BACTERIAL EXPRESSION OF AN EFFECTOR PROTEIN OF YELLOW RUST PATHOGEN AND A RESISTANCE PROTEIN OF WHEAT AND CHARACTERIZATION OF THE EFFECTOR PROTEIN

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SAYIT MAHMUT ERDOĞAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
BIOTECHNOLOGY

JANUARY 2019

Approval of the thesis:

BACTERIAL EXPRESSION OF AN EFFECTOR PROTEIN OF YELLOW RUST PATHOGEN AND A RESISTANCE PROTEIN OF WHEAT AND CHARACTERIZATION OF THE EFFECTOR PROTEIN

submitted by **SAYIT MAHMUT ERDOĞAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Biotechnology Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Assoc. Prof. Dr. Can Özen
Head of Department, **Biotechnology** _____

Prof. Dr. Mahinur S. Akkaya
Supervisor, **Biotechnology, METU** _____

Assist. Prof. Dr. Salih Özçubukçu
Co-Supervisor, **Chemistry Dept., METU** _____

Examining Committee Members:

Assoc. Prof. Dr. Çağdaş Devrim Son
Biology, Middle East Technical University _____

Prof. Dr. Mahinur S. Akkaya
Biotechnology, METU _____

Prof. Dr. Semra İde
Physical Engineering Dept. Hacettepe University _____

Assoc. Prof. Dr. Yeşim Soyer
Department of Food Engineering, METU _____

Assist. Prof. Dr. Urartu Özgür Şafak Şeker
UNAM, Bilkent University _____

Date: 16.01.2019

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Sayıt Mahmut Erdoğan

Signature:

ABSTRACT

BACTERIAL EXPRESSION OF AN EFFECTOR PROTEIN OF YELLOW RUST PATHOGEN AND A RESISTANCE PROTEIN OF WHEAT AND CHARACTERIZATION OF THE EFFECTOR PROTEIN

Erdoğan, Sayıt Mahmut
Doctor of Philosophy, Biotechnology
Supervisor: Prof. Dr. Mahinur S. Akkaya
Co-Supervisor: Assist. Prof. Dr. Salih Özçubukçu

January 2019, 122 pages

Yellow rust is one of the most important wheat disease encountered in large parts of the world and in our country caused by a pathogen which is called as *Puccinia striiformis* f. sp. tritici (Pst). Rapid alterations in the pathogen virulency can make previously resistant varieties susceptible to the disease. Finding out common or species-specific genes, participated in the plant-pathogen interactions will provide an understanding of the biological mechanisms of the disease. Within the scope of this thesis followings were carried out: Coding region of TaYr10 protein, which resists Avr containing pathogen races, were separated into its domains and these domains were expressed in *E.coli* expression strains and purified by Immobilized Metal Affinity Chromatography (IMAC), Ultrafiltration and Size Exclusion Chromatography (SEC).

Similarly, coding gene of the candidate effector protein of *Puccinia striiformis* f. sp. tritici (Pst), pstSCR1 which causes disease was expressed in *E.coli* expression strains and purified by IMAC, ultrafiltration and SEC. Among the studied proteins, only PstSCR1 protein achieved the requisite amount and purity to further study the structural analysis. Using the pure PstSCR1 protein, structural analyses were carried

out by SAXS and NMR and as a result, 3D structure of the protein was identified in nanometer levels by SAXS. Additionally, secondary structure of the PstSCR1 protein was identified by CDS and the melting point of the protein was determined DSF. The data obtained in these studies provide important clues at the molecular level for the future studies about the disease mechanism of the pathogen and possible effector-resistance protein interactions.

.

Keywords: Yellow (stripe) rust of wheat, Puccinia striiformis, PstSCR1 effector, SAXS, Protein NMR

ÖZ

SARI PAS PATOJENİNE AİT BİR EFEKTÖR PROTEİNİN VE BUĞDAYA AİT BİR DİRENÇ PROTEİNİNİN BAKTERİ İÇİNDE İFADESİ VE EFEKTÖR PROTEİNİN KARAKTERİZASYONU

Erdoğan, Sayıt Mahmut
Doktora, Biyoteknoloji
Tez Danışmanı: Prof. Dr. Mahinur S. Akkaya
Ortak Tez Danışmanı: Dr. Öğr. Üyesi Salih Özçubukçu

Ocak 2019, 122 sayfa

Sarı pas, *Puccinia striiformis* f. sp. *tritici* (Pst) patojeninin neden olduğu, dünyanın büyük bölümünde ve ülkemizde buğdayda rastlanan en önemli hastalıklardan birisidir. Patojen popülasyonunun virulanlığındaki hızlı değişimler daha önce dirençli olan varyeteleri hassas hale getirebilir. Bitki-patojen etkileşmesinde görevli ortak veya türe özgü genlerin bulunması bu genlerin görev aldığı biyolojik mekanizmaların anlaşılmasını sağlayacaktır. Bu tez kapsamında şunlar yapılmıştır: Buğdayda, Avr içeren patojen ırklarına karşı direnç gösteren Tayr10 proteini domainlerine ayrılmış ve bu domainleri kodlayan bölgeler *E.coli* ifade suşları içinde ifade edilmiş ve yeterli ifade miktarına ulaşan domainler İmmobilize Metal Afinite Kromatografisi (IMAC), Ultrafiltreleme ve boyut dışlama kromatografisi (SEC) ile saflaştırılmıştır.

