proteins and the recruited co-receptor protein BAK1, in order to fully understand the mechanism of the first layer of defence in plants.

1.3 Research aim and objectives

The primary goal of this work is to understand PTI of tomato plant mediated by pattern recognition receptor CORE using bioinformatics approaches. Baring that is mind, the following achievements was intended:

- To construct a model of plant immune receptor CORE and bacterial PAMP csp22 and to validate the model.
- To construct the docking complex of CORE, csp22 and BAK1 to analyze the interactions between them.
- Analyze the docked complex by comparing with the present PRR-PAMP-Coreseptor complex crystalline structure of FLS2-flg22-BAK1.

1.4 Literature review

This chapter presents the overview of plant pattern triggered immunity mediated by the pattern recognition receptor CORE, on members of the Solanaceae family. At the end of the chapter different tools used for this study are briefly described.

1.4.1 Introduction to plant immune system

To date it is understood that plants employ two layers of defense mechanisms to impart immunity to them against invading pathogens. The first is knows as pattern triggered immunity (PTI) and the second layer known as effector triggered immunity (ETI). This process of defense mechanism was clearly illustrated by Jonathan in 2006 through a zigzag model. According to their model, plant confers this immunity following four

phases. In the first phase, different pathogen associated molecular patterns (PAMPs) or microbes-associated molecular patterns (MAMPs) are recognized by different pattern recognition receptors of plant. This recognition of PAMPs by PRR results in PTI. Some successful pathogen can successfully evade the PTI of plant which results in phase two effector susceptibility (ETS) (Fig 1.2). In this case pathogens deploy effectors and escape PTI. These effectors are again recognized by nucleotide binding leucine rich receptors (NB-LRRs) which activates ETI, the third phase of the zigzag model. In the final phase, the pathogen gains new effectors which can again suppress ETI (Fig 1).

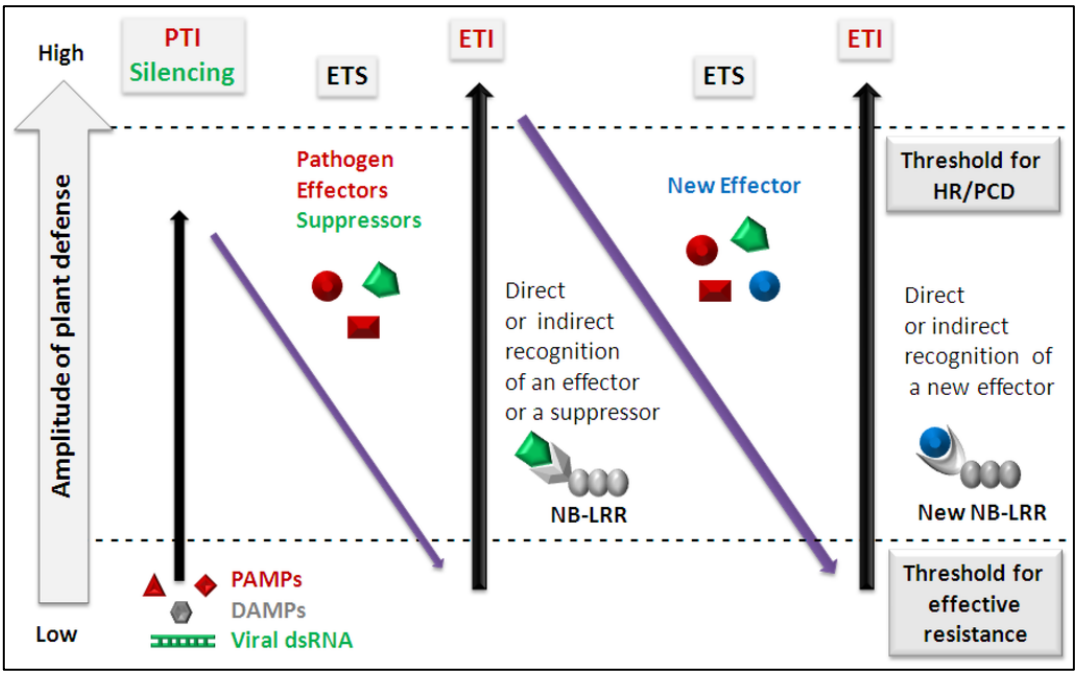


Fig 1.2: ZigZag model of plant immune system (Jones and Dangl, 2006)

1.4.2 Pattern triggered immunity (PTI)

The first layer of defense against invading pathogens is known as Pattern triggered immunity (PTI) (Jones and Dangl, 2006; Ronald and Beutler, 2010; Tsuda *et al.*, 2009). Plants contain an arsenal of receptor proteins on the cell surface membranes, and these proteins ultimately play a vital role in PTI. These receptors are able to lock on to specific pathogen associated molecular patterns (PAMPs) or microbial associated molecular

pattern (MAMPs) which different invasive microorganisms such as bacteria and fungi secret. These PAMPs and MAMPs are recognized by the LRR region of PRRs which are mainly of two types. One is receptor like kinase which has kinase domain at the end and another one is receptor like protein which does not have any kinase domain (Zipfel, 2014). Receptor like kinase has four main regions which are leucine rich repeat (LRR), a single pass transmembrane domain (TM), one juxtamembrane domain (JM) and an intracellular kinase domain (Song *et al.*, 1995). The LRR binds with PAMPs/MAMPs (Figure 1.3) and through the TM and JM the signal is transferred to the inner side of the cell by kinase domain and PTI is activated. During this event a co-receptor protein is recruited which is required for the fill activation of PTI. (Stefanie, 2017)

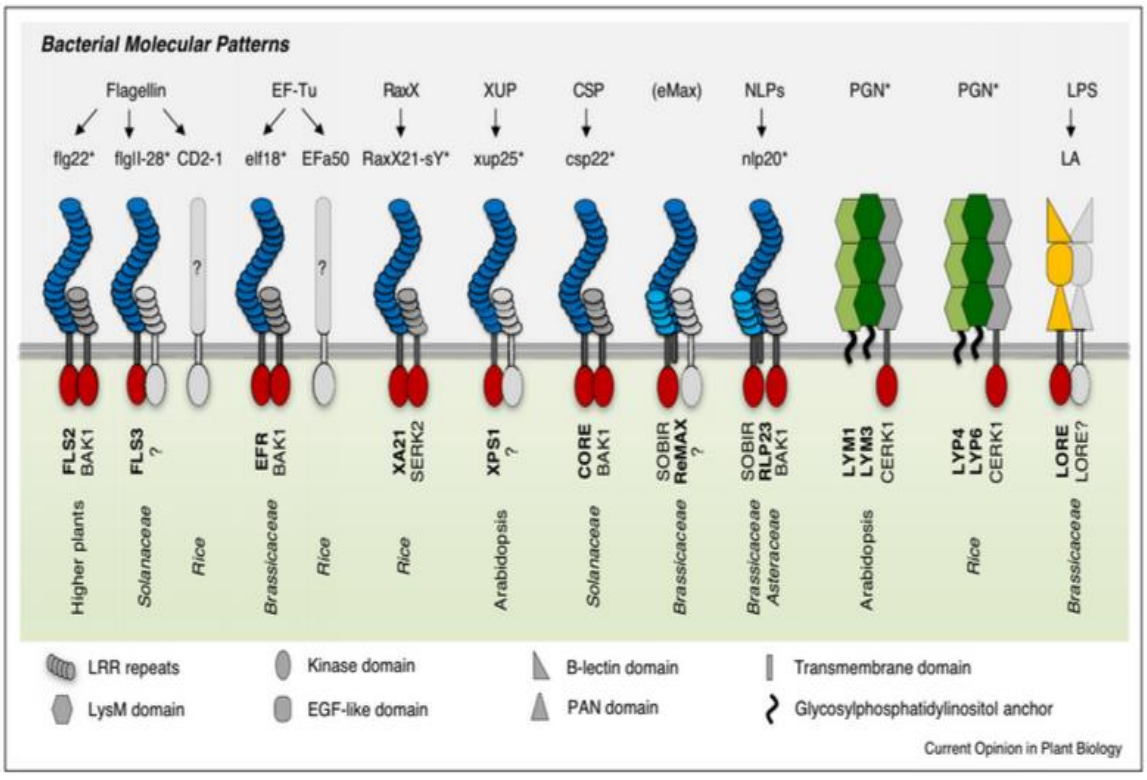


Fig 1.3: Different PRRs along with their recognized PAPMs in bacteria (Stefanie, 2017)

1.4.3 Effector triggered immunity (EIT)

Successful pathogens can avoid the pattern triggered immunity and secrets effectors by its type 3 secretion system (Thomma *et al.*, 2011). Plant can also avoid doing harm by effectors by its resistant proteins R. Most of these proteins are intracellular receptor proteins of the nucleotide binding leucine rich repeat (NB-LRR). Effector triggered immunity occur more quickly than pattern triggered immunity (Jones and Dangl, 2006; Tao *et al.*, 2003; Tsuda and Katagiri, 2010). These effectors were previously known as avirulence factors (Bent and Mackey, 2007; Chisholm *et al.*, 2006). This ETI is also known as gene-for-gene hypothesis where both gene product from plant and pathogen interacts with each other in receptor-ligand manner (Schürch *et al.*, 2004).

1.4.4 CORE mediated pattern triggered immunity

CORE is a cold shock protein receptor kinase which shows which recognizes the PAMP csp22. The responsiveness to csp22 actually helped subsequent identify the receptor kinase CORE as the PRR for csp22, as described by Lei *et al.* in 2016. CORE acts as a genuine receptor with high affinity and specificity for csp22, as identified by its heterologous expression in *A. thaliana* (Lei *et al.*, 2016).

The CORE protein contains 22 leucine rich repeat (LRR) domains with a six AA long island domain in the middle of the 11^{th} LRR domain. It is flanked by the N-terminal (Nt) LRR and C-terminal (Ct) LRR domains on either side. The Ct LRR is on the other side joined to the outer juxtamembrane (JM) domain, which along with the inner JM domain sandwich the transmembrane membrane (TM) domain, with the kinase domain on the other side of the inner JM domain (Lei *et al.*, 2016) (Fig 1.4). This structure was found to be very similar to the structure of EFR, the bacterial EF-Tu found in members of the Brassicaceae family such as *A. thaliana* (Zipfel *et al*., 2006) and Xa21, the bacterial receptor kinase found in the rice for RaxX21-sY (Song *et al.*, 1995; Pruitt *et al.*, 2015), respectively.

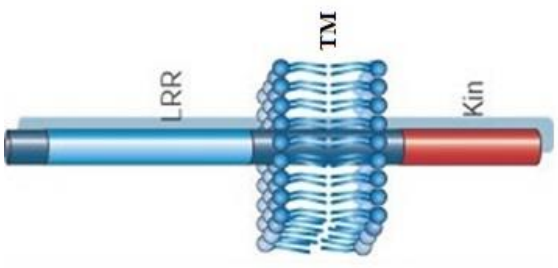


Fig 1.4: Positions of the different domains for the CORE protein (Dangl and Jones, 2001)

A comparative study revealed the similarities and differences the CORE protein AA sequence had with the EFR sequence (Fig 1.5), and it was seen that the LRR region shows about 64% sequence similarity and the kinase domain with about 50% similarity (Lei *et al.*, 2016). The remaining differences in the sequences are thus thought to be the cause for the specificity of the receptors to their respective MAMPs.

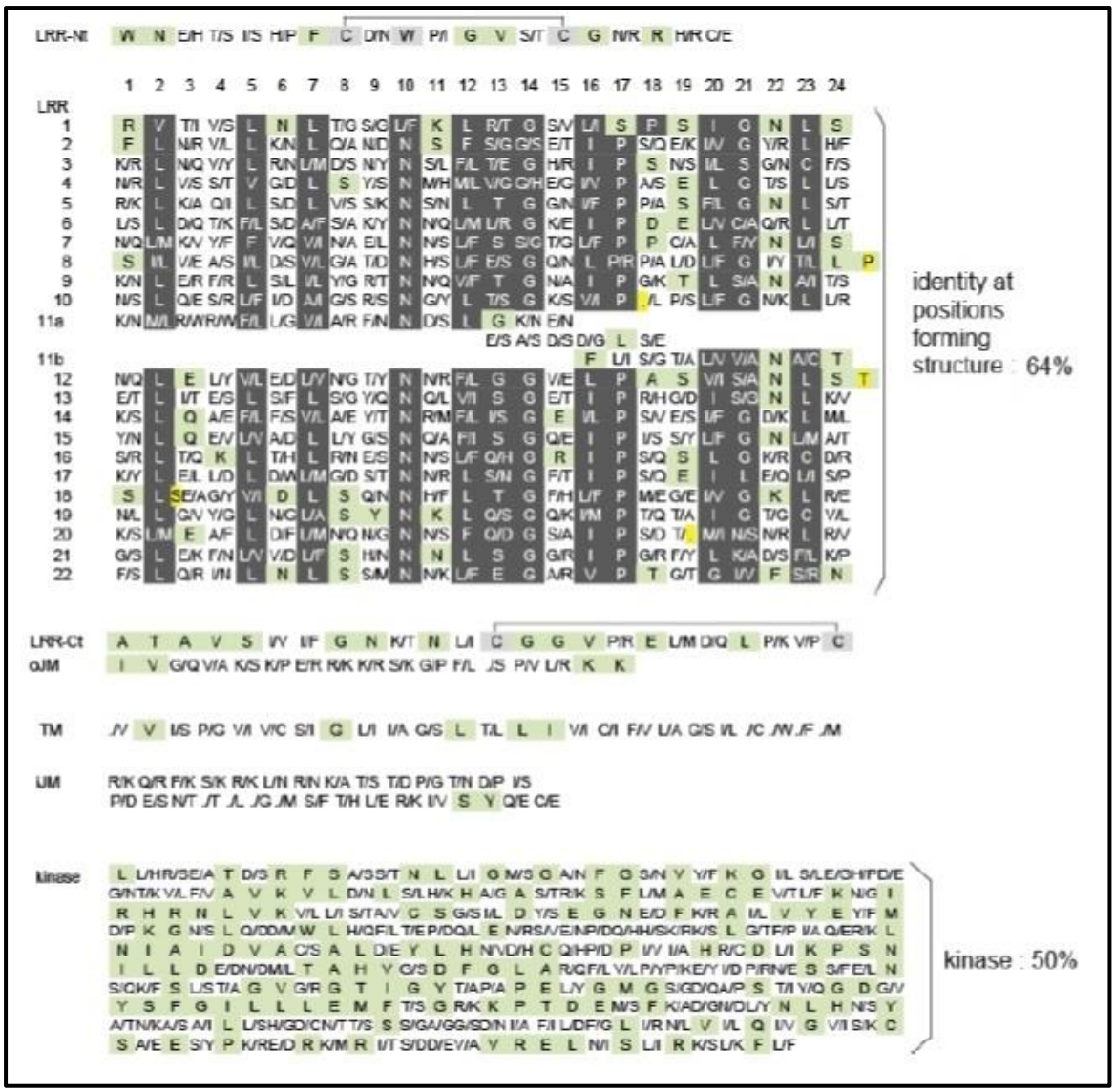


Fig 1.5: Comparison of the primary structures of the LRR receptor kinases encoded by CORE (Solyc03g096190) from tomato and EFR (At5g20480) from Arabidopsis. Single letters indicate positions with identical amino acids (AA, green underlay) while two letters separated by "/" indicate divergent AA residues (first letter denoting CORE AA and the later denoting EFR AA), respectively. Positions with deletions by "." or insertions of single AA in the repeats are highlighted in yellow (Lei *et al.*, 2016).

1.4.5 Csp22 as activator of CORE

The highly conserved nucleic acid binding motif RNP-1 of bacterial cold shock proteins (CSPs) was identified about fifteen years ago, but never found in plants outside the Solanaceae family (Felix *et al.*, 2003), for instance in Arabidopsis or rice. Bacterial secretions which were able to rapidly lower the incubation temperature by more the 10 °C (cold shock) were named CSPs (Bae *et al.*, 2000). As these proteins are naturally found in the bacterial cytoplasm and are membrane impermeable, it was surprising to see that plants contained receptors which were able to detect them but specificity of the activity of csp22 (the RNP-1 epitope with PAMP activity) strongly suggested the presence of a PRR protein located on plant surfaces able to perceive this PAMP (Lei *et al.*, 2016)

1.4.6 BAK1 Co-receptor regulated the CORE mediated immunity

Similar to EFR, which is a structural homologue to CORE, the binding of csp22 to CORE LRR also triggers the formation of a heterodimer with BAK1, acting as a co-receptor to facilitate the PTI mechanism (Chinchilla et al., 2009; Postma *et al.*, 2016). This is also consistent with the FLS2 mediated immunity, which also recruits BAK1 as a co-receptor in a ligand dependent manner (Sun *et al.*, 2013).

1.4.7 Other characterized pattern triggered immunity

The FLS2 complex was the first PRR to be fully characterized with its activity with flg22 studied in details as the crystalline complex was produced. On the attachment of the flg22 to the immunogenic epitope of FLS2, a co-receptor protein BAK1 also attaches to form a heterodimer essentially forming the activated complex responsible for medicating PTI (Fig 1.6) (Zipfel, 2014; Chinchilla *et al.*, 2006; Felix *et al.*, 1999; Sun *et al.*, 2013). The PRR FLS2 was first discovered in *Arabidopsis Thialiana* (Gómez-Gómez and Boller, 2000), following which it was also identified in tobacco, tomato, rice and grapevine (Hann and Rathjen, 2007; Robatzek *et al.*, 2007; Takai *et al.*, 2008; Trdá *et al.*, 2014).

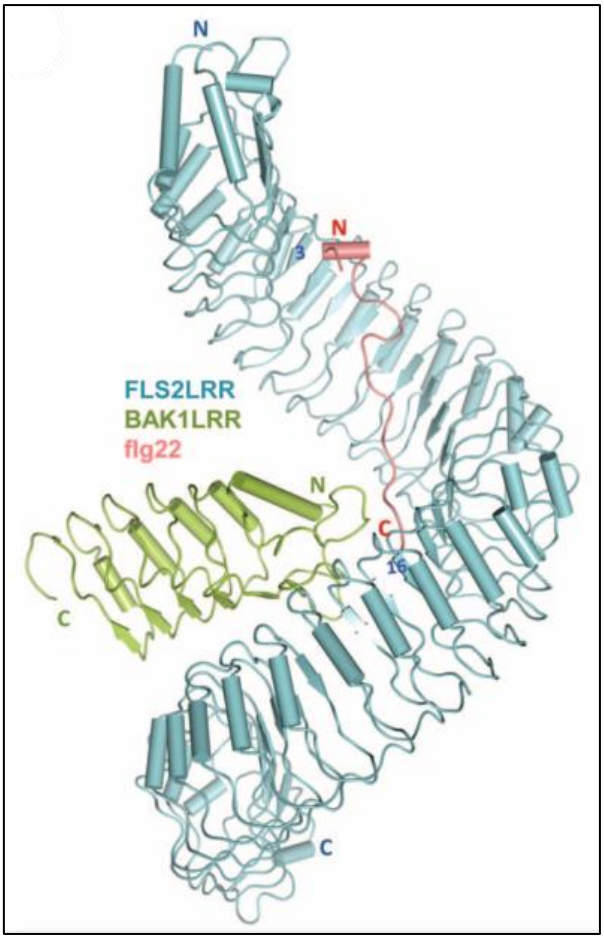


Fig 1.6: Cartoon structure showing the binding pattern of FLS2 (green) with its PAMP flf22 (pink) and co-receptor BAK1 (Sun *et al.*, 2013)

Flg22 binds to PRR FLS2 on the concave surface by crossing 14 LRR domains (LRR3 to LRR6) and the Ct of flg22 gets trapped in between the FLS2 LRR and BAK1 co-receptor. Interactions of flg22 with FLS2LRR can be divided into two parts separated by a kink (flg22 Asn10 and Ser11) in the central region of the peptide (Figure 2.11). Before the kink, the N-terminal seven residues bind to FLS2 LRR2 to LRR6 (FLS2LRR2-6) (Figure 1.7 A). Both hydrogen bonds and hydrophobic contacts mediate flg22 interaction with FLS2LRR. Flg22 Leu3 inserts into a hydrophobic pocket of FLS2 (Figure 1.7 B). In addition to hydrophobic contacts, FLS2 Arg152 and FLS2 Tyr148 also engage hydrogen bonds with flg22 Gln1 and flg22 Leu3, respectively (Sun *et al.*, 2013)

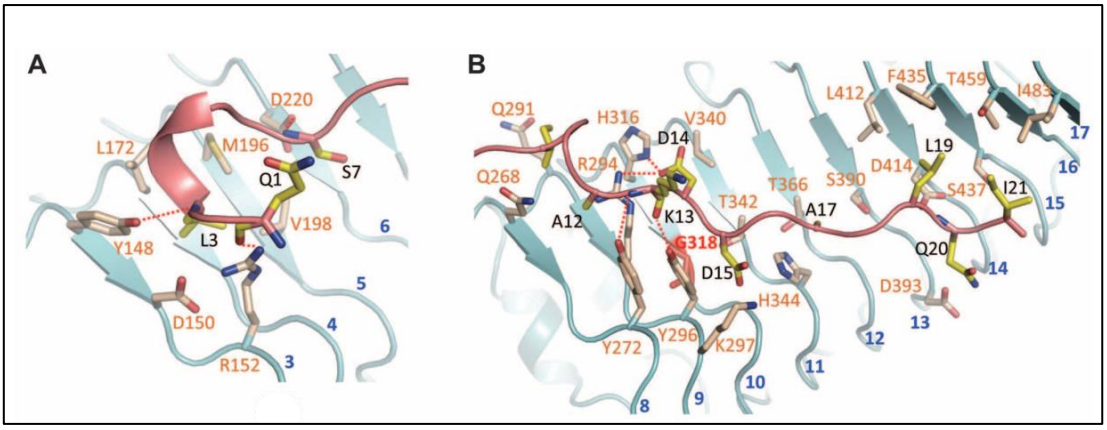


Fig 1.7: Molecular interaction of FLS2 (blue) and flg22 (pink) (Sun *et al.*, 2013)

Similar to FLS2, EFR which was previous seen to show remarkable similarities to the CORE protein, also interacts with the bacterial elongation factor Tu (EF-Tu). So the conserved Nt acetylated epitope, elf18, the first 18 AA of EF-Tu binds to EFR and mediates PTI (Kunze *et al.*, 2004; Zipfel *et al.*, 2006).

1.4.8 Computational approach for protein 3D structure prediction

X-ray crystallography, NMR-spectroscopy and dual polarization interferometry are definitively the tools of choice for producing protein structures. But due to their extremely high expenses, their utility is limited. As a result several computational methods for protein structure determination have been developed. There are chiefly two main classifications- single template modelling (STM) and another is multiple template modelling. As the names suggest, STMs and MTMs use only one or more than one template(s) or reference structure(s), respectively; to compare the AA sequence provided, and produce 3D structures of the proteins with respect to the AA sequence provided.

In our case both STM and MTM succeeded in producing the whole protein, but with varying degrees of accuracy. It was also observed that the MTM tools tended to perform better, as opposed to STM tools. In the course of our study we ended up using three different modelling tools for the modeling of the CORE ectodomain, and the protein models they produced, namely- Muster, HHpred, IntFOLD and I-TASSER.

Muster being the only STM among the four tools and it relies of threading/fold detection to predict the protein (Wu and Zhang, 2008). I-TASSER (Fig 1.8) and IntFOLD (Fig 1.9) followed suit and also applied the threading method to attain their models, but unlike Muster they both use multiple templates (Roy *et al.*, 2010; Daniel *et al.* 2011). I-TASSER even incorporates *ab initio* (Fig 1.10) besides threading, so allow an even more thorough model. HHpred on the other hand relied on homology modelling (Fig 1.11) for producing the model (Al-Lazikani *et al.*, 2001; García-Sánchez *et al.*, 2000).

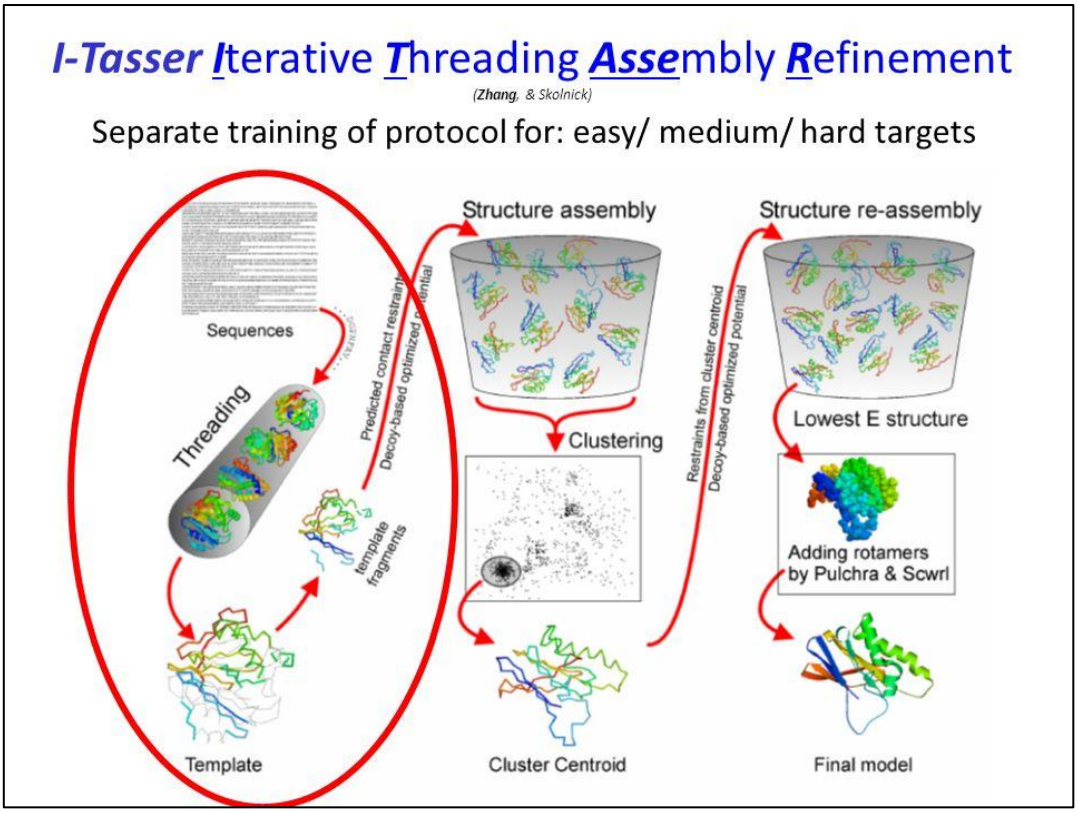


Fig 1.8: Steps of threading modelling by I-TASSER tool (Zhang Lab)

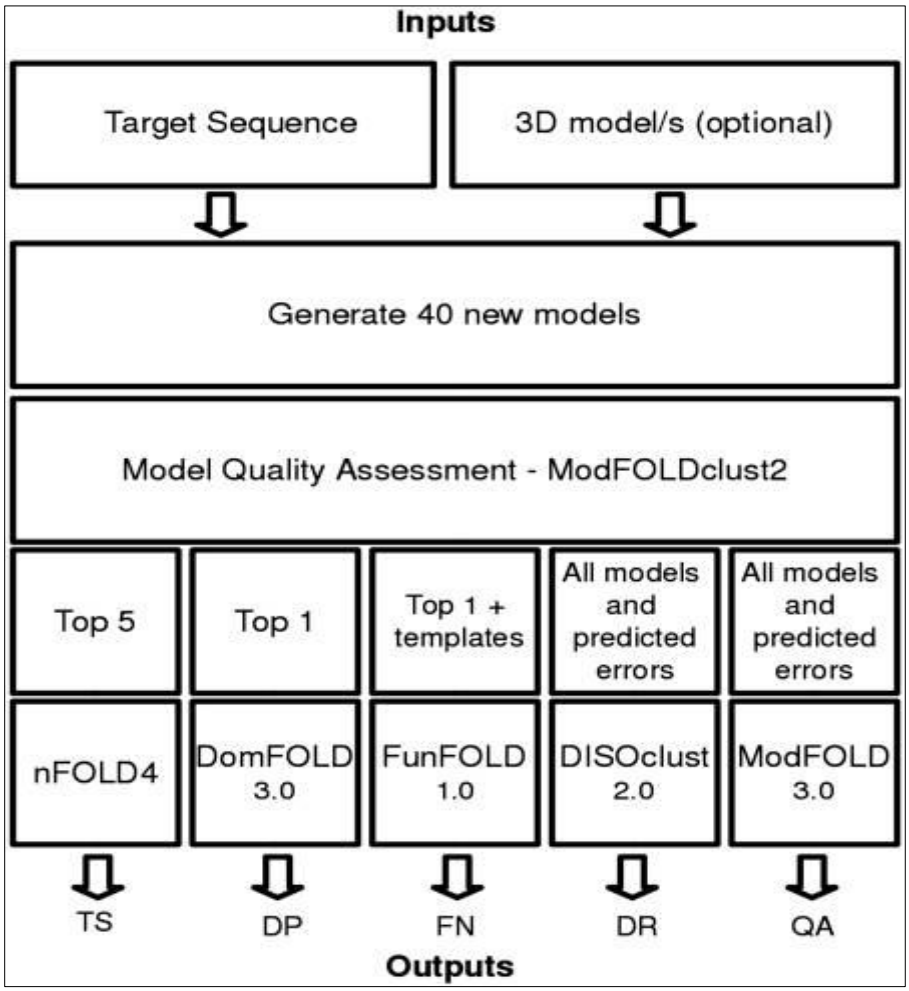


Fig 1.9: IntFOLD's mechanism used to model CORE LRR ectodomain (Daniel *et al.* 2011).

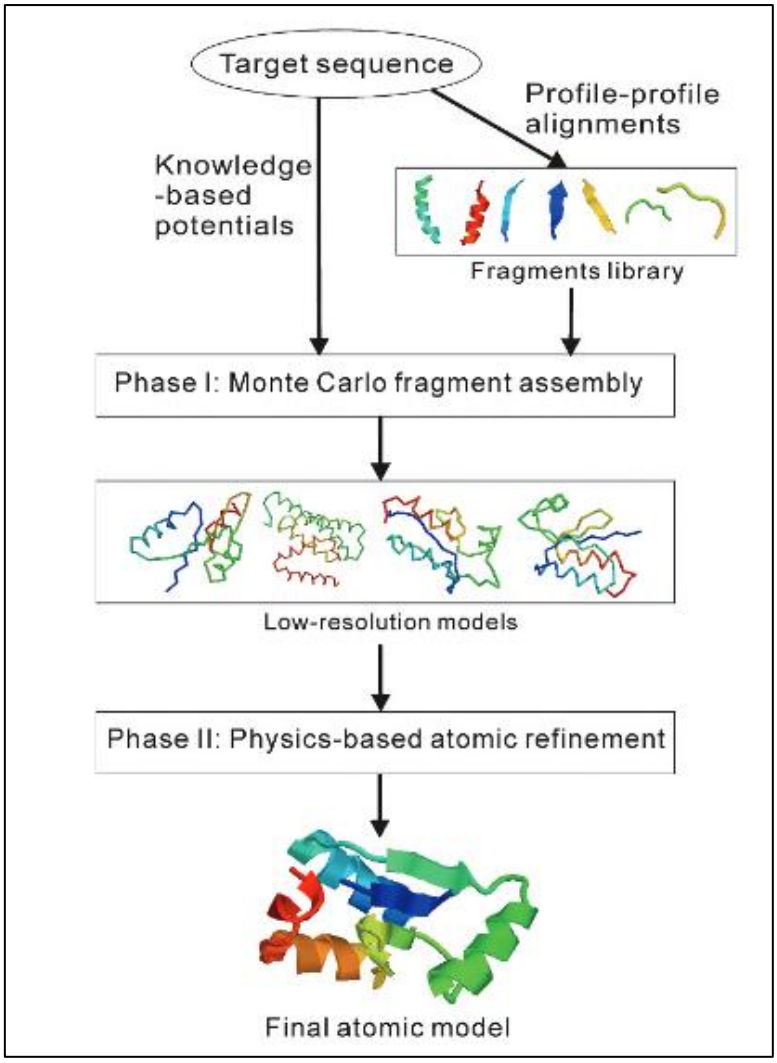


Fig 1.10: *Ab initio* modelling protocol of Rosetta tool (Rohl *et al.*, 2004)

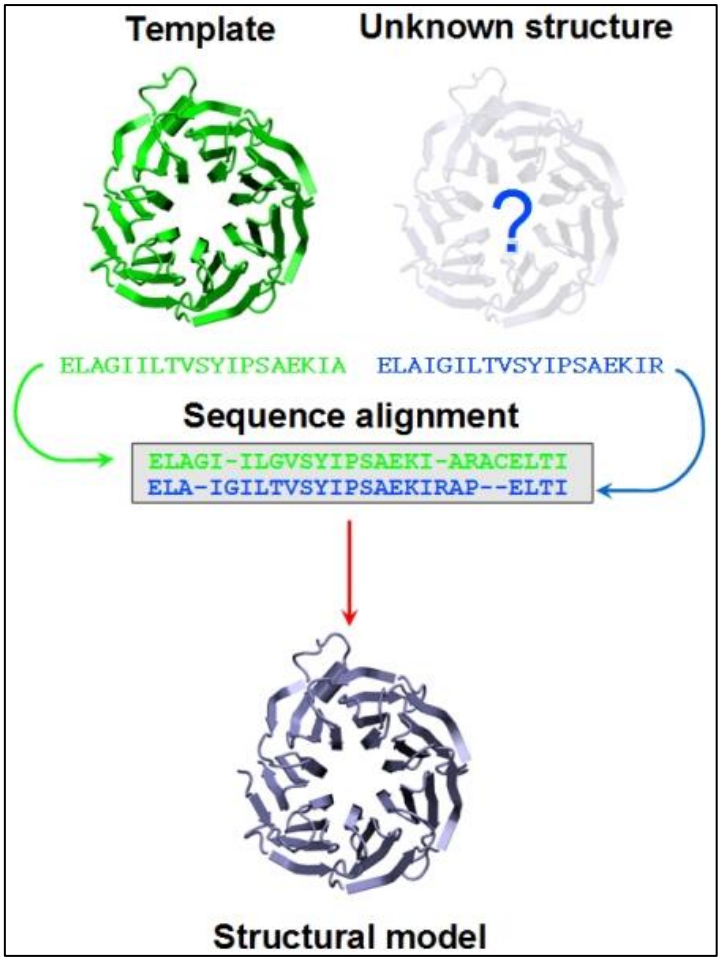


Fig 1.11: Principles of Homology modelling

1.4.9 Model validation

To test the structural integrity of the modelled proteins various bioinformatics tools were employed. They were- Verify3D, ERRAT, Ramachandran distribution plot generated by the RAMPAGE server and GROMACS software suite (Laskowski *et al.*, 1993)

The statistics of non-bonded interactions between different atom types ERRAT tool is used whereas Verify 3D analyzes the compatibility of an atomic 3D protein model with its own primary amino acid sequence. The angle of rotation of the residues was analyzed and they were positioned in an allowed or disallowed region, to see the precision of rotation of the structure. The GROMACS software suite was used to run molecular dynamics simulations to analyse the biomolecular system and conformations of a protein.

1.4.10 Protein-protein docking

Docking refers to the method of prediction of orientation of one molecule with another molecule to form a stable complex. Molecules can be proteins, nucleic acids, carbohydrates etc. The receiving molecule is known as receptor molecule most commonly a protein. The partner molecule is known as ligand which binds with the receptor molecule.

After looking through various docking tools in order to observe the interactions between the proteins, the online docking tool ClusPro was used (Kozakov *et al*, 2017; Kozakov *et a*, 2013; Kozakov *et al*, 2006; Comeau *et al*, 2004). As protein-protein interactions are very important in understanding the mechanism of action of the receptors, the data was keenly analyzed to reach conclusions.

1.4.11 Molecular dynamics simulation

Molecular dynamic (MD) simulation is becoming one of the most important and popular technique in the theoretical study of molecules since last decades. It is the computational method of research which connects the knowledge of macroscopic world with the microscopic study by the theory of statistical mechanics. MD simulation gives the result of detailed information on the fluctuation and conformational change of protein and also being used to determine the structure, dynamics and thermodynamics of biological molecules and their complexes (Allen, 2004). It is also important for the study of different biological processes in plants and animals by analysing the protein stability, conformational changes, protein folding and dynamic done by MD simulation. Commonly used MD simulation tools are GROMACS, AMBER, CHARMM and NAMD etc.

1.5 Scope and Limitations of the Study

To begin with we analyzed the primary and secondary structure of the LRR of PRR CORE and PAMP csp22 which helped us with the second stage of our study, which was modelling them using different tools and verifying the structural integrity of these models using other validation tools. In the next stage these models were subjected to a molecular dynamics system using the GROMACS software suite. Then the root mean square deviation (RMSD), root mean square fluctuation (RMSF) and radius of gyration (Rg) were measured and discussed. Following this, the PRR, PAMP and co-receptor CORE, csp22 and BAK1, respectively; where docked to form a three protein complex using the online tool ClusPro. The docked complex was subjected to MD simulation following the same protocol and the same analyses were conducted on it. As the LRRs of the PRRs are responsible for the binding and triggering of the PTI, our model of the LRR of CORE was sufficient to understand the interactions that take place in the complex.

There is no *in silico* study reference of MD simulation of PTI proteins. Thus, the reference structure was limited to FLS2 crystalline structure, FLS2 complex with its PAMP flg22 and crystal BAK1 having PDB ID 4MNA and 4MN8, respectively.

CHAPTER 2

MATERIALS AND METHODS

2.1 Introduction

This chapter consists of the methodology used for sequence based analysis of the CORE protein, followed by the steps for modelling of the ectodomain domain (the LRR region which is responsible for the interaction with the PAMP) of the CORE protein and also the PAMP csp22. Finally the procedures followed to run the molecular dynamics simulations and observe protein interactions via docking, are also described here.

2.2 Modelling of pattern recognition receptor CORE

Intensive modelling methodology was followed for the modelling of CORE ectodomain. Different modelling approaches such as single template modelling (STM) and multiple template modelling (MTM) were followed to model the CORE LRR ectodomain. It was seen that the four tools – HHpred toolkit, I-TASSER, IntFOLD and Muster; were able to model the CORE protein's LRR region with relative similarity.

Muster was the only tool which was able to generate a satisfactory model being a single template modelling tool using the threading method. Whereas the rest of the models were generated using multiple template modelling tools – HHpred used multiple template homology modelling, I-TASSER used threading and *ab intio* method, and finally IntFOLD used accuracy self-estimate (ASE) scores and refinement based on multiple template modelling.

2.2.1 Sequence based analysis and delineation of domain boundary

The Amino acid (AA) sequence of the target CORE protein was retrieved from Uniprot KB with the accession number K4BJ41/ Solyc03g096190 (Lei Wang *et al*, 2016). To have an initial idea about the physio-chemical properties and secondary structure CORE, primary structure was used to analyze for predicting physio-chemical properties using ProtParam tool (Gasteiger *et al.*, 2005) and the secondary structure was predicted using PSIPRED (Buchan *et al.*, 2010). To identify the conserved region of the sequence, ConSurf tool (Armon *et al.*, 2001) was used. To investigate the domain architecture, InterPro (Hunter *et al.*, 2009) was used. TMHMM (Emanuelsson *et al.*, 2007) was used for predicting the transmembrane region.

2.2.2 Single template modelling

NCBI BLASTp (Mahram and Herbordt, 2010) analysis of CORE protein AA sequence was carried out against Protein Data Bank (PDB) using default parameter values to search for the suitable template for CORE single template modelling. NCBI BLASTp suggested 4mn8_A (chain A of crystal structure FLS2-Bak1-flg22 complex) as the best template for modelling of CORE. This is confirmed by the template covering 100% of the protein with 36% identity. Then different single template modelling approach was carried out (Table 2.1) using different single template modelling approaches were carried out using SWISS-MODEL (Schwede *et al.*, 2003), RaptorX (Källberg *et al*, 2012), Spark-X (Huang *et al.*, 2014), Muster (Wu and Zhang, 2008) and PSPS (Chen *et al*, 2006).

Table 2.1: Single template approach for modelling of CORE LRR ectodomain

Tool	**Modelling Method**	**Template**
SWISS-MODEL	Homology	4mn8A
RaptorX	Threading	4mn8A
FFAS-3D	Threading	4mn8A
FFAS03	Threading	4mn8A
Sparks-X	Threading	4mn8A
Muster	Threading	4mn8A
PSPS	Homology	4mn8A

2.2.3 Multiple template modelling

Different multiple template modelling approaches were also carried out for the modelling of CORE receptor protein. Phyre2 intensive modelling (Kelley *et al.*, 2015), I-TASSER (Roy *et al.*, 2010), HHpred toolkit (Söding *et al*, 2005), AIDA (Xu *et al.*, 2014) and IntFOLD (Daniel *et al*, 2011), where used. Phyre2 is based on *ab initio*, I-TASSER is based on threading and *ad initio* and AIDA is homology based multi-template modelling server.

Table 2.2: Multiple template approach for modelling of CORE LRR ectodomain

R/P ID	**Tool**	**Modelling Method**	**Template/s**
CORE_Hhpred11	HHpred	Homology	4mnA_A+5hyx_B+5gr9_B
CORE_AIDA	AIDA	Homology	4mn8+4mnA
CORE_I-TASSER	I-TASSER	Ab initio and Threading	5gijB+4mn8A+5hyxB
CORE_Phyre2	Phyre2 Intensive	Ab initio	4mnaA+4mn8A+5gijB
CORE_IntFOLD	IntFOLD	Threading	4mn8+4mnA

2.2.4 Structural validation

To evaluate the structural and geometrical consistency and reliability of the modelled proteins, several approaches were adopted. ERRAT (Wallner and Elofsson, 2003) was used to study the non-bonded interactions between different atom types while, Verify 3D (Liithy *et al.*, 1992) was subjected to assess the compatibility of the atomic models with its own AA sequence. To study the geometrical consistency of the modelled proteins, Ramachandran plot generated from RAMPAGE (Laskowski *et al.*, 1993) were assessed. The protein quality was also visually analysed by PyMOL tool (DeLano, 2002). Detailed analysis of the complex was done to identify the structural details of the CORE LRR ectodomain.

2.2.5 Molecular dynamics simulation of CORE ectodomain and csp22

To refine and obtain the stable structure of CORE protein, protein modelled by HHpred toolkit, I-TASSER, Muster and IntFOLD tools were subjected to molecular dynamics (MD) simulation with GROMACS (Van Der Spoel *et al.,* 2005) software suite. The OPLS united force field was used to run the simulations. Before running the simulation, the systems were solvated, neutralized, energy minimized and equilibrated. In case of solvation, the proteins were taken into a cubic box with a minimum distance 1Å between the protein surfaces and edges. Then the boxes with these proteins inside were solvated with SPC water model (van der Spoel *et al.*, 1998). The systems were neutralized with genion tool of GROMACS before energy minimization. Then the systems were equilibrated for 1 ns NPT ensemble followed by 1 ns NVT ensemble maintaining a constant 1 atm pressure and 300 K temperature, respectively. Finally a 20 ns MD simulation was carried out for each system. The same procedures were also followed for MD simulation of PAMP csp22 except csp22 simulation run was set for a 100 ns period. To treat the long range electrostatic interactions, particle mesh Ewald (PME) method was applied. Root mean square deviation (RMSD), radius of gyration (Rg), energy and root mean square fluctuations (RMSFs) were calculated using GROMACS tools to monitor conformational changes over the simulation time.

2.3 Docking of CORE, its PAMP csp22 and co-receptor BAK1

The best LRR structure (model produced by I-TASSER) was energy minimized and equilibrated. After 2 ns NPT equilibration followed by 1 ns NVT equilibration the energy minimized and equilibrated structure was given as initial protein structure. Same procedure was followed in case of csp22 protein. But in case of co-receptor BAK1 X-ray crystallographic structure (PDB ID: 4MN8) was obtained. Then CORE LRR, csp22 and BAK1 structures were subjected to docking with the protein-protein docking tool ClusPro (Kozakov *et al*, 2017; Kozakov *et a*, 2013; Kozakov *et al*, 2006; Comeau *et al*, 2004). For multiple protein docking first CORE was docked with the csp22 and then best docked structure complex was used for further docking with BAK1.

From docking result, best predictions were selected and interactions were analysed using PyMOL (DeLano, 2002).

2.4 Molecular dynamics simulation of docked complexes

After docking, each complex was subjected to run molecular dynamic (MD) simulation with GROMACS (Van Der Spoel *et al.*, 2005) software suite. The OPLS united force field was used to run the simulations. Before running the simulation, the systems were solvated, neutralized, energy minimized and equilibrated. In case of solvation, the proteins were taken into a cubic box with a minimum distance 1Å between the protein surfaces and edges. Then the boxes with these protein complexes inside were solvated with SPC water model (van der Spoel *et al.*, 1998). The systems were neutralized with genion tool of GROMACS before energy minimization. Then the systems were equilibrated for 1 ns NPT ensemble followed by 1 ns NVT ensemble maintaining a constant 1atm pressure and 300 K temperature, respectively. Finally a 20 ns MD simulation was carried out for each systems and root mean square deviation (RMSD), root mean square fluctuation (RMSF) and radius of gyration (Rg) was done. Also after 20 ns of simulation the complexes were subjected to further analysis by PyMOL tool.

2.5 Comparative study between CORE ectodomain complexes with FLS2 complex

To compare the binding mechanism between CORE docked complexes and FLS2-flg22-BAK1 crystal complex, the crystal structure of FLS2-flg22-Bak1 (PDB ID: 4mn8A) was obtained from protein data base. Then, PyMOL tool was used for comparative study of binding conformation between CORE complexes and FLS2 PRR complex.

2.6 Summary

This chapter illustrates the detailed methodology of protein modelling used in this study. Also model validation, docking protocol, MD simulation protocol is also described at the end of the chapter.

CHAPTER 3

RESULTS AND DISCUSSIONS

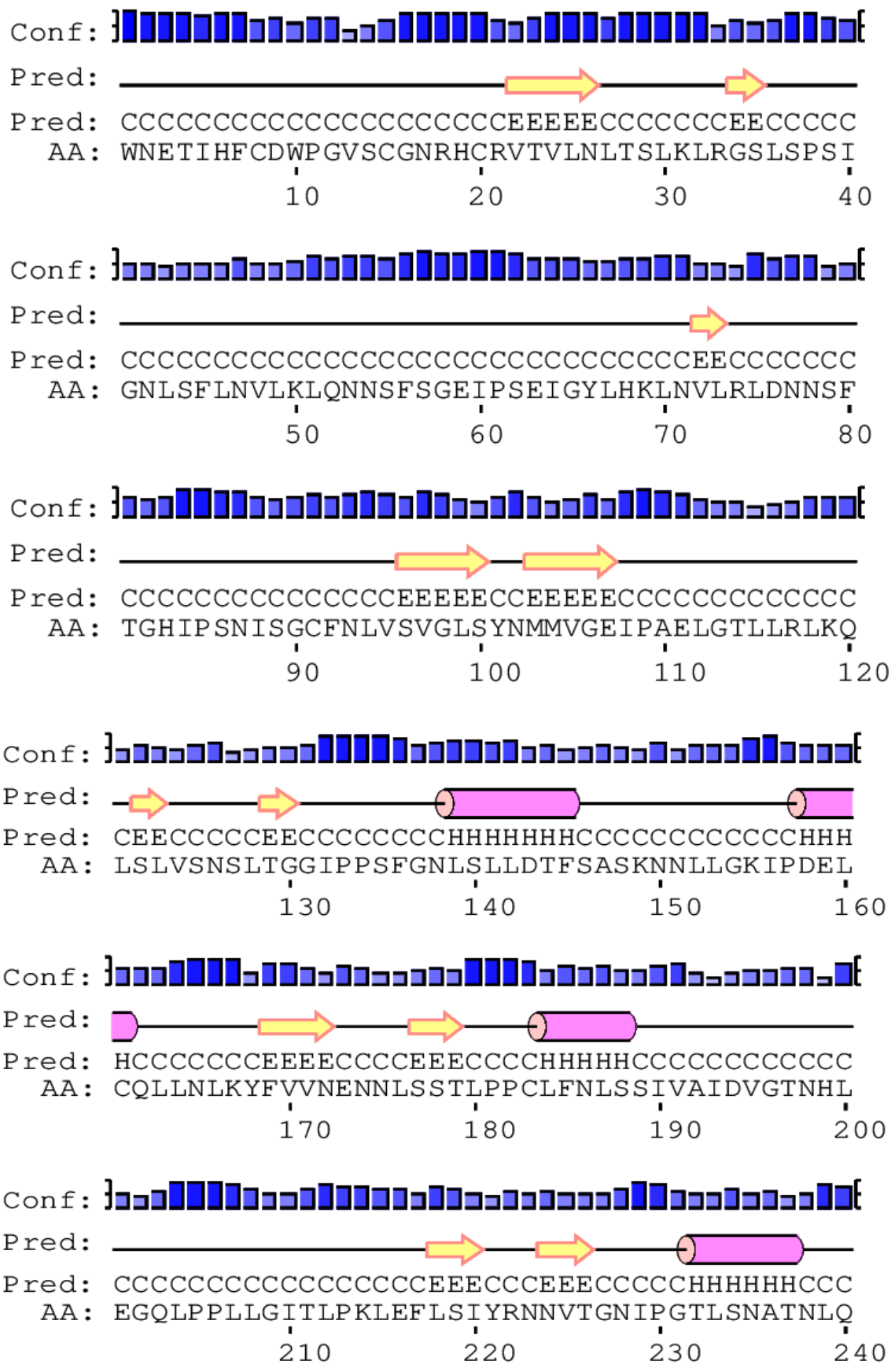
3.1 Introduction

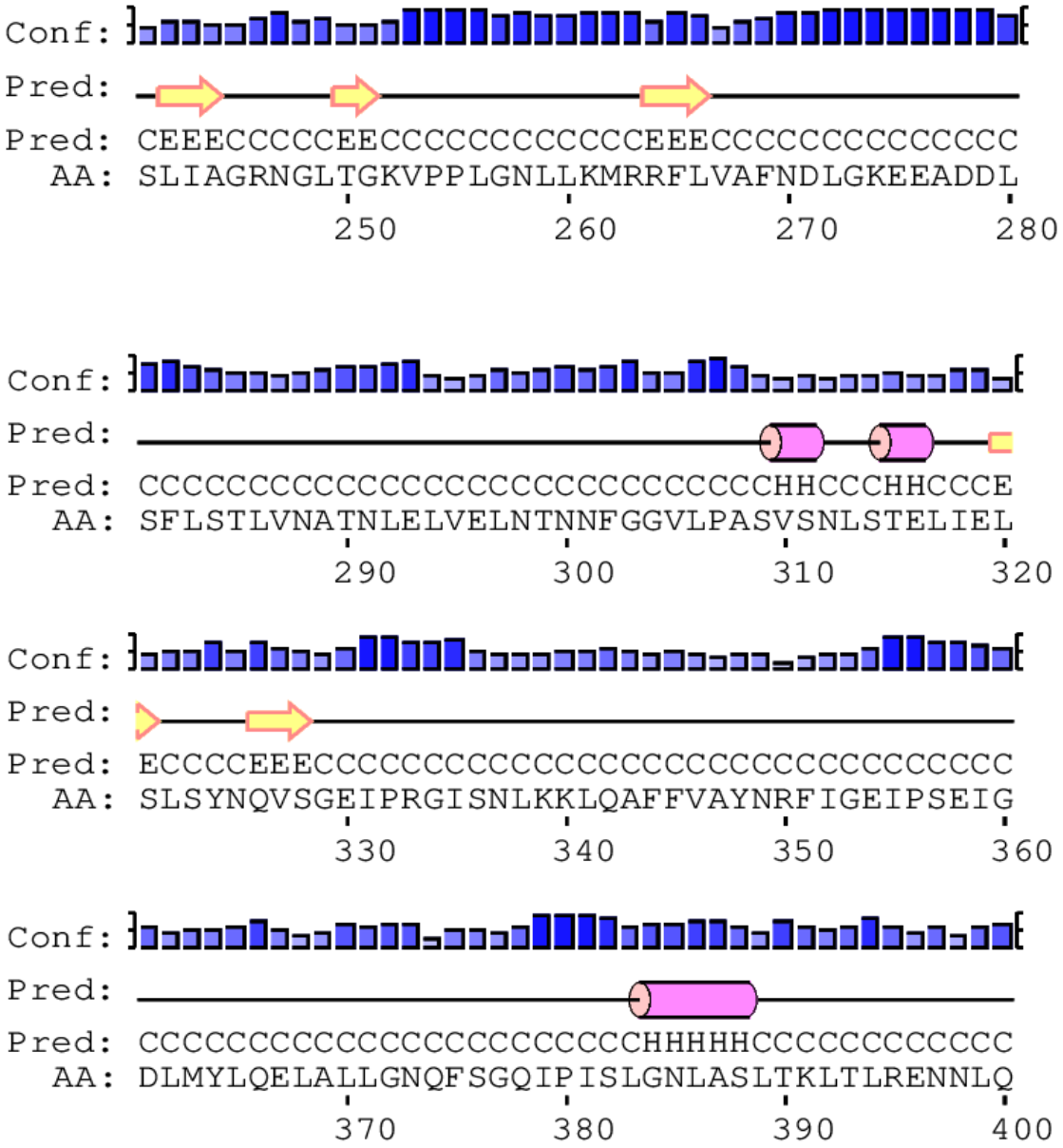
This chapter presents first the sequence based analysis results and modeling results. Then the interactions between the plant PRR CORE with the PAMP csp22 and co-receptor BAK1 are illustrated and analyzed. Finally the interactions between the proteins in our three protein complex are compared to FLS2 mediated immunity, at the end of this chapter.

3.2 Sequence based analysis and delineation of domain boundaries

At first the physico-chemical properties of the CORE protein ectodomain was analyzed by using ProtParam, which revealed that the CORE LRR domain consisted of 580 AA and has a molecular weight of 62.5 kDa. The isoelectric point (pI) was seen to be as 5.75 consistent with the slightly acidic property of the protein and the aliphatic index was found to be 113.59 which indicated the stability of CORE in a wide range of temperatures. The instability index was seen to be below 40 at about 29.33 proving the protein to be quite stable. Finally a GRAVY value of 0.129 told us that the protein was polar in nature.

The secondary structure predicted using PSIPRED (Fig 3.1) showed that the CORE protein ectodomain is mainly composed of alpha helix, beta sheets (strands) and coils.


Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCEEEEECCCCCCCCEECCCCC
AA: WNETIHFCDWPGVSCGNRHCRVTVLNLTSLKLRGSLSPSI
10
20
30
40
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEECCCCCCC
AA: GNLSFLNVLKLQNNSFSGEIPSEIGYLHKLNVLRLDNNSF
50
60
70
80
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCEEEEECCEEEEECCCCCCCCCCCCCC
AA: TGHIPSNISGCFNLVSVGLSYNMMVGEIPAELGTLLRLKQ
90
100
110
120
Conf:
Pred:
Pred: CEECCCCCEECCCCCCCCCHHHHHHHCCCCCCCCCCCCHHH
AA: LSLVSNSLTGGIPPSFGNLSLLDTFSASKNNLLGKIPDEL
130
140
150
160
Conf:
Pred:
Pred: HCCCCCCCEEEECCCCEECCCCHHHHHCCCCCCCCCCCC
AA: CQLLNLKYFVVNENNLSSTLPPCLFNLSSIVAIDVGTNHL
170
180
190
200
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCEEECCCEECCCCCHHHHHHCCC
AA: EGQLPPLLGITLPKLEFLSIYRNNVTGNIPGTLSNATNLQ
210
220
230
240


Conf:
Pred:
Pred: CEEECCCCCEECCCCCCCCCCCCEEECCCCCCCCCCCCCC
AA: SLIAGRNGLTGKVPPLGNLLKMRRFLVAFNDLGKEEADDL
250
260
270
280
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHCCCHHCCCE
AA: SFLSTLVNATNLELVELNTNNFGGVLPASVSNLSTELIEL
290
300
310
320
Conf:
Pred:
Pred: ECCCCEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: SLSYNQVSGEIPRGISNLKKLQAFFVAYNRFIGEIPSEIG
330
340
350
360
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCHHHHHCCCCCCCCCCC
AA: DLMYLQELALLGNQFSGQIPISLGNLASLTKLTLRENNLQ
370
380
390
400

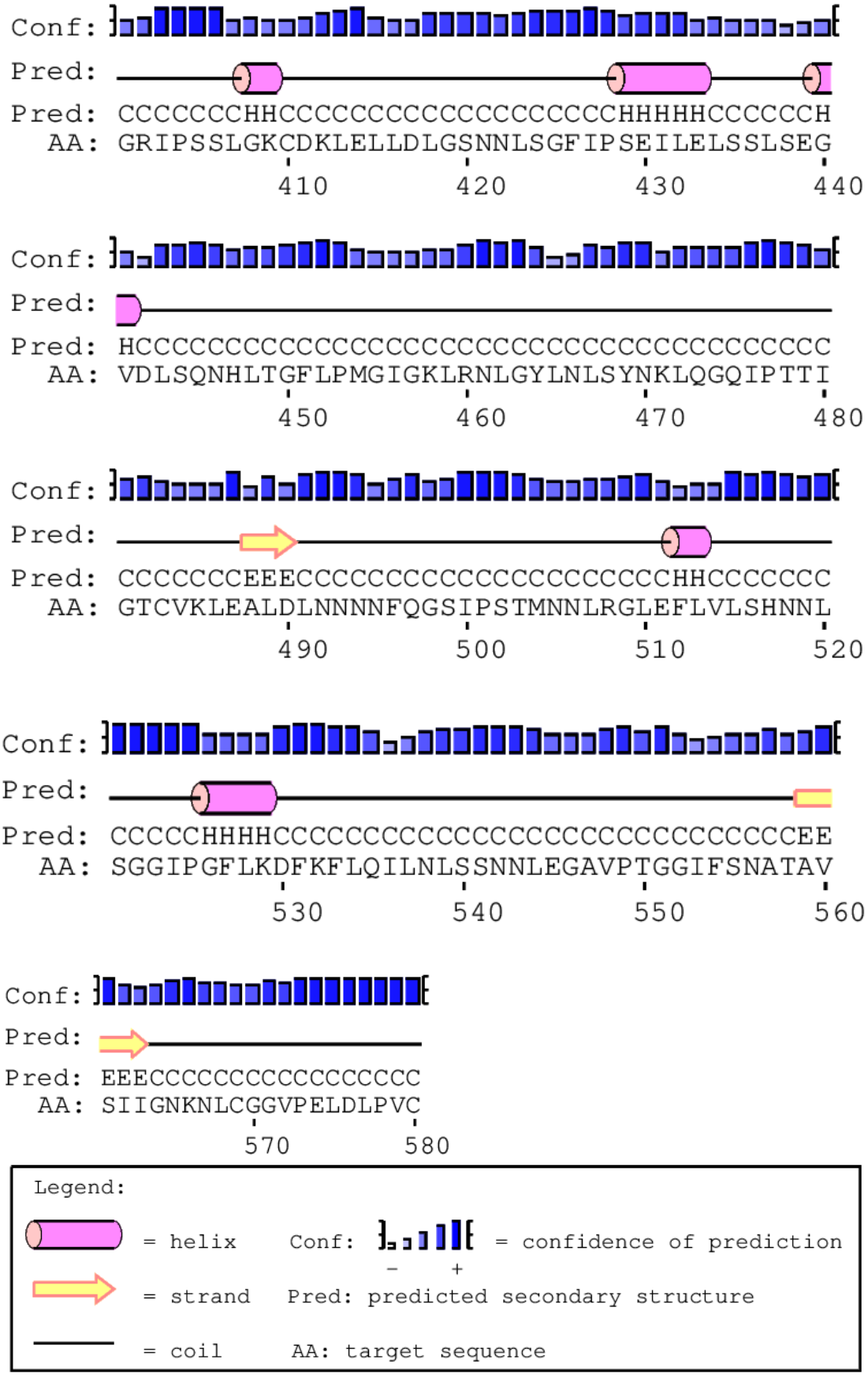


Fig 3.1: Secondary structure predicted by PSIPRED

Then InterPro was used to analyse the domain architecture of the LLR ectodomain of the CORE protein (Fig 3.2). It was seen that there were three leucine-rich domain superfamilies, six leucine-rich repeat typical sub-types and three leucine-rich repeat domains on the CORE ectodomain.

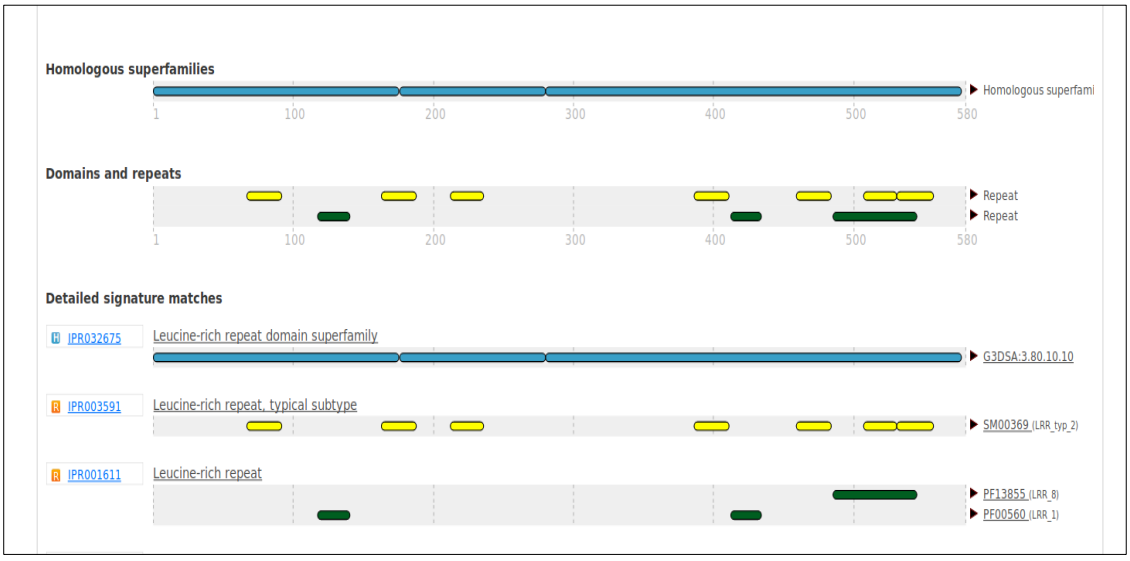
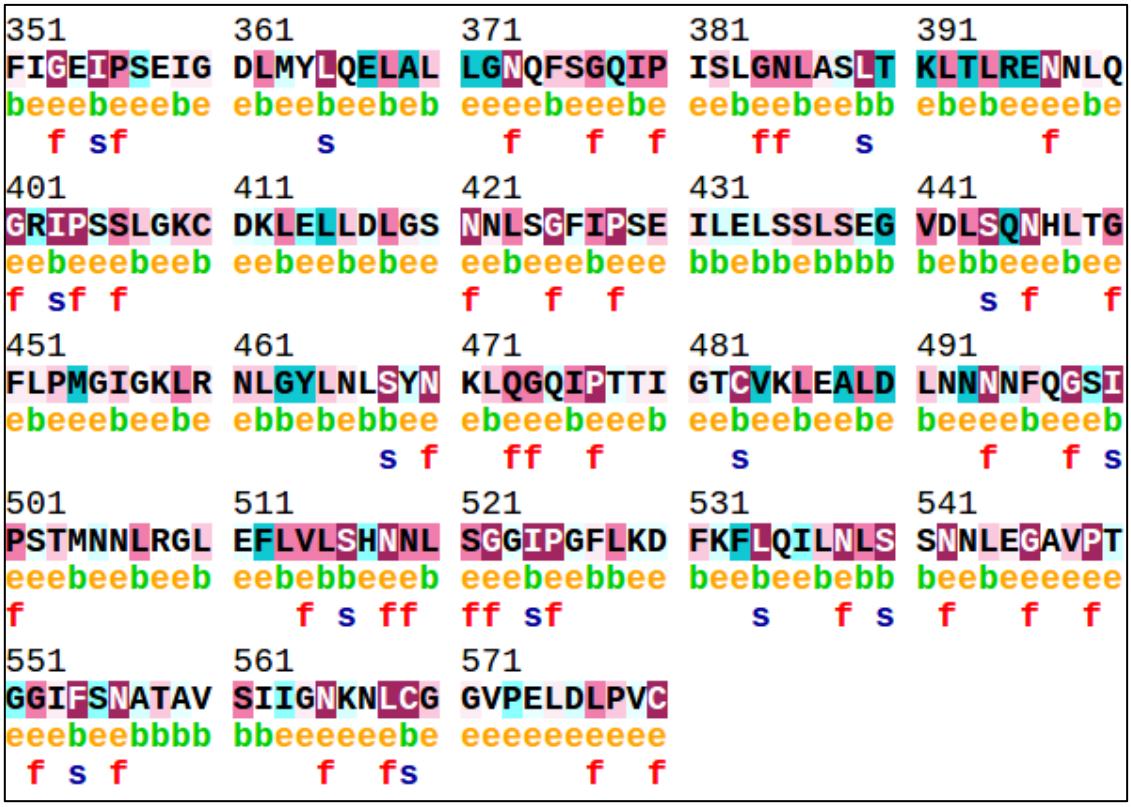


Fig 3.2: Domain architecture analysis by InterPro

Then the ConSurf tool was used to predict the conserved region of the CORE protein's LRR region. The results show that most of the amino acids are exposed residues according to the neural-network algorithm with an even distribution of buried residues, predicted functional residues and predicted structural residues (Fig 3.4).


1 11 21 31 41
WNETIHFCDW PGVSCGNRHC RVTVLNLTSL KLRGSLSPSI GNLSFLNVLK
51 61 71 81 91
LQNNSFSGEI PSEIGYLHKL NVLRLDNNSF TGHIPSNISG CFNLVSVGLS
101 111 121 131 141
YNMMVGEIPA ELGTLLRLKQ LSLVSNSLTG GIPPSFGNLS LLDTFSASKN
151 161 171 181 191
NLLGKIPDEL CQLLNLKYFV VNENNLSSTL PPCLFNLSSI VAIDVGTNHL
201 211 221 231 241
EGQLPPLLGI TLPKLEFLSI YRNNVTGNIP GTLSNATNLQ SLIAGRNGLT
251 261 271 281 291
GKVPPLGNLL KMRRFLVAFN DLGKEEADDL SFLSTLVNAT NLELVELNTN
301 311 321 331 341
NFGGVLPASV SNLSTELIEL SLSYNQVSGE IPRGISNLKK LQAFFVAYNR


351 361 371 381 391
FIGEIPSEIG DLMYLQELAL LGNQFSGQIP ISLGNLASLT KLTLRENNLQ
401 411 421 431 441
GRIPSSLGKC DKLELLDLGS NNLSGFIPSE ILELSSLSEG VDLSQNHLTG
451 461 471 481 491
FLPMGIGKLR NLGYLNLSYN KLQGQIPTTI GTCVKLEALD LNNNNFQGSI
501 511 521 531 541
PSTMNNLRGL EFLVLSHNNL SGGIPGFLKD FKFLQILNLS SNNLEGAVPT
551 561 571
GGIFSNATAV SIIGNKNLCG GVPELDLPVC

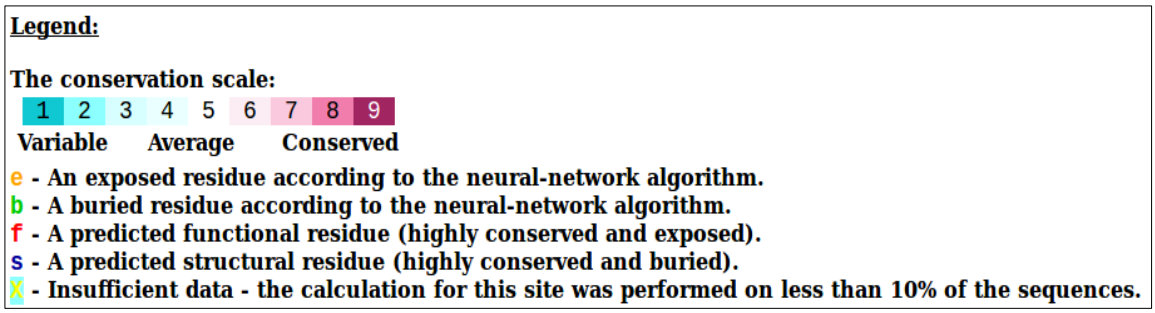


Fig 3.3: Prediction of conserved regions using the ConSurf tool.

3.3 Single template modelling

On conducting a Protein Blast (BLASTp) on the NCBI server many suitable templates complimentary to the CORE LRR sequence turned up. Among which, the most closely linked one proved to be of 4MNA_A (the free ectodomain of the FLS2 crystalline complex) with the highest score of 823, an e-value of 3e-101 which was the lowest and a sequence identity of 36%. The HHpred server also showed similar results with 100% probability of a match with the 4MNA_A template.

Multiple single template modelling tools were employed in attempts of modelling the CORE protein's ectodomain. All the tools were able to construct the LRR region of the protein with varying success. PSPS, Muster and Swiss Model tried to predict the protein structure using homology modelling; whereas RaptorX, Spark X, FFAS03 and FFAS-3D modelled the protein using local meta-threading server (LOMETS) produced multiple structures, of which the best models were chosen for further validation.

3.4 Multiple template modelling

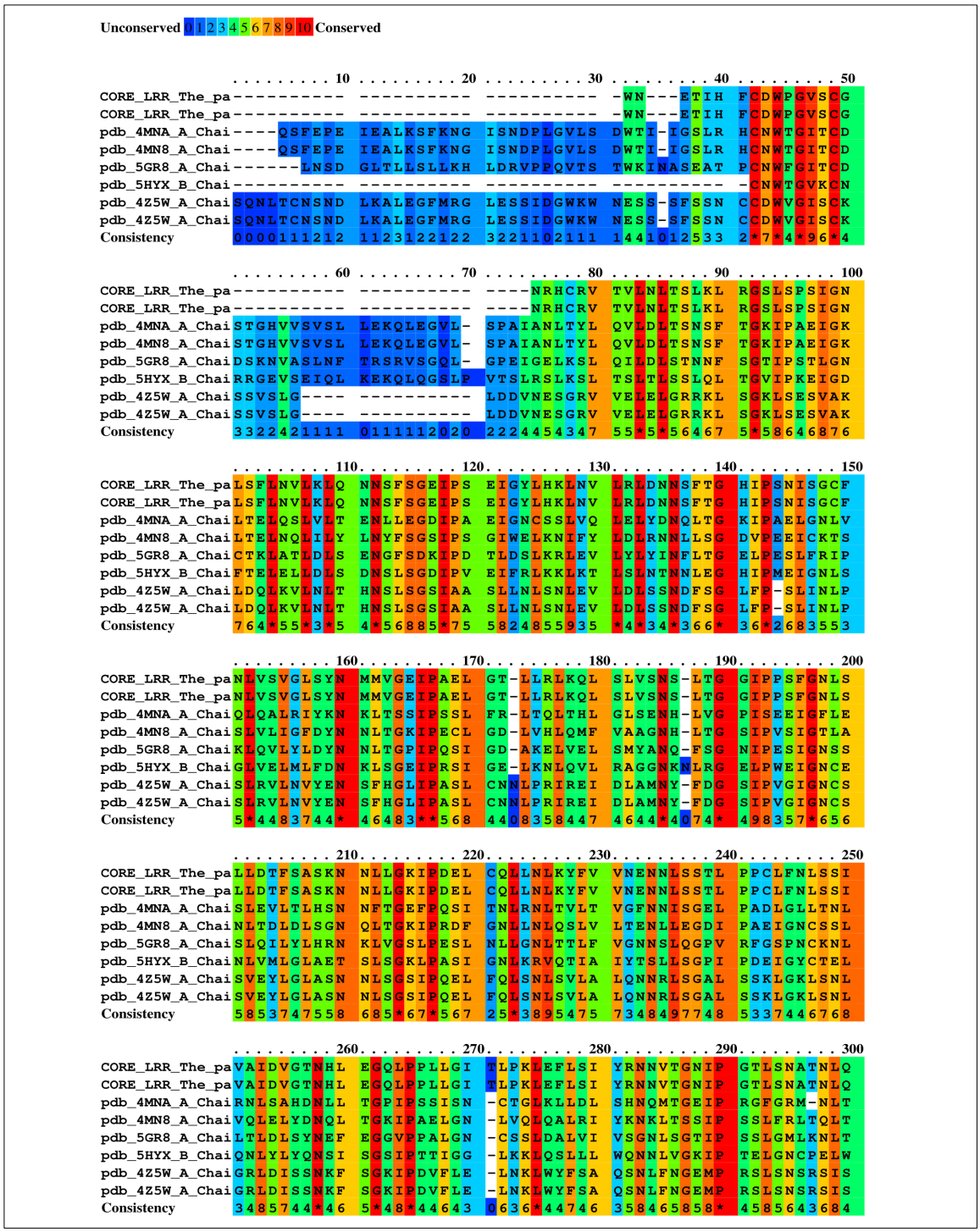
I-TASSER selected top ten threading templates according to the highest Z-score of each threading alignment of LOMENTS (Roy *et al*, 2011) from thousands of threading alignments. Finally according to the lowest C-scores five models were generated and the model with the lowest C-score was chosen for further validation.

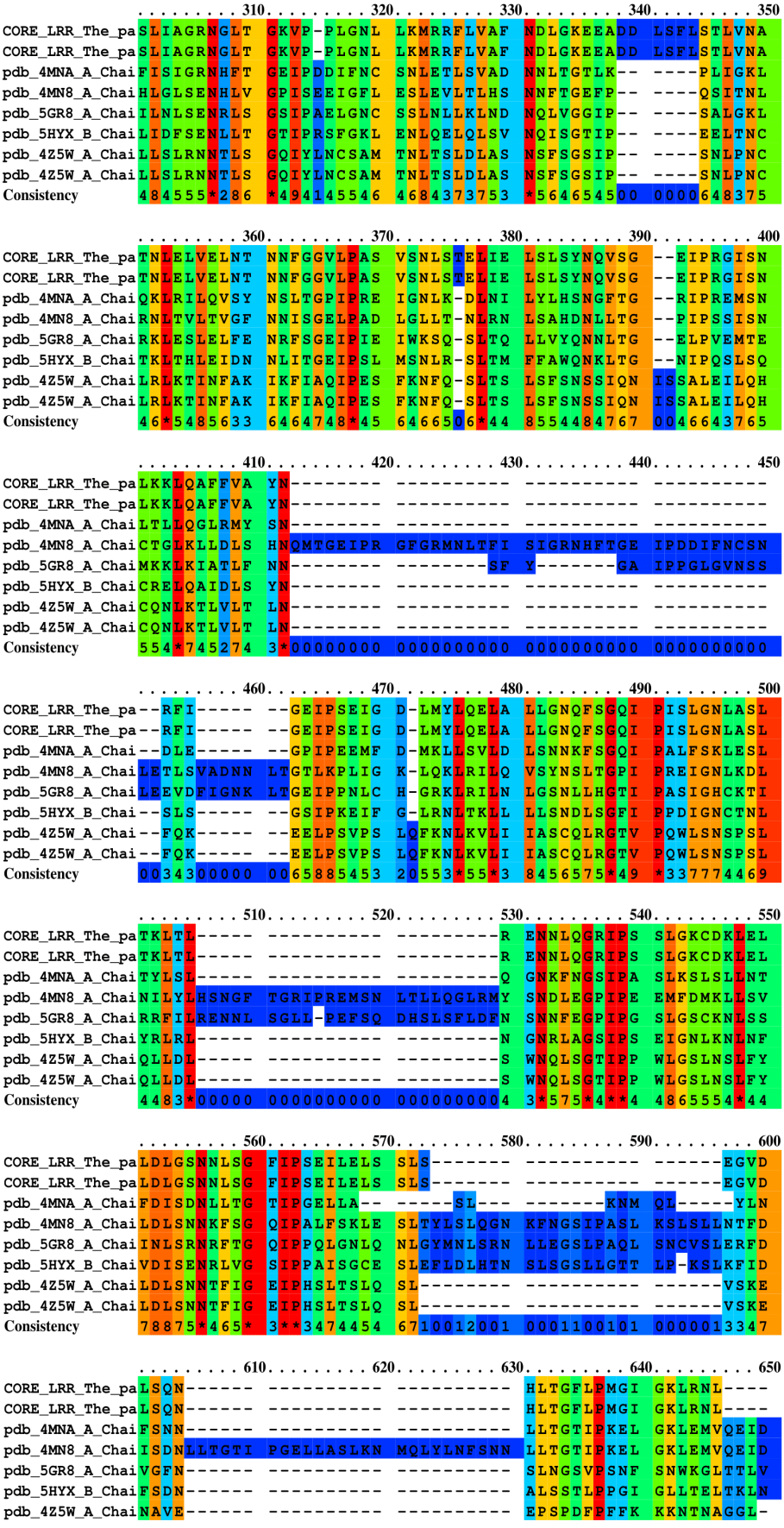
Multiple sequence alignment (MSA) using the Praline tool (Simossis and Heringa, 2005) using the top five templates (Fig 3.5), according to lowest e-value, from NCBI BLASTp was carried out to see sequences similarity it had with the query sequence (CORE LRR 580 AA sequence) (Table 3.1). Then in the HHpred server the top five templates were selected, as it is the only tool which allows the templates to be selected manually, and for modelling in all possible combinations. Only the top five templates were used as it was stated in the HHpred literature that no more than five templates should be used to acquire the best possible results. On trying all possible combinations, we finally got 30 models and all were constructed using the HHpred toolkit (Table 3.2)

Table 3.1: Protein Blast results obtained from the NCBI database

PDB ID	**Max. Score**	**E-value**	**Q C (%)**	**Idn. (%)**	**Template Short Identity**
4MNA_A	321	3e-101	98	36	Chain A, Crystal Structure of the Free FLS2 Ectodomain
4MN8_A	287	4e-87	97	35	Chain A, Crystal Structure of flg22 in complex with the FLS2 and BAK1 ectodomains
5GR8_A	268	1e-80	97	34	Chain A, Crystal Structure of Pepr1-atpep1
5HYX_B	247	2e-73	97	34	Chain B, Plant Peptide Hormone Receptor Rgfr1 in Complex with Rgf1
4Z5W_A	218	2e-62	98	34	Chain A, The Plant Peptide Hormone Receptor

Max. Score, Maximum Score; Q C, Query Coverage; Idn., Identity.


Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved
CORE_LRR_The_pa
pdb_4MNA_A_Chai
pdb_4MN8_A_Chai
pdb_5GR8_A_Chai
pdb_5HYX_B_Chai
pdb_4Z5W_A_Chai
Consistency


CORE_LRR_The_pa
CORE_LRR_The_pa
pdb_4MNA_A_Chai
pdb_4MN8_A_Chai
pdb_5GR8_A_Chai
pdb_5HYX_B_Chai
pdb_4Z5W_A_Chai
pdb_4Z5W_A_Chai
Consistency

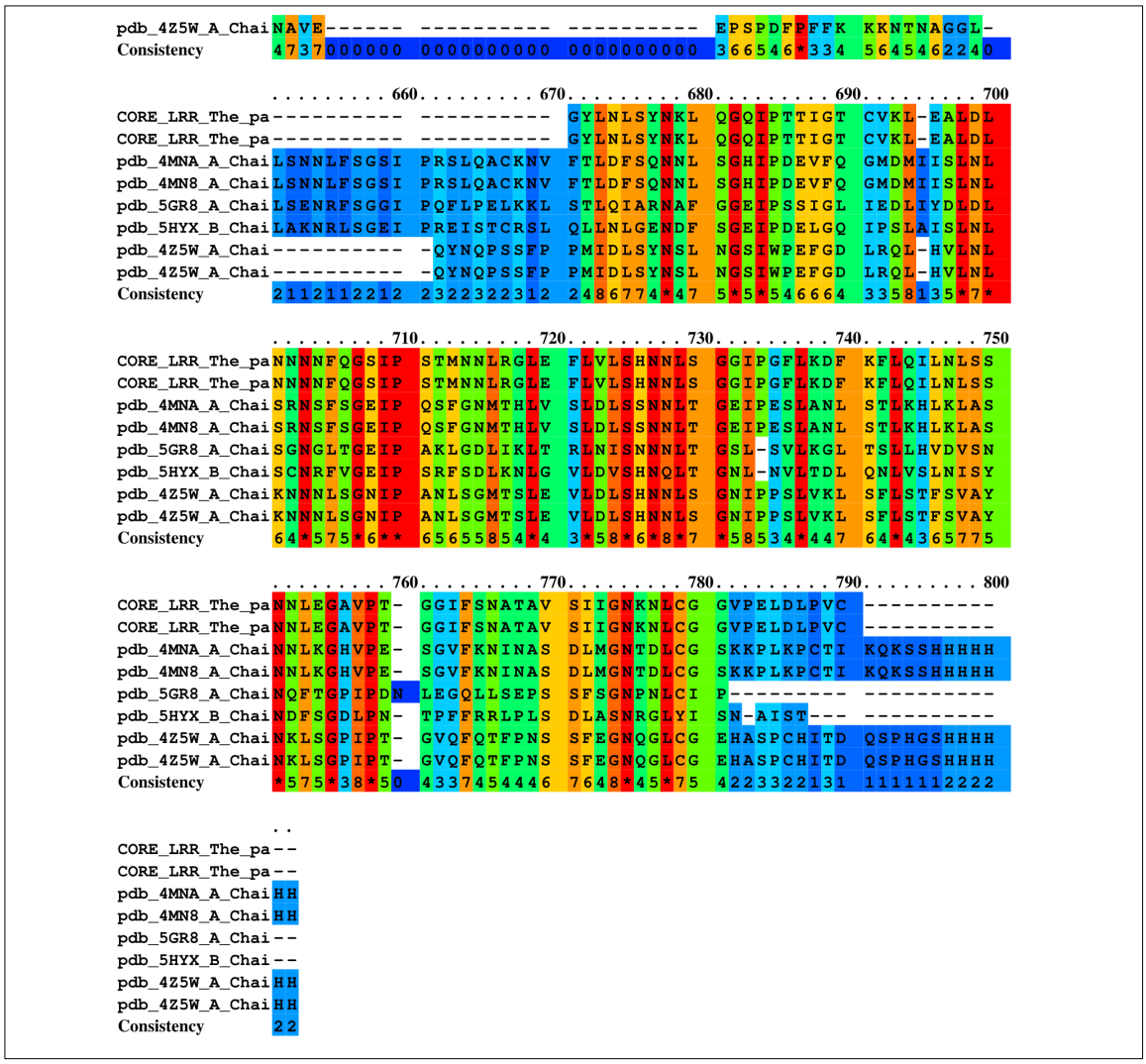


Fig 3.4: MSA of the CORE LRR sequence with the top five templates from NCBI BLASTp results.

Table 3.2: Top five templates turned up on uploading the CORE LRR sequence to the HHpred server

PDB ID hit	Name	Probability	E-value
4MNA_A	LRR receptor-like serine/threonine-protein kinase FLS2; FLS2, plant immunity, Leucine-rich repeat; HET: NAG; 3.998A {*Arabidopsis thaliana*}	100	1.1e-42
5GR8_A	Leucine-rich repeat receptor-like protein kinase; PEPR1, DAMP, PRR, AtPEP1., TRANSFERASE; HET: NAG; 2.587A {*Arabidopsis thaliana*}	100	4e-41
5HYX_B	ASP-PTR-TRP-LYS-PRO-ARG-HIS-HIS-PRO-HYP- ARG-ASN-ASN, Probable LRR receptor-like serine/threonine-protein; Plant Receptor, TRANSFERASE; HET: PTR, NAG; 2.56A {*Arabidopsis thaliana*}	100	1.3e-40
4Z5W_B	plant peptide hormone receptor; receptor, HORMONE; HET: NAG, TYS; 2.2A {*Daucus carota*}	100	4.5e-38
5GR9_B	Leucine-rich repeat receptor-like protein kinase; LRR receptor, extracellular domain, TDR; HET: NAG; 2.647A {*Arabidopsis thaliana*}	100	8.8e-41

On analyzing the validation scores (Appendix A) and also visualizing each model, it was quite evident that Model-11 (4MNA_A+5HYX_B+5GR9_B) was the most accurate representation of the CORE LRR ectodomain out of all the 30 models which were generated. With a 95.16% score on Verify 3D, indicating excellent compatibility of the model with its own AA sequence; the ERRAT score of 59.54 shows that the generated

model is robust; and finally the Ramachandran distribution shows about 98.3% (89.1%+9.2%) of the protein is in the allowed region. Thus this was selected for further treatments.

Phyre2 was only used to conduct its intensive modelling, where it used twenty templates (selected automatically by the server) based on maximum confidence, heuristics, and alignment coverage and percentage identity. All the models were constructed with over 95% confidence.

AIDA predicted the LRR domain using the templates 4MNA_A and 4MN8_A, and the final model was considered for further validation.

IntFOLD was also used for its template based modelling based of accuracy self-estimate score and refinement. This integrates the ModFOLD6_rank method for scoring the multiple-template models that were generated using a number of alternative sequence-structure alignments.

3.5 Structural validation

On visually analyzing the seven models generated by the single template modelling tools (Fig 3.7) it was evident that most tools were struggling to produce the entire protein. FFAS03, FFAS-3D, PSPS and Swiss Model showed major gaps in the middle of the LRR domain, where RaptorX and Spark X had difficulty forming the two terminals and showed a loss coil in the middle region of the protein. Only Muster was able to produce a structure that could be said to be consistent with the FLS2 ectodomain.

Following the visual analysis, different structural validation tools were used to quantitatively analyse the structure. Verify 3D was used to analyse the compatibility of the model with its own AA sequence, ERRAT was used to analyse the interactions between the AA and understand how robust the structure is, and finally the Ramachandran distribution plot showed the stability (Table 3.4) of the protein according to the rotational symmetry of the structure generated based on which the AA are labeled to be in allowed and disallowed regions.

On running the test it was seen that FFAS02 and FFAS-3D scored 75.49% and 78.11%, respectively, in Verify 3D, meaning that these models did not give acceptable results (>80). At the same time, the same two models failed to produce any results for ERRAT whereas all the others met the requirement (>50). And although none of the models acquired the ideal scores in the Ramachandran Plot (Favored region ~ 98% and Allowed region~2%) using the RAMPAGE server (Lovell *et al.*, 2003), Muster showed the most promise with there being more than 90% in the favored region.

Table 3.3: Validation scores for CORE LRR models made using single template models

Modelling tool	Verify 3D	ERRAT	Ramachandran Plot Summary from RAMPAGE (%)		
	(%)		FR	AR	OR
Muster	93.62	61.53	90.1	8.5	1.4
PSPS	87.07	67.66	84.9	14	1
Raptor X	89.14	74.26	85.6	12.8	1.6
Spark X	92.93	62.24	89.1	9.5	1.4
Swiss model	89.42	70.73	87.3	11.7	1.1
FFAS03	75.49	Failed	85.3	13.1	1.6
FFAS-3D	78.11	Failed	85	13.4	1.6

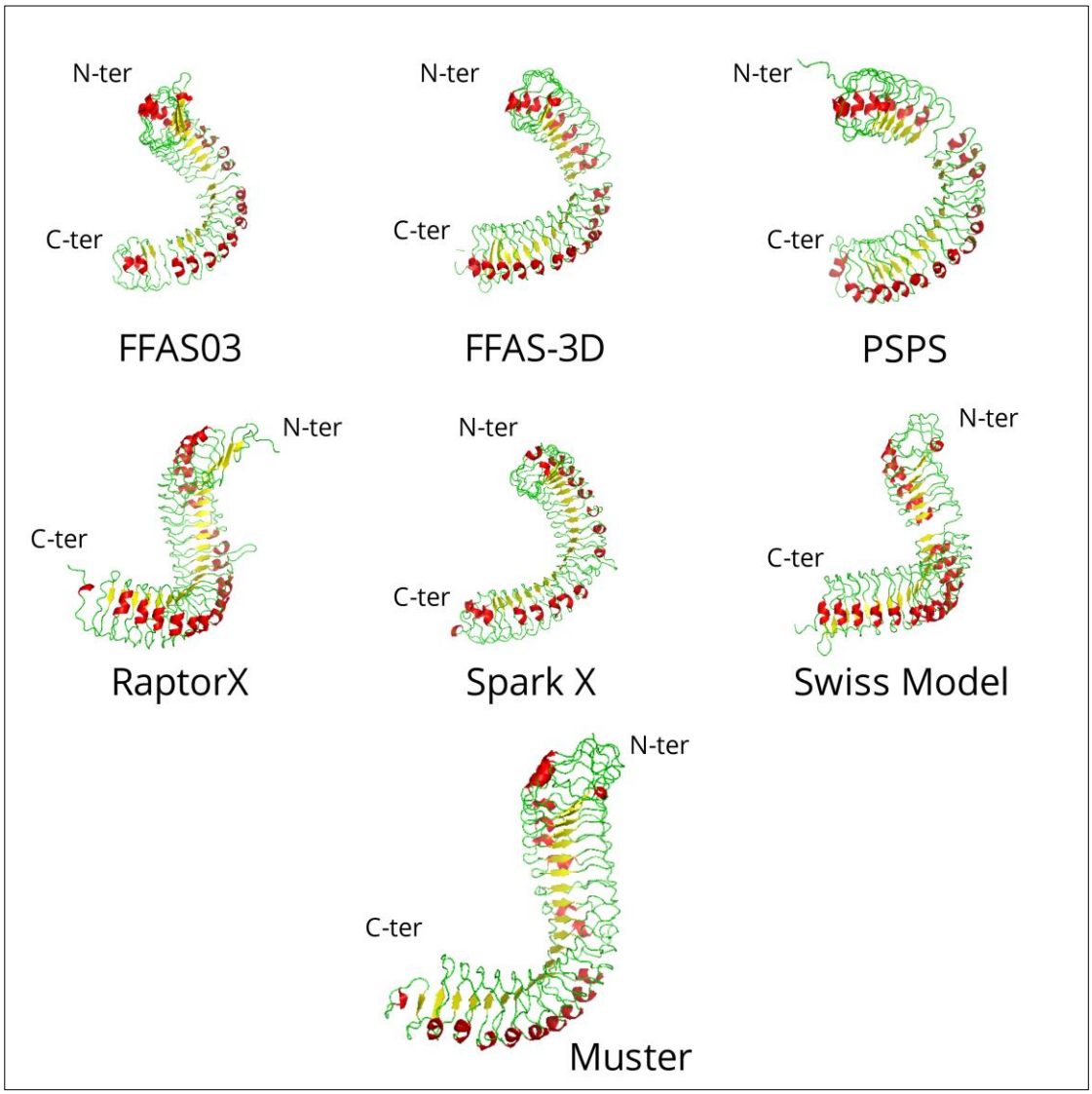


Fig 3.5: Models of the CORE LRR domain generated using Single Template Modelling tools

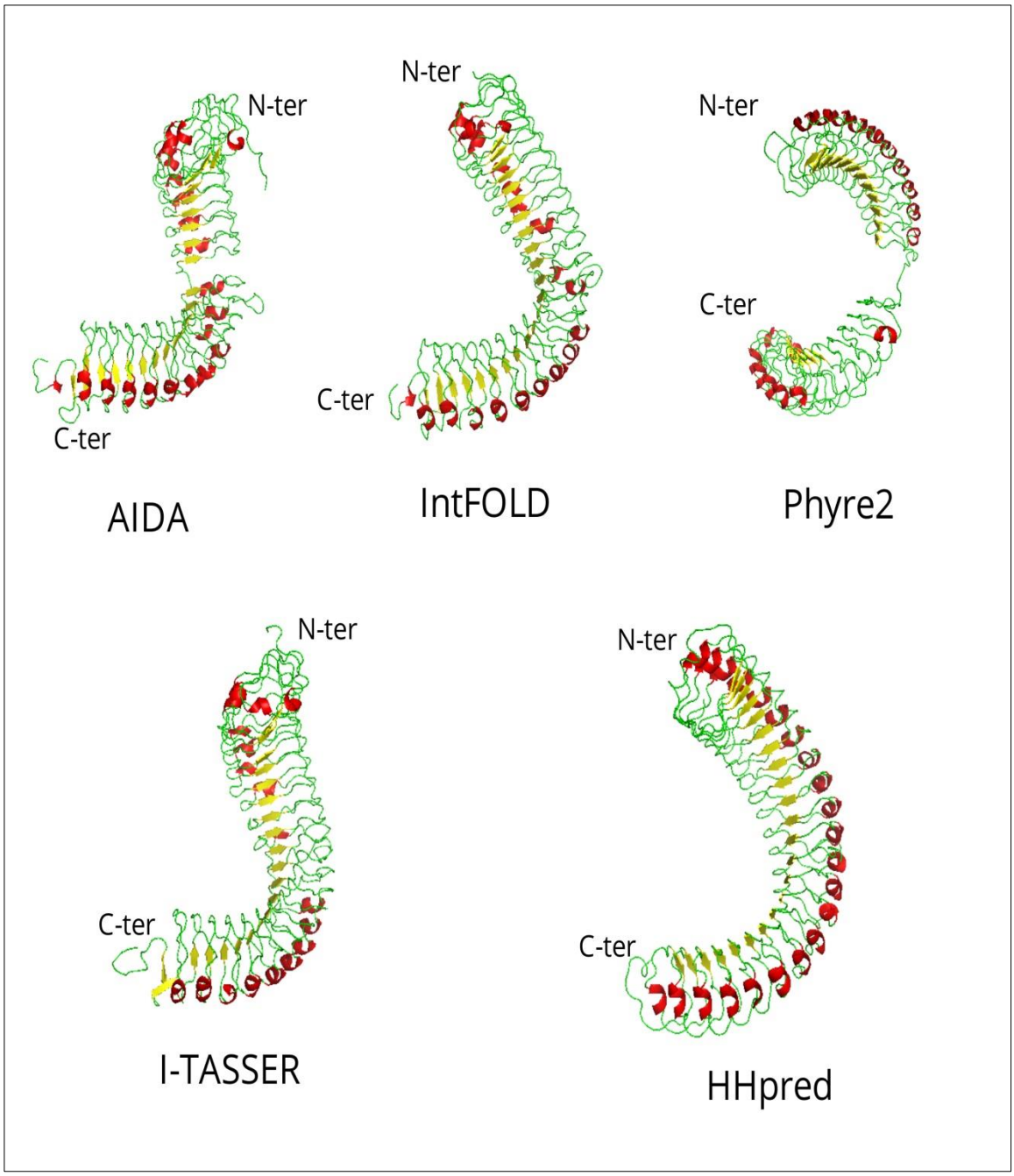


Fig 3.6: Models of the CORE LRR domain generated using Multiple Template Modelling tools

The multiple template modelling tools performed much better than the single template modelling tools (Fig 3.8), as can be visually verified. With the exception of two major gaps in the middle of the protein structure in those constructed by Phyre2 and AIDA, the others were able to construct the full ectodomain of the CORE protein. While HHpred

and I-TASSER both produced very well constructed proteins, the one produced by HHpred is seen to be a little less consistent with the LRR domain of the FLS2 crystalline complex, whereas the I-TASSER model shows a remarkable similarity.

Then similar to the single template models these models were also subjected to the same validation tools and the results were analyzed (Table 3.8). It is seen that all the models scored over 90% in Verify 3D and so qualified with great remarks, expect for the model produced by Phyre2 (intensive modelling) (70%) which failed to across the 80% acceptable score. This showed that the AA have good compatibility with their own 3D structure. All the models scored acceptable amounts on ERRAT (>50) proving that the models had robust structures. The models also have most of their residues in the allowed region (more than 95%) of the Ramachandran distribution plot with again Phyre2 having the highest amount of residues in the outlier region (3.5).

Table 3.4: Validation scores for CORE LRR models made using multiple template models

Modelling tool	Verify 3D	ERRAT	Ramachandran Plot Summary from RAMPAGE (%)		
	(%)		FR	AR	OR
Phyre2 (intensive)	70	75.72	84.9	10.6	4.5
IntFOLD	100	62.24	89.4	9	1.6
AIDA	93.97	64.34	84.6	12.6	2.8
I-TASSER	97.59	73.51	79.9	17.1	2.9
HHpred (model-11)	95.16	59.54	89.1	9.2	1.7

After visual analysis and quantitative structural validation was done, the models with the best overall scores were short listed (Table 3.7 and 3.8 – marked with blue) subjected to molecular dynamics (MD) simulation to identify which is the most accurate model of the CORE LRR ectodomain.

As for the PAMP csp22, due to the amino acid sequence being so small, among all the tool used only I-TASSER, HHpred, Quark and PepFOLD were able to model the PAMP. But then also gave some controversial results. For instance, none of the models scored any points on Verify 3D (all got zero) whereas on the other hand scored pretty high scores on ERRAT and Ramachandran distribution plot. Having already taken into account that this is a very short AA sequence, these are to be expected and so unable to identify the most accurate model of csp22, all were subjected to MD simulations.

Table 3.5: Validation scores for csp22 models constructed using various types of modelling tools

Modelling tool	Verify 3D	ERRAT	Ramachandran Plot Summary from RAMPAGE (%)		
	(%)		FR	AR	OR
HHpred	0	78.57	95	5	0
Quark	0	100	70	20	10
I-TASSER	0	100	80	15	5
PepFOLD	0	87.5	100	0	0

3.6 Molecular dynamics simulation of CORE LRR ectodomain and csp22 proteins

To better understand and verify the structures modelled by the various single and multiple template modelling tools, and also to get a clear idea of the different characteristics, a 20 ns molecular dynamics simulation was run. The results of the simulation were depicted in the form of root mean square deviation (RMSD), root mean square fluctuation (RMSF) and Radius of Gyration (Rg) graphs, and then analyzed. For the CORE LRR ectodomain, most models gave similar results for RMSD, RMSF and Rg, after the initial discrepancies, showing similar stability and compactness of the different models.

All the RMSD values of the different models were compiled and a studied (Fig 4.9). It was seen that all the models fluctuated around the relatively constant point of 0.4 nm. Among which, the HHpred model reached the highest RMSD value of about 0.6 nm at around 14 ns in the, curiously HHpred was also the model to acquire the lowest value at around the 7 ns. So, it can be assumed that this change in the RMSD was due to the protein not adhering to a fixed conformation throughout the simulation period. The other models by I-TASSER, IntFOLD and Muster, also gave similar results but did not deviate as much from the 0.4 nm value. But on finer observation it was observed that the model produced by I-TASSER was the most stable across the 20 ns simulation time. Showing the least amount of deviation and also having the lowest overall average RMSD value, as can be observed on the graph.

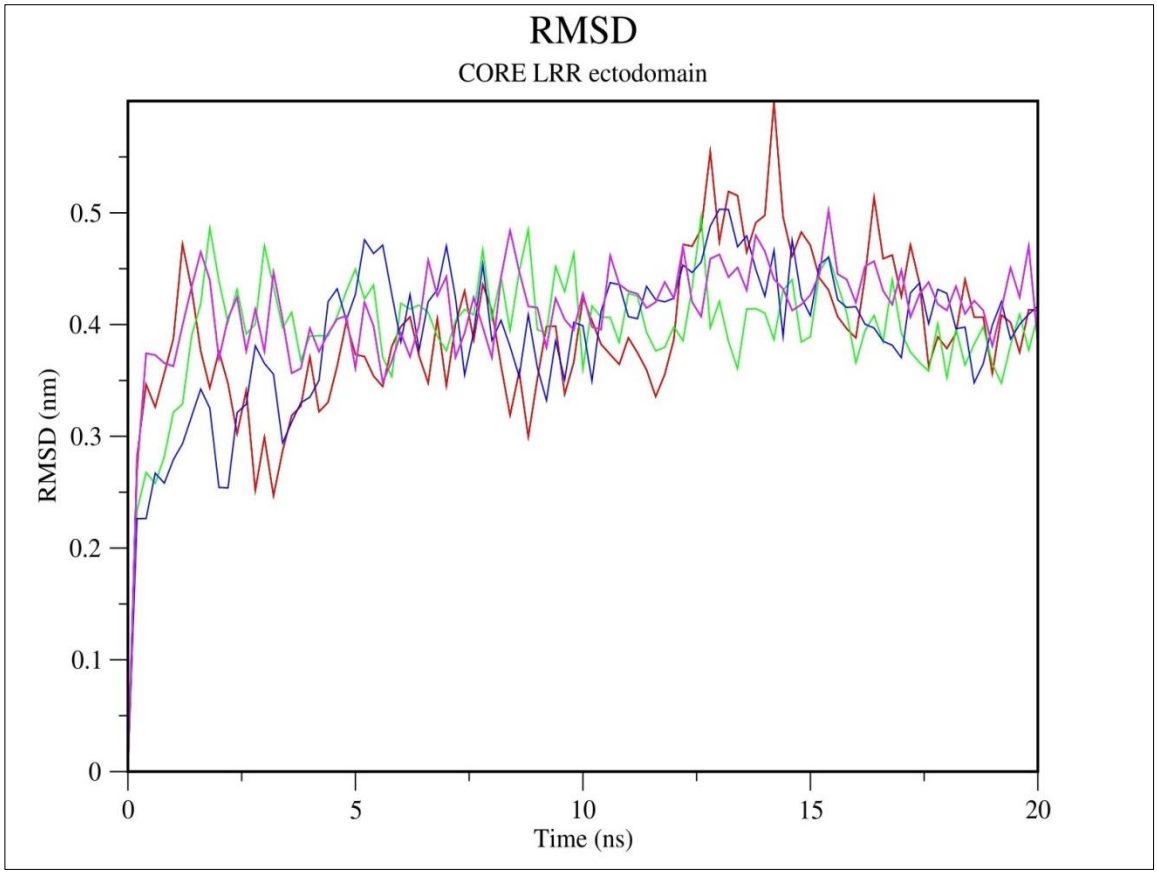


Fig 3.7: RMSD graphs generated by the four different models of the CORE LRR ectodomain. The green, red, blue and pink represent the models by I-TASSER, HHpred, IntFOLD and Muster, respectively.

Similar to the RMSD analysis, the RMSF of all the models was also observed (Fig 3.10). Surprisingly the HHpred models showed the maximum fluctuation nearing the end when approaching close to 9000 atoms, despite giving the impression of having the most stable terminals among all the generated models. Muster on the other hand showed a sizable fluctuation at around the very center of the protein, at around the 4400 atom cluster. Both I-TASSER and IntFOLD showed relatively less fluctuations thorough out the structure with I-TASSER having a net higher fluctuation at the starting (0-200 atoms) and IntFOLD having a net higher fluctuation at the far end (7500-8200 atoms).

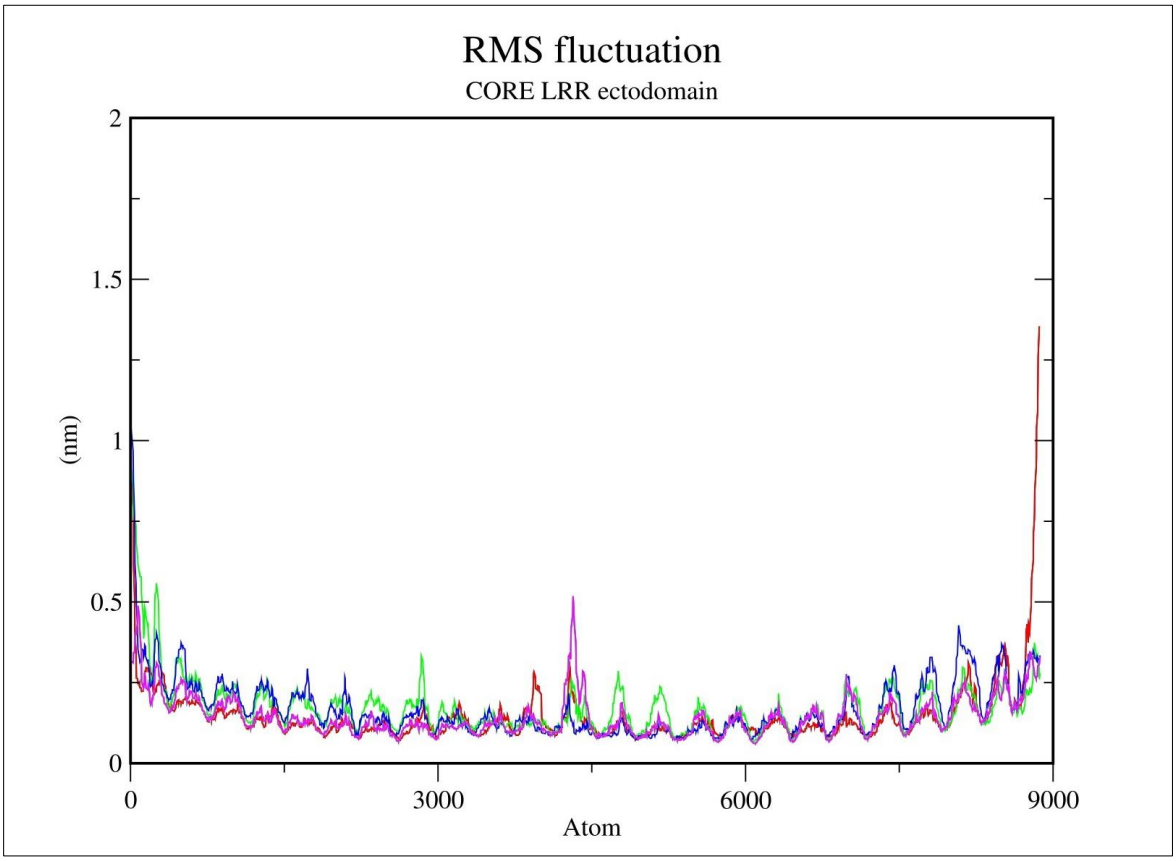


Fig 3.8: RMSF graphs generated by the four different models of the CORE LRR ectodomain. The green, red, blue and pink represent the models by I-TASSER, HHpred, IntFOLD and Muster, respectively.

The Radius of Gyration values of the different models were more mixed and so had no conclusive good model. While I-TASSER had the highest Rg vale peaking to about 3.565 nm thrice, at 2 ns, 5 ns and 7.5 ns, subsequently. At the same time I-TASSER was also the one to have the smallest deviation in the Rg value over the simulation period, proving that the model constructed by I-TASSER was the one which had the least uncoiling over the 20 ns simulation period (Fig 3.11). Completely contradictory to this, we also see the HHpred model giving the lowest Rg values but also the structure showing the maximum fluctuation, meaning that the structure is more prone to uncoiling over the simulation period.

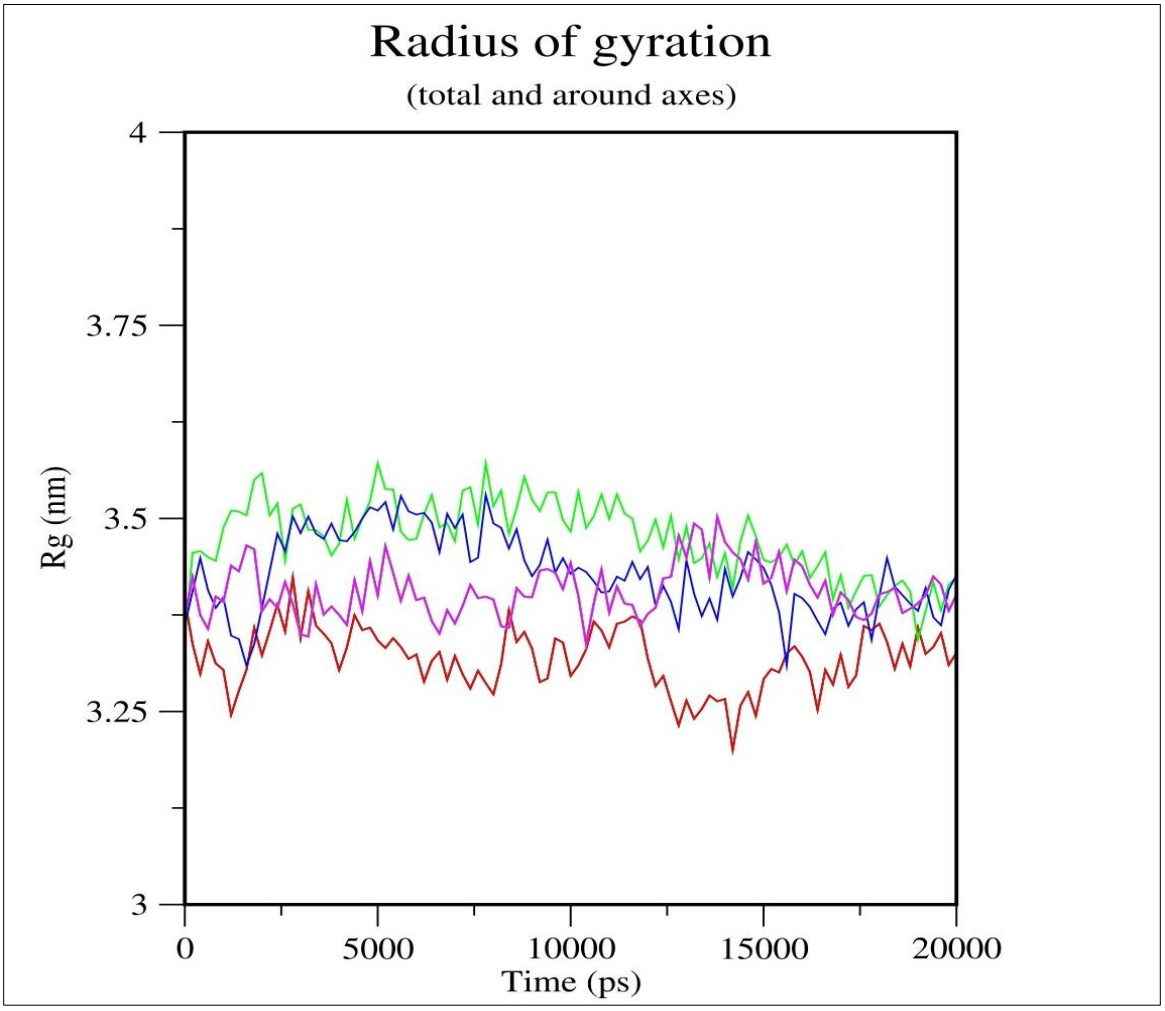


Fig 3.9: Rg graphs generated by the four different models of the CORE LRR ectodomain. The green, red, blue and pink represent the models by I-TASSER, HHpred, IntFOLD and Muster, respectively.

The same conditions were applied to the PAMP csp22, but due to its small structure the values were fluctuating a lot which is to be expected from a small protein consisting of only 22 AA. The RMSD graph showed that the I-TASSER generated model consistently has the lowest values with the minimum amount of fluctuation till 60 ns at an average value of around 0.15 nm, which is a remarkable amount of stability for a small protein (Fig 12 a). But following that its value spikes up and reaches its maximum value of 0.25 nm. Which is still very low, but then the RMSD value of the PepFOLD model took the place of the lowest value at about 0.17 nm. But the PepFOLD graph shows much more fluctuations compared to that of I-TASSER. The models generated by HHpred and Quark consistently gave a relatively high value showing much fluctuation as well, showing instability.

Similarly the RMSF (Fig 3.12 b) and Rg (Fig 3.12 c) values were also plotted on to graphs and analyzed over the 20 ns simulation period. As expected all the models showed relatively high fluctuation in both cases, and so any conclusive remark was difficult to make. But on finer observation it was seen the RMSF values of the I-TASSER and PepFOLD models were again comparatively lower, whereas when it came to the Rg value I-TASSER had a lower value proving the structure was a bit more compact among the two.

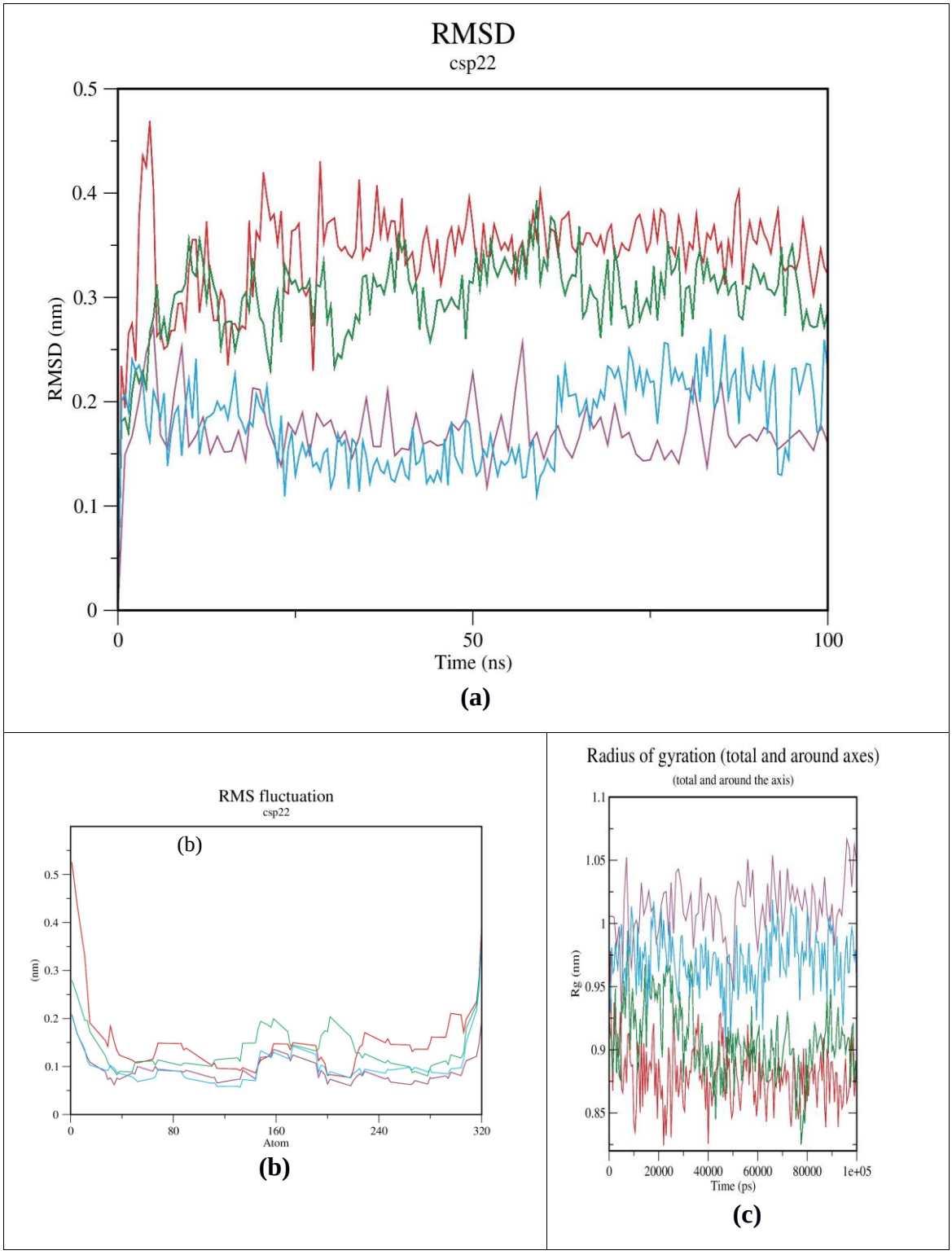


Fig 3.10: Graphs generated to analyse the structures of csp22. The red, green, purple and cyan colors are for the HHpred, Quark, PepFOLD and I-TASSER models, respectively. (a) RMSD analysis; (b) RMSF analysis; (c) Rg analysis

In lights of the molecular dynamics findings the CORE LRR ectodomain model and the csp22 models generated by I-TASSER were chosen for the purposed of observing the molecular interaction between these two proteins, in the presence of the co-receptor BAK1; which a consistent observational study as was done to observe the interactions between PRR FLS2 with PAMP flg22 in the presence BAK1, in the crystalline structure.

3.7 Molecular interaction of CORE with PAMP csp22 and co-receptor BAK1

Significantly different bonds were observed in the interaction between the CORE LRR ectodomain, csp22 and BAK1 proteins before and after the molecular dynamics simulation. In which hydrogen bonds (H-bonds), hydrophobic interactions, ionic interactions, cation-Pi interactions and aromatic interactions in the three protein complex were the most prominent. The detailed interactions between these PTI components are discussed below in details. The interactions which were seen both before and after the 20 ns molecular dynamic simulation were the probable prominent bonds and so were highlighted and analyzed (Fig 3.13).

Looking at the H-bond interactions being established, we can see that the Arginine residues (ARG) in the positions 222 and 146 of the A (CORE LRR) and C (BAK1) chains, respectively; were prominent contributors to the number of H-bonds being formed. ARG 222 of the A chain formed bonds with ASP 170 on the C chain, whereas the ARG 146 of the C chain formed bonds with the ASN 175 of the A chain. Other interaction between ASN 151 of A with TYR of C was also observed. And only one interaction between the CORE LRR domain and csp22 was observed at LYS 12 of B with TYR 324 of A (Table 3.10).

At the same time it was observed that no H-bonds between the chains A and C were seen to be present with chain B (csp22), both before and after the simulation. Some hydrophobic interactions were also observed between PHE 16 and PRO 19 of chain B with PRO 191 and ILE 192 of chain C. Whereas the residues of chain A- LEU 153, PHE 269, LEU 294 and LAL 369 interacted with residues TYR 100, LEU 188, VAL 5 and ALA 10 of chains C and B, subsequently (Table 3.11).

After that only one ionic interaction (within 6 Angstroms) and one Aromatic interaction (within 4.5 and 7 Angstroms) were observed before and after the 20 ns MD simulation (Table 3.12 and Table 3.14). At the same time it was observed that no Cation-Pi interactions between any of the chains were observed after the simulation (Table 3.13).

Table 3.6: H-bonds formed by CORE LRR domain with csp22 and BAK1

Protein-Protein Side Chain-Side Chain Hydrogen Bonds								
Before MD Simulation								
DONOR				**ACCEPTOR**				**Distance**
Position	**Chain**	**Residue**	**Atom**	**Position**	**Chain**	**Residue**	**Atom**	**(A^o)**
129	A	THR	OG1	77	C	ASN	ND2	3.11
151	**A**	**ASN**	**ND2**	**124**	**C**	**TYR**	**OH**	**2.98**
151	**A**	**ASN**	**ND2**	**124**	**C**	**TYR**	**OH**	**2.98**
175	A	ASN	ND2	124	C	TYR	OH	2.77
175	A	ASN	ND2	124	C	TYR	OH	2.77
197	A	THR	OG1	170	C	ASP	OD1	2.85
222	**A**	**ARG**	**NH2**	**170**	**C**	**ASP**	**OD1**	**2.77**
222	**A**	**ARG**	**NH2**	**170**	**C**	**ASP**	**OD1**	**2.77**
222	**A**	**ARG**	**NH2**	**170**	**C**	**ASP**	**OD2**	**2.81**
222	**A**	**ARG**	**NH2**	**170**	**C**	**ASP**	**OD2**	**2.81**
12	B	LYS	NZ	183	C	ASN	OD1	3.24
77	C	ASN	ND2	129	A	THR	OG1	3.11
77	C	ASN	ND2	129	A	THR	OG1	3.11
124	**C**	**TYR**	**OH**	**151**	**A**	**ASN**	**ND2**	**2.98**
124	C	TYR	OH	175	A	ASN	ND2	2.77
143	C	ARG	NH1	224	A	ASN	OD1	2.67

143	C	ARG	NH1	224	A	ASN	OD1	2.67
143	C	ARG	NH2	224	A	ASN	OD1	2.69
143	C	ARG	NH2	224	A	ASN	OD1	2.69
146	**C**	**ARG**	**NE**	**175**	**A**	**ASN**	**OD1**	**3.42**
146	**C**	**ARG**	**NE**	**175**	**A**	**ASN**	**ND2**	**3.11**
146	**C**	**ARG**	**NH2**	**175**	**A**	**ASN**	**OD1**	**2.67**
146	C	ARG	NH2	175	A	ASN	OD1	2.67
146	**C**	**ARG**	**NE**	**199**	**A**	**HIS**	**NE2**	**3.09**
190	C	THR	OG1	221	A	TYR	OH	2.87
149	A	LYS	NZ	22	B	GLY	O	2.76
221	A	TYR	OH	1	B	ALA	O	3.15
222	A	ARG	NH1	190	C	THR	O	2.74
222	A	ARG	NH1	190	C	THR	O	2.74
246	A	ARG	NH2	13	B	GLY	O	2.7
246	A	ARG	NH2	13	B	GLY	O	2.7
246	A	ARG	NH1	187	C	SER	O	2.66
246	A	ARG	NH1	187	C	SER	O	2.66
246	A	ARG	NH1	188	C	LEU	O	2.79
246	A	ARG	NH1	188	C	LEU	O	2.79
264	A	ARG	NE	3	B	GLY	O	2.89
1	B	ALA	N	194	A	ASP	OD1	3.3
1	B	ALA	N	194	A	ASP	OD2	3.14
1	B	ALA	N	219	A	SER	OG	2.87
5	B	VAL	N	319	A	GLU	OE1	3.18
5	B	VAL	N	319	A	GLU	OE2	3.05
12	**B**	**LYS**	**N**	**324**	**A**	**TYR**	**OH**	**2.9**

After MD Simulation								
DONOR				**ACCEPTOR**				**Distance**
Position	**Chain**	**Residue**	**Atom**	**Position**	**Chain**	**Residue**	**Atom**	**(A^o)**
151	**A**	**ASN**	**OD1**	**124**	**C**	**TYR**	**OH**	**2.67**
151	**A**	**ASN**	**OD1**	**124**	**C**	**TYR**	**OH**	**2.67**
222	**A**	**ARG**	**NH1**	**170**	**C**	**ASP**	**OD2**	**2.55**
222	**A**	**ARG**	**NH1**	**170**	**C**	**ASP**	**OD2**	**2.55**
222	**A**	**ARG**	**NH2**	**170**	**C**	**ASP**	**OD2**	**3.42**
222	**A**	**ARG**	**NH2**	**170**	**C**	**ASP**	**OD2**	**3.42**
348	A	TYR	OH	11	B	GLU	OE2	2.56
124	**C**	**TYR**	**OH**	**151**	**A**	**ASN**	**OD1**	**2.67**
143	C	ARG	NE	201	A	GLU	OE1	3.36
143	C	ARG	NE	201	A	GLU	OE2	2.78
143	C	ARG	NH2	201	A	GLU	OE1	2.59
143	C	ARG	NH2	201	A	GLU	OE1	2.59
146	**C**	**ARG**	**NE**	**175**	**A**	**ASN**	**ND2**	**3.32**
146	**C**	**ARG**	**NH1**	**175**	**A**	**ASN**	**ND2**	**3.08**
146	**C**	**ARG**	**NH1**	**175**	**A**	**ASN**	**ND2**	**3.08**
146	**C**	**ARG**	**NH2**	**199**	**A**	**HIS**	**ND1**	**2.77**
146	C	ARG	NH2	199	A	HIS	ND1	2.77
246	A	ARG	NH2	15	B	GLY	O	3.04
246	A	ARG	NH2	15	B	GLY	O	3.04
324	A	TYR	OH	10	B	ALA	O	2.65
12	**B**	**LYS**	**N**	**324**	**A**	**TYR**	**OH**	**3.15**
13	B	GLY	N	324	A	TYR	OH	3.34

Table 3.7: Hydrophobic Interactions formed by CORE LRR domain with csp22 and BAK1

Hydrophobic Interactions within 5 Angstroms					
Before MD Simulation					
Position	**Residue**	**Chain**	**Position**	**Residue**	**Chain**
1	ALA	B	192	ILE	C
14	PHE	B	180	ILE	C
14	PHE	B	182	VAL	C
14	PHE	B	186	PHE	C
14	PHE	B	189	PHE	C
14	PHE	B	191	PRO	C
14	PHE	B	194	PHE	C
16	PHE	B	180	ILE	C
16	PHE	B	182	VAL	C
16	**PHE**	**B**	**191**	**PRO**	**C**
17	ILE	B	191	PRO	C
19	**PRO**	**B**	**191**	**PRO**	**C**
19	**PRO**	**B**	**192**	**ILE**	**C**
103	MET	A	125	LEU	C
153	**LEU**	**A**	**100**	**TYR**	**C**
217	PHE	A	2	VAL	B
221	TYR	A	1	ALA	B
243	ILE	A	2	VAL	B
266	LEU	A	2	VAL	B
269	**PHE**	**A**	**188**	**LEU**	**C**
294	**LEU**	**A**	**5**	**VAL**	**B**
343	ALA	A	5	VAL	B

345	PHE	A	10	ALA	B
345	PHE	A	5	VAL	B
345	PHE	A	7	TRP	B
345	PHE	A	8	PHE	B
347	ALA	A	10	ALA	B
369	**ALA**	**A**	**10**	**ALA**	**B**
After MD Simulation					
Position	**Residue**	**Chain**	**Position**	**Residue**	**Chain**
7	TRP	B	182	VAL	C
16	PHE	B	182	VAL	C
16	**PHE**	**B**	**191**	**PRO**	**C**
19	**PRO**	**B**	**191**	**PRO**	**C**
19	**PRO**	**B**	**192**	**ILE**	**C**
153	**LEU**	**A**	**100**	**TYR**	**C**
217	PHE	A	1	ALA	B
221	TYR	A	192	ILE	C
243	ILE	A	17	ILE	B
266	LEU	A	17	ILE	B
266	LEU	A	5	VAL	B
269	**PHE**	**A**	**188**	**LEU**	**C**
294	**LEU**	**A**	**5**	**VAL**	**B**
324	TYR	A	8	PHE	B
345	PHE	A	8	PHE	B
369	**ALA**	**A**	**10**	**ALA**	**B**
369	ALA	A	8	PHE	B
371	LEU	A	10	ALA	B

Table 3.8: Ionic interactions formed by CORE LRR domain with csp22 and BAK1

Ionic Interactions within 6 Angstroms					
Before MD Simulation					
Position	**Residue**	**Chain**	**Position**	**Residue**	**Chain**
222	**ARG**	**A**	**170**	**ASP**	**C**
264	ARG	A	20	ASP	B
After MD Simulation					
Position	**Residue**	**Chain**	**Position**	**Residue**	**Chain**
201	GLU	A	143	ARG	C
222	**ARG**	**A**	**170**	**ASP**	**C**
395	ARG	A	11	GLU	B

Table 3.9: Cation-Pi Interactions formed by CORE LRR domain with csp22 and BAK1

Cation-Pi Interactions within 6 Angstroms							
Before MD Simulation							
Position	**Residue**	**Chain**	**Position**	**Residue**	**Chain**	**D(cation-Pi)**	**Angle**
7	TRP	B	391	LYS	A	5.14	160.73
After MD Simulation							
No Cation-Pi interactions were observed at the end of the 20 ns molecular dynamics simulation.							

Table 3.10: Aromatic interactions formed by CORE LRR domain with csp22 and BAK1

Aromatic-Aromatic Interactions within 4.5 and 7 Angstroms							
Before MD Simulation							
Residue	**Position**	**Chain**	**Residue**	**Position**	**Chain**	**D(centroid-centroid)**	**Dihedral Angle**
14	PHE	B	194	PHE	C	6.87	57.02
345	PHE	A	7	TRP	B	5.3	101.76
345	**PHE**	**A**	**8**	**PHE**	**B**	**5.46**	**109.86**
After MD Simulation							
Residue	**Position**	**Chain**	**Residue**	**Position**	**Chain**	**D(centroid-centroid)**	**Dihedral Angle**
324	TYR	A	8	PHE	B	6.93	100
345	**PHE**	**A**	**8**	**PHE**	**B**	**4.81**	**152.48**

Chain A is for CORE LRR ectodomain; Chain B is for csp22; and Chain C is for BAK1; the interactions in bold were seen both before and after the 20 ns MD simulation

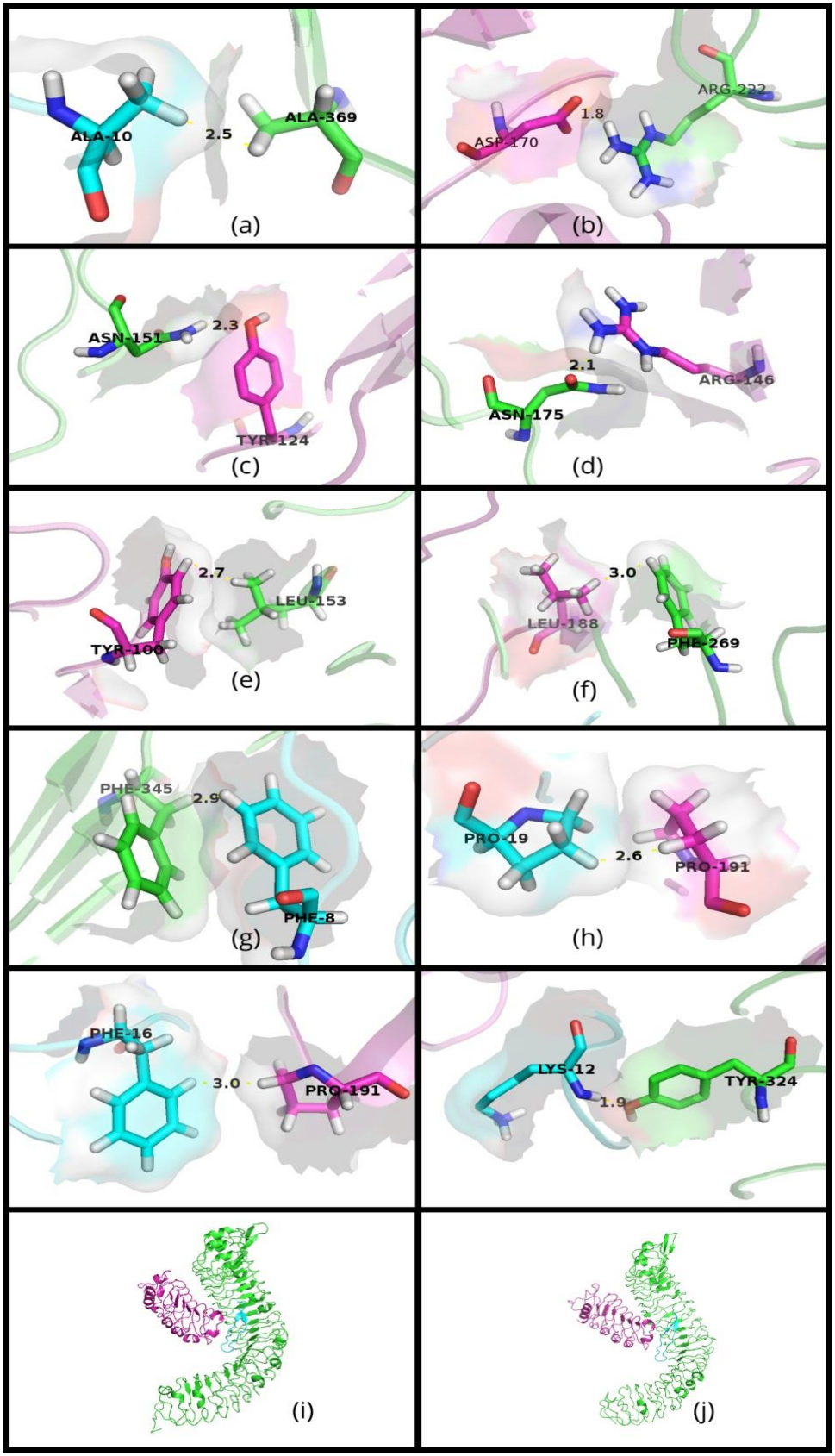


Fig 3.11: (a-h) Molecular interactions between CORE LRR (green), csp22 (cyan) and BAK1 (purple); (i) Cartoon representation of the complex of CORE LRR (green), csp22 (cyan) and BAK1 (purple) before the simulation and (j) after the simulation.

3.8 Comparative study between CORE-csp22-BAK1 complex and FLS2-flg22-BAK1 crystal structure

On observing the three protein (CORE LRR, csp22 and BAK1) docked complex and analyzing it, much structural similarities where seen with the FLS2 LRR-flg22-BAK1 crystal structure in its binding behavior. It can be seen that the PAMP csp22 binds to one lateral side of the LRR of CORE with the co-receptor BAK1 binding head first on to them, forming a heterodimer.

Just as it is observed in the FLS2-BAK1 complex that a heterodimer formation is induced by the PAMP flg22 (Sun *et al.*, 2013), for the CORE-BAK1 complex also the heterodimer is established induced by the PAMP csp22 (ref. Diff – PNAS).

Also, there are lots of hydrogen bonds formed between FLS2 and flg22 and between FLS2 and Bak1 protein which contribute significantly for in case of binding. A total 31 H-bonds formed between FLS2 and PAMP flg22 and 27 H-bonds formed between FLS2 and co-receptor Bak1 protein. Among these, FLS2 Tyr272 and Tyr296 and flg22 Lys13 contribute significantly to the interactions around this interface (Fig 3.14). (Sun *et al.*, 2013)

On analyzing the CORE LRR, csp22 and BAK1 complex, we observed that there was a total of 42 H-bonds of which 30 of the bonds were between CORE LRR and BAK1 whereas on 12 H-bonds were between CORE LRR and csp22. And no H-bond interactions were observed between the PAMP and the co-receptor. The residues ARG222 and ARG146 of the A (CORE LRR) and C (BAK1) chains, respectively; were prominent contributors to the number of H-bonds being formed. ARG222 on chain A formed bonds with ASP170 on the C chain, whereas the ARG 146 of the C chain formed bonds with the ASN 175 of the A chain. There was only one H-bond interaction between the CORE LRR domain (TYR 324) and csp22 (LYS 12) which was observed both before and after the simulation. Other similar hydrophobic, ionic and aromatic interactions were also observed, but very few were seen after the simulation (Table 3.11, 3.12, 3.13 and 3.14) (Fig 3.13)

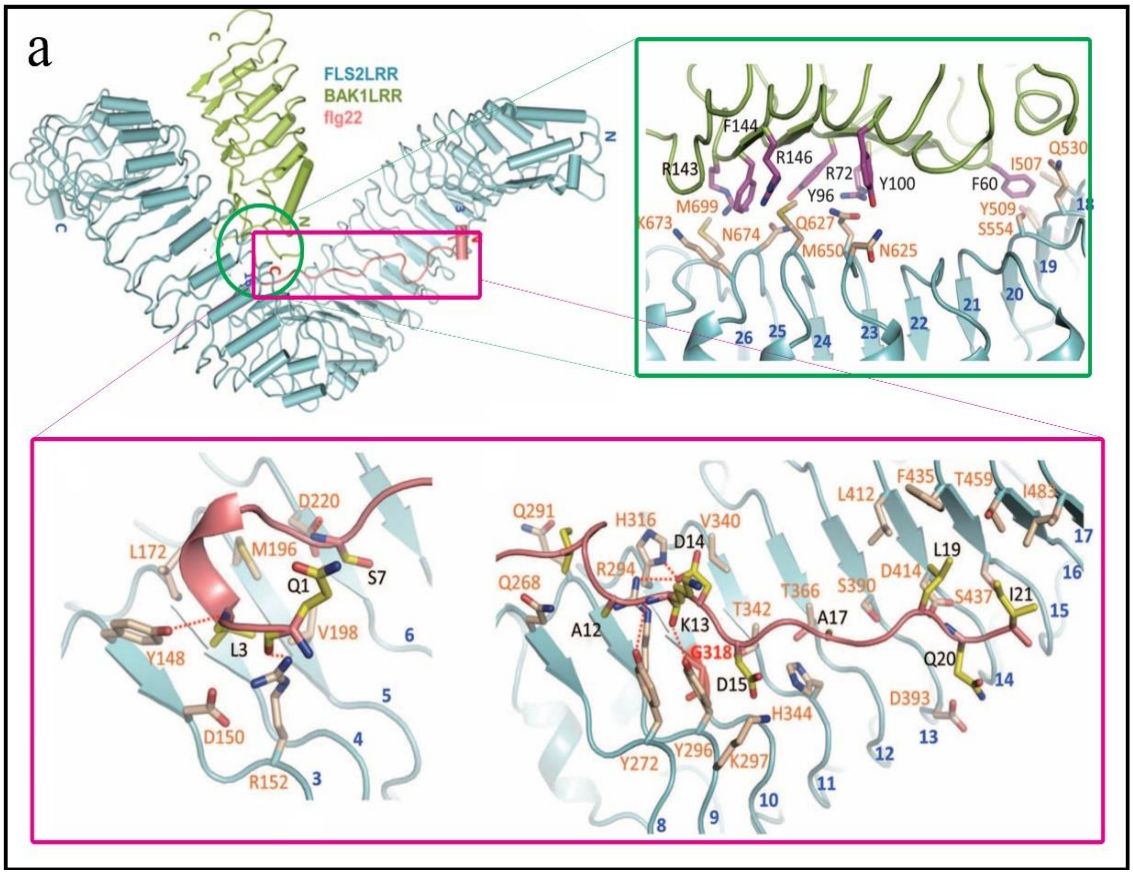


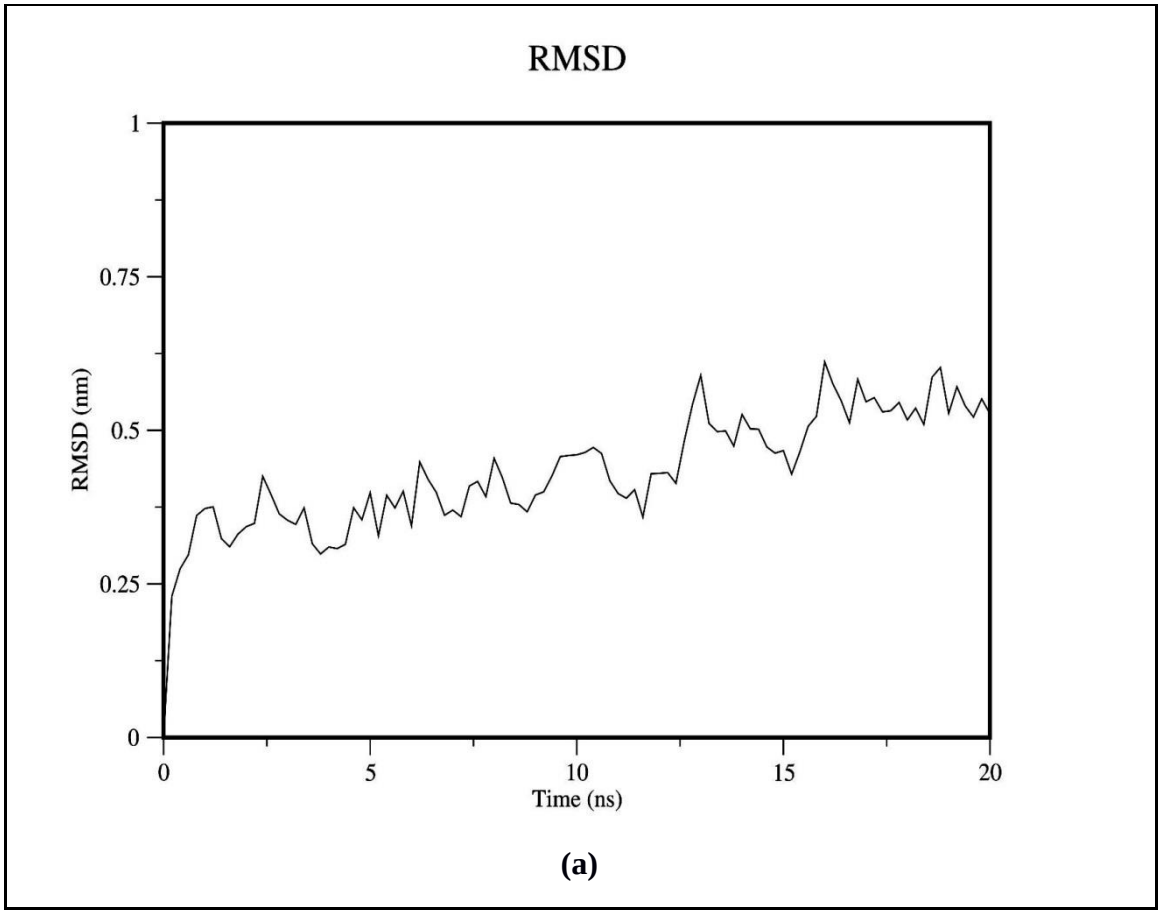
Fig 3.12: Binding method in FLS2 complex showing various interactions for FLS2 (cyan), flg22 (pink) and BAK1 (green).

3.9 Molecular dynamics simulation of CORE LRR ectodomain, csp22 and BAK1 complex

RMSD and RMSF for backbone atoms and radius of gyration (Rg) of C-alpha atoms were calculated for the complex of CORE LRR with csp22 and BAK1. The RMSD graph (Fig 3.15 a) shows an initial spike going up to 0.35 nm in the first 2.5 ns. After that the overall average RMSD value was seen to be 0.35 nm with a fluctuation of about +/-0.5 nm though out the rest of the 20 ns simulation. The highest value was seen to be reached at around 5 ns, which was of about 0.4 nm. From these values it can be concluded that the three protein complex formed after csp22 binds to the CORE LRR aided by co-receptor BAK1 is a very stable structure. Proving that the modelled proteins (CORE LRR and csp22) are very accurate and the produce realistic interactions as expected *in vivo*.

The RMSF graph (Fig 3.15 b) produced showed maximum fluctuation in the beginning and in the starting of the last quartile of the graph. The maximum fluctuation reached was about 0.6 nm, by about 100 atoms in the 8800 to 8900 atoms interval. This fluctuation might have been due to some conformational changes cause by the interaction of the different proteins.

The Rg values for C-alpha atom also showed that the complex was quite active over the simulation period but at the same time was not uncoiling itself. The values ranged from about 3.125 to 3.45 nm over the 20 ns (20000 ps) simulation time. Reaching a maximum value at around 16 ns of 3.45 nm and then came back down (Fig 3.15 c).



(a)

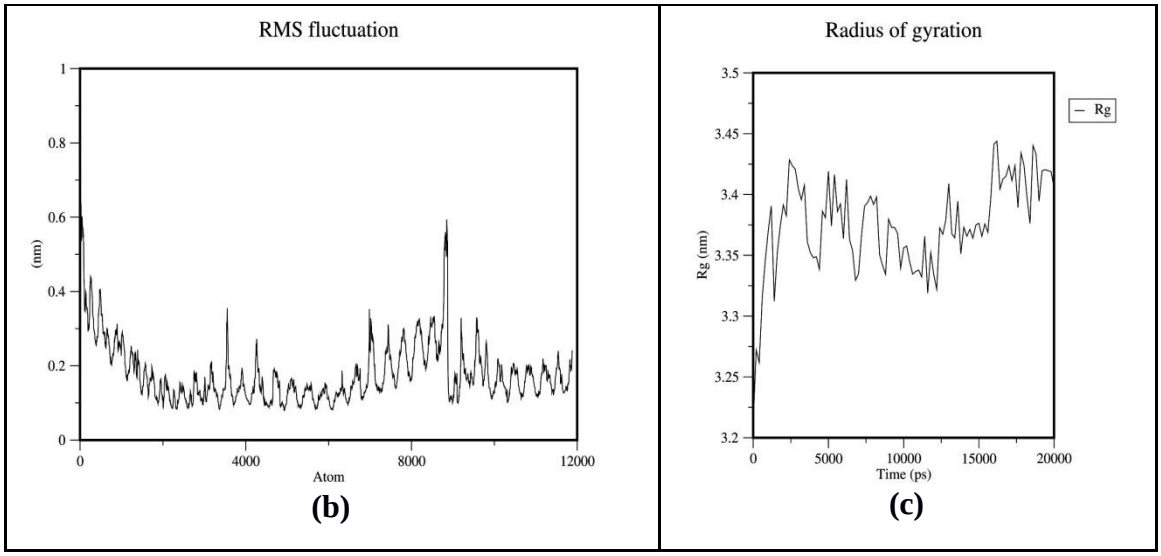


Fig 3.13: The following graph were generated for the CORE LRR, csp22 and BAK1 three protein complex- (a) RMSD graph; (b) RMSF; (c) Radius of gyration (Rg)

3.10 Summery

This chapter elaborately describes all the steps followed in order to achieve a full understand of the interaction between the CORE ectodomain with the PAMP csp22, with the help of the co-receptor BAK1; along with the molecular dynamics study of the PRR and PAMP before observing the interactions and also that of the structural dynamics after the docking. Finally, the interactions were quantified and compared to see the overall changes which took place during the 20 ns MD simulation. It was observed that there was a significant loss in interaction over all, which included hydrogen bonding, hydrophobic interactions, ionic interactions, cation-Pi interactions and aromatic interaction (Table 3.15). It was concluded that this might have been a result of the proteins changing their conformation over the simulation period.

Table 3.11: Summary of interactions among CORE ectodomain, csp22 and BAK1 in the complex

Interaction between	**H-bond**		**Hydrophobic**		**Ionic**		**Cation-Pi**		**Aromatic**	
	B. MD	**A. MD**	**B. MD**	**A. MD**	**B. MD**	**A. MD**	**B. MD**	**A. MD**	**B. MD**	**A. MD**
CORE_csp22	11	6	12	11	1	1	1	0	2	2
CORE_BAK1	30	16	3	2	1	2	0	0	0	0
BAK1_csp22	1	0	13	5	0	0	0	0	1	0
Total	42	22	28	18	2	3	1	0	3	2

B. MD, Before molecular dynamics simulation; A. MD, After molecular dynamics simulation

CHAPTER 4

CONCLUSIONS AND RECOMMENDATIONS

4.1 Conclusions

The main objectives of this study were to model the ectodomain of the CORE protein for analyzing the CORE LRR mediated pattern triggered immunity in plant. The study concludes to the follows:

1. The CORE LRR modelled by I-TASSER gave the best model of the CORE protein's ectodomain. In the intensive modelling study it could be seen that for PRR ectodomain proteins, the multiple template modelling tools yielded better results compared to single template modelling approach.
2. The binding mechanism of the csp22 with the CORE LRR domain and then the BAK1 joining up to form a heterodimer is just consistent with the crystalline structure of FLS2 complex, which is the only solved crystal structure of these types of interactions.

Results presented in this thesis show the evidence that the objectives of the research have been successfully achieved.

This study is the first ever *in silico* approach towards modelling the ectodomain of the CORE protein and to observe its interaction with its PAMP csp22 and co-receptor BAK1.

4.2 Recommendations for Future Works

This research can be further improved by adopting some measures as following:

1. The study can be improved by running the molecular dynamics simulation for a much longer time (micro/millisecond), which would allow more conclusive conclusions to be drawn from the study, as a better understanding of the protein's nature might be understood.
2. The interaction of PRR CORE can be observed with other PAMPs, such as csp15, to better understand the difference in activity for different but closely related PAMPs.
3. The interaction of the PRR-PAMP complex with a mutated co-receptor can be observed to see how mutations at certain residues or clusters effect their interactions, and their ability to trigger patter-triggered immunity.

REFERENCES

Al-Lazikani, B., Jung, J., Xiang, Z., and Honig, B. (2001). Protein structure prediction. *Current opinion in chemical biology, 5*(1), 51-56.

Allen, M. P. (2004). Introduction to molecular dynamics simulation. *Computational soft matter: from synthetic polymers to proteins, 23*, 1-28.

Armon, A., Graur, D., and Ben-Tal, N. (2001). ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *Journal of molecular biology, 307*(1), 447-463.

Bates, P. A., Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. (2001). Enhancement of protein modelling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins: Structure, Function, and Bioinformatics, 45*(S5), 39-46.

Bae W., Xia, B., Inouye, M. and Severinov, K. (2000). Escherichia coli CspA-family RNA chaperones are transcription antiterminators. *Proc. Natl Acad. Sci. USA* 97, 7784–7789.

Bent A. F., and Mackey, D. (2007). Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. *Annu. Rev. Phytopathol., 45*, 399-436.

Buchan, D. W., Ward, S., Lobley, A. E., Nugent, T., Bryson, K., and Jones, D. T. (2010). Protein annotation and modelling servers at University College London. *Nucleic acids research*, gkq427.

Buenavista, M. T., Roche, D. B., and McGuffin, L. J. (2012). Improvement of 3D protein models using multiple templates guided by single-template model quality assessment. *Bioinformatics, 28*(14), 1851-1857.

Chen, C.-C., Hwang, J.-K., and Yang, J.-M. (2006). 2: protein structure prediction server. *Nucleic acids research, 34*(suppl 2), W152-W157.

Chinchilla, D., Bauer, Z., Regenass, M., Boller, T., and Felix, G. (2006). The Arabidopsis receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception. *The Plant Cell, 18*(2), 465-476.

Chisholm, S. T., Coaker, G., Day, B., and Staskawicz, B. J. (2006). Host-microbe interactions: shaping the evolution of the plant immune response. *Cell, 124*(4), 803-814.

Combet, C., Jambon, M., Deléage, G., and Geourjon, C. (2002). Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics, 18*(1), 213-214.

Comeau SR, Gatchell DW, Vajda S, Camacho CJ. (2004). ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Research*.

Comeau SR, Gatchell DW, Vajda S, Camacho CJ. (2004). ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*.

Dangl, J. L., and Jones, J. D. (2001). Plant pathogens and integrated defense responses to infection. *nature, 411*(6839), 826-833.

Dardick, C., and Ronald, P. (2006). Plant and animal pathogen recognition receptors signal through non-RD kinases. *PLoS Pathog, 2*(1), e2.

DeLano, W. L. (2002). The PyMOL molecular graphics system.

Di Matteo, A., Federici, L., Mattei, B., Salvi, G., Johnson, K., Savino, C., et al. (2003). The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proceedings of the National Academy of Sciences, 100*(17), 10124-10128.

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols, 2*(4), 953-971.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M., Eramian, D., Shen, M. y., et al. (2006). Comparative protein structure modelling using Modeller. *Current protocols in bioinformatics*, 5.6. 1-5.6. 30.

Felix, G., Duran, J. D., Volko, S., and Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *The Plant Journal, 18*(3), 265-276.

Fiser, A. (2004). Protein structure modelling in the proteomics era.

García-Sánchez, M. I., Díaz-Quintana, A., Gotor, C., Jacquot, J.-P., Miguel, A., and Vega, J. M. (2000). Homology predicted structure and functional interaction of ferredoxin from the eukaryotic alga Chlamydomonas reinhardtii with nitrite reductase and glutamate synthase. *JBIC Journal of Biological Inorganic Chemistry, 5*(6), 713-719.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005). *Protein identification and analysis tools on the ExPASy server*: Springer.

Gómez-Gómez, L., and Boller, T. (2000). FLS2: An LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. *Molecular cell, 5*(6), 1003-1011.

Hann, D. R., and Rathjen, J. P. (2007). Early events in the pathogenicity of Pseudomonas syringae on Nicotiana benthamiana. *The Plant Journal, 49*(4), 607-618.

Hothorn, M., Belkhadir, Y., Dreux, M., Dabi, T., Noel, J. P., Wilson, I. A., et al. (2011). Structural basis of steroid hormone perception by the receptor kinase BRI1. *Nature, 474*(7352), 467-471.

Huang, Y. J., Mao, B., Aramini, J. M., and Montelione, G. T. (2014). Assessment of template-based protein structure predictions in CASP10. *Proteins: Structure, Function, and Bioinformatics, 82*(S2), 43-56.

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., et al. (2009). InterPro: the integrative protein signature database. *Nucleic acids research, 37*(suppl 1), D211-D215.

Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., and Godzik, A. (2005). FFAS03: a server for profile–profile sequence alignments. *Nucleic acids research, 33*(suppl 2), W284-W288.

Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature, 444*(7117), 323-329.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., et al. (2012). Template-based protein structure modelling using the RaptorX web server. *Nature protocols, 7*(8), 1511-1522.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015). The Phyre2 web portal for protein modelling, prediction and analysis. *Nature protocols, 10*(6), 845-858.

Kozakov D, Beglov D, Bohnuud T, Mottarella S, Xia B, Hall DR, Vajda, S. (2013) How good is automated protein docking? *Proteins: Structure, Function, and Bioinformatics*.

Kozakov D, Brenke R, Comeau SR, Vajda S. (2006). PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins*.

Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S. (2017). The ClusPro web server for protein-protein docking. *Nature Protocols*;12(2):255-278

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography, 26*(2), 283-291.

Lei W., Markus A., Elias E., Ursula F. Damaris K., and Georg F. (2016). The pattern-recognition receptor CORE of Solanaceae detects bacterial cold-shock protein. *Nature Plants*, 2:16185

Liithy, R., Bowie, J., and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature, 356*(6364), 83-85.

Lobley, A., Sadowski, M. I., and Jones, D. T. (2009). pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics, 25*(14), 1761-1767.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., et al. (2003). Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics, 50*(3), 437-450.

Mahram, A., and Herbordt, M. C. (2010). *Fast and accurate NCBI BLASTP: acceleration with multiphase FPGA-based prefiltering.* Paper presented at the Proceedings of the 24th ACM International Conference on Supercomputing, 73-82.

Medzhitov, R., and Janeway, C. A. (1997). Innate immunity: the virtues of a nonclonal system of recognition. *Cell, 91*(3), 295-298.

Offord, V., Coffey, T., and Werling, D. (2010). LRRfinder: a web application for the identification of leucine-rich repeats and an integrative Toll-like receptor database. *Developmental & Comparative Immunology, 34*(10), 1035-1041.

Park, C.-J., Caddell, D. F., and Ronald, P. C. (2012). Protein phosphorylation in plant immunity: insights into the regulation of pattern recognition receptor-mediated signaling. *Frontiers in plant science, 3*, 177.

Postma, J. *et al.* (2016). Avr4 promotes Cf-4 receptor-like protein association with the BAK1/SERK3 receptor-like kinase to initiate receptor endocytosis and plant immunity. *New Phytol*. 210, 627–642

Pruitt, R. N. *et al*. (2015). The rice immune receptor XA21 recognizes a tyrosine-sulfated protein from a gram-negative bacterium. Sci. Adv. 1, e1500245.

Robatzek, S., Bittel, P., Chinchilla, D., Köchner, P., Felix, G., Shiu, S.-H., et al. (2007). Molecular identification and characterization of the tomato flagellin receptor LeFLS2, an orthologue of Arabidopsis FLS2 exhibiting characteristically different perception specificities. *Plant molecular biology, 64*(5), 539-547.

Rohl, C. A., Strauss, C. E., Misura, K. M., and Baker, D. (2004). Protein structure prediction using Rosetta. *Methods in enzymology, 383*, 66-93.

Ronald, P. C., and Beutler, B. (2010). Plant and animal sensors of conserved microbial signatures. *Science, 330*(6007), 1061-1064.

Rost, B., Yachdav, G., and Liu, J. (2004). The predictprotein server. *Nucleic acids research, 32*(suppl 2), W321-W326.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols, 5*(4), 725-738.

Schaftenaar, G., and Noordik, J. H. (2000). Molden: a pre-and post-processing program for molecular and electronic structures. *Journal of computer-aided molecular design, 14*(2), 123-134.

Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., et al. (2011). Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European biophysics journal, 40*(7), 843-856.

Schürch, S., Linde, C. C., Knogge, W., Jackson, L. F., and McDonald, B. A. (2004). Molecular population genetic analysis differentiates two virulence mechanisms of the fungal avirulence gene NIP1. *Molecular Plant-Microbe Interactions, 17*(10), 1114-1125.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003). SWISS-MODEL: an automated protein homology-modelling server. *Nucleic acids research, 31*(13), 3381-3385.

Sheridan, R., Fieldhouse, R. J., Hayat, S., Sun, Y., Antipin, Y., Yang, L., et al. (2015). EVfold. org: Evolutionary Couplings and Protein 3D Structure Prediction. *bioRxiv*, 021022.

Simossis, V. A., and Heringa, J. (2005). PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic acids research, 33*(suppl 2), W289-W294.

Söding, J., Biegert, A., and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research, 33*(suppl 2), W244-W248.

Song, W.-Y., Wang, G.-L., Chen, L.-L., Kim, H.-S., Pi, L.-Y., Holsten, T., et al. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science, 270*(5243), 1804-1806.

Stefanie R. (2017). Sensing of molecular patterns through cell surface immune receptors. *Current Opinion in Plant Biology.* 28:68-77

Sun, Y., Li, L., Macho, A. P., Han, Z., Hu, Z., Zipfel, C., et al. (2013). Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. *Science, 342*(6158), 624-628.

Takai, R., Isogai, A., Takayama, S., and Che, F.-S. (2008). Analysis of flagellin perception mediated by flg22 receptor OsFLS2 in rice. *Molecular Plant-Microbe Interactions, 21*(12), 1635-1642.

Thomma, B. P., Nürnberger, T., and Joosten, M. H. (2011). Of PAMPs and effectors: the blurred PTI-ETI dichotomy. *The plant cell, 23*(1), 4-15.

Trdá, L., Fernandez, O., Boutrot, F., Héloir, M. C., Kelloniemi, J., Daire, X., et al. (2014). The grapevine flagellin receptor VvFLS2 differentially recognizes flagellin-derived epitopes from the endophytic growth-promoting bacterium Burkholderia phytofirmans and plant pathogenic bacteria. *New Phytologist, 201*(4), 1371-1384.

Tsuda, K., and Katagiri, F. (2010). Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Current opinion in plant biology, 13*(4), 459-465.

Tsuda, K., Sato, M., Stoddard, T., Glazebrook, J., and Katagiri, F. (2009). Network properties of robust immunity in plants. *PLoS Genet, 5*(12), e1000772.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *Journal of computational chemistry, 26*(16), 1701-1718.

van der Spoel, D., van Maaren, P. J., and Berendsen, H. J. (1998). A systematic study of water models for molecular simulation: derivation of water models optimized for use with a reaction field. *The Journal of chemical physics, 108*(24), 10220-10230.

Wallner, B., and Elofsson, A. (2003). Can correct protein models be identified? *Protein Science, 12*(5), 1073-1086.

Wallner, B., Fang, H., and Elofsson, A. (2003). Automatic consensus-based fold recognition using Pcons, ProQ, and Pmodeller. *Proteins: Structure, Function, and Bioinformatics, 53*(S6), 534-541.

Wiederstein, M., and Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research, 35*(suppl 2), W407-W410.
Wu, S., and Zhang, Y. (2008). MUSTER: improving protein sequence profile–profile alignments by using multiple sources of structure information. *Proteins: Structure, Function, and Bioinformatics, 72*(2), 547-556.
Xu, D., Jaroszewski, L., Li, Z., and Godzik, A. (2013). FFAS-3D: improving fold recognition by including optimized structural features and template re-ranking. *Bioinformatics*, btt578.
Xu, D., Jaroszewski, L., Li, Z., and Godzik, A. (2014). AIDA: ab initio domain assembly server. *Nucleic acids research, 42*(W1), W308-W313.
Xu, D., and Zhang, Y. (2012). Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins: Structure, Function, and Bioinformatics, 80*(7), 1715-1735.
Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature methods, 12*(1), 7-8.
Yang, Y., Faraggi, E., Zhao, H., and Zhou, Y. (2011). Improving protein fold recognition and template-based modelling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics, 27*(15), 2076-2082.
Zhou, H., and Skolnick, J. (2009). Protein structure prediction by pro-Sp3-TASSER. *Biophysical journal, 96*(6), 2119-2127.
Zipfel, C. (2014). Plant pattern-recognition receptors. *Trends in immunology, 35*(7), 345-351.
Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J. D., Boller, T., et al. (2006). Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. *Cell, 125*(4), 749-760.

APPENDIX A

Validation scores for all the models constructed using the HHpred toolkit.

Models	Verify 3D	ERRAT	Ramachandran Plot Summary from RAMPAGE (%)		
	(%)		FR	AR	OR
1	92.92	66.37	87.9	9.7	2.4
2	94.3	64.8	92	6.8	1.2
3	94.1	63.05	88.4	9.9	1.7
4	94.13	63.75	89.3	8.5	2.3
5	91.19	63.17	91.2	7.5	1.4
6	93.61	61.12	89.9	9	1
7	95.34	65.5	88	10.7	1.2
8	94.47	64.45	89.6	8.8	1.6
9	94.82	68.3	88.4	9.4	2.3
10	92.4	64.91	90.3	7.8	1.9
11	95.16	59.54	89.1	9.2	1.7
12	89.98	67.78	90.5	7.6	1.9
13	95.51	57.97	89.6	9	1.4
14	93.78	60.7	87.9	10.4	1.7
15	90.33	63.57	88.9	9.7	1.6
16	96.2	67.95	89.3	8.7	2.1
17	94.73	59.36	88.7	8.1	3.2
18	91.88	65.32	91.3	6.4	2.3
19	87.05	62.7	89.4	9	1.6
20	94.3	62.63	89.8	8.5	1.7
21	94.96	56.44	89.7	8.7	1.6
22	93.74	58.38	86.9	11.2	1.9
23	92.92	58.38	88.9	9	2.1

24	91.88	63.05	88.4	9.5	2.1
25	87.46	65.82	86	12.4	1.6
26	87.28	66.49	90.8	7.1	2.1
27	91.83	68.96	88.8	9.6	1.6
28	88.05	62.08	88.7	9.2	2.1
29	90.23	63.41	90.7	7.7	1.6
30	92.33	66.25	85.1	12.9	1.9

APPENDIX B

Different bioinformatics tools used in the study

Serial No.	Tool	Used for
1	ProtParam	Primary Structure Prediction
2	PSIPRED	Secondary Structure Prediction
3	ConSurf	Conserved Region Prediction
4	InterPro	Domain Prediction
6	NCBI BLASTp	Sequence Alignment
8	SWISS-MODEL	Homology Modelling
9	Prospect2	Threading Modelling
10	FFAS-3D	Threading Modelling
11	FFAS03	Threading Modelling
12	Sparks-X	Threading Modelling
13	Muster	Threading Modelling
14	AIDA	Homology Modelling
15	I-TASSER	Threading Modelling
16	Phyre-2	Threading Modelling
17	Raptor-X	Threading Modelling
18	HHpred	Homology Modelling
19	Quark	Threading Modelling
20	IntFOLD	Threading Modelling
21	ERRAT	Structure Validation
22	Verify-3D	Structure Validation
23	RAMPAGE (Ramachandran plot)	Structure Validation
24	GROMACS	MD Simulation
25	PyMOL	Molecular Visualisation