Aynı şekilde, *Puccinia striiformis* f. sp. *tritici* (Pst) patojeninin buğdayda hastalığa neden olan aday efektörü PstSCR1 proteinini kodlayan geni *pstSCR1 E.coli* ifade suşları içinde ifade edilerek IMAC, Ultrafiltreleme ve SEC ile saflaştırılmıştır. Çalışılan proteinler içerisinde yapı tayini için elde edilmesi gerekli olan miktar ve saflığa ulaşan protein olan saf PstSCR1 proteini ile, SAXS (Küçük açılı X-ışını

saçılması) ve NMR (Nükleer Manyetik Rezonans) yapı analizleri yapılarak proteinin nanometre düzeyinde üç boyutlu yapısı tayin edilmiştir. Ayrıca, PstSCR1 proteininin sekonder yapısı Dairesel Dikroizm Spektroskopisi (CDS) yöntemiyle belirlenmiş ve proteinin erime noktası (Tm) Diferansiyel Taramalı Florometri (DSF) ile tespit edilmiştir. Elde edilen bilgiler patojenin hastalık oluşturma mekanizması ve olası efektör-direnç proteini etkileşim çalışmaları için moleküler düzeyde önemli ipuçları sağlamaktadır.

Anahtar Kelimeler: Yellow (stripe) rust of wheat, Puccinia striiformis, PstSCR1 effector, SAXS, Protein NMR

To my family

ACKNOWLEDGMENTS

I would like to present my deepest gratitude to my advisor Prof. Dr. Mahinur S. Akkaya for her continuous support and valuable discussions during my PhD study and precious contribution to my thesis. I am very satisfied for all of the subjects and methods that I learned from her and the experiences that I gained from her laboratory and therefore feel very fortunate because of doing my PhD study under her supervision.

I would like to thank Assist. Prof. Dr. Salih Özçubukçu for being my co-supervisor and for his valuable contributions to my thesis.

I would like to express my special appreciation to Assist. Prof. Dr. Mehmet Şen for accepting me to his laboratory in the University of Houston, Texas, USA to carry out significant part of this thesis in his laboratory and supervising me about the experiments and methods throughout my study.

I am grateful to Prof. Dr. Semra İde (Hacettepe University) and Prof. Dr. U-Ser Jeng (NSSRC - Taiwan) for carrying out SAXS analysis of my samples in NSRRC - Taiwan and their valuable contribution to my thesis.

I-would-like to thank my TMC members; Assoc. Prof. Dr. Yeşim Soyer (METU) and Assist. Prof. Dr. Urartu-Şeker (Bilkent University - UNAM) and for their valuable recommendations and criticism on the thesis.

I want to thank to the laboratory members in METU; Çağlar, Bayantes, Ayşe, and Zemran and the ones in the University of Houston; Pragya, Collins, Omar and Josh. I appreciate their suggestions and brainstorming on my research and for their friendship.

I specially thank to my father, mother and siblings for their endless support and encouragement throughout my life.

I was supported by TÜBİTAK during my PhD study within the scope of TÜBİTAK (2211-C) scholarship programme. I would like to acknowledge TÜBİTAK and METU-BAP financial supports

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AB: Antibody

Avr: Avirulence

BAK1: Brassinosteroid Insensitive 1-Associated Kinase 1

β -ME: beta-Mercaptoethanol

Da: Dalton

DNA: Deoxyribonucleic acid

dNTP: Deooxy-nucleotidetriphosphate

EDTA: Ethylene diamine tetraacetic acid

e/p: enzyme/protein ratio

ETI: Effector triggered immunity

F-primer: Forward primer

GOI: Gene of interest

HR: Hypersensitive response

IMAC: Immobilized metal affinity chromatography

IP: Immunoprecipitation

iRNA: Interfering RNA

kb: Kilobase

kDa: Kilo dalton

Kan: Kanamycin

LRR: Leucine rich repeat

LB: Lysogeny broth

MAPKs: Mitogen-activated protein kinases

M: Molar

mL: Milliliter

MW: Molecular weight

MWCO: Molecular weight cut-off value

NB: Nucleotide binding

ng: Nanogram

Ni-NTA: Nickel-nitrilotriacetic (Ni bound to agrose bead by chelation using nitriloacetic acid beads)

ON: Over night

p: plasmid

PAMP: Pathogen-associated molecular pattern

pmol: Picomole

PMSF: Phenylmethylsulfonyl-fluoride (protease inhibitor)

PRR: Pattern recognition receptor

POI: Protein of interest

Pst: *Puccinia striiformis* f. sp. *tritici*

PTI: PAMP triggered immunity

PVDF: Polyvinylidene fluoride

R: Resistance

Rg: Radius of gyration

R-Primer: Reverse primer

RT: Room temperature

SAR: Systemic acquired resistance

SEC: Size-Exclusion (Gel Filtration) Chromatography

s/n: Signal to noise ratio

Taq: Thermus aquaticus

TDNA: Transfer DNA

TIR: Toll, interleukin-1 receptor

Tm: Melting point

TPI: Tagged-protein immunoprecipitation

T3SS: Type 3 secretion system

WB: Western blot

CHAPTER 1

INTRODUCTION

1.1 Yellow rust disease of wheat

Yellow rust also known as stripe rust disease of wheat had been a globally devastating epidemic caused by *Puccinia striiformis f. sp. tritici* (*Pst*) occurring in several wheat fields which have moist and cool climate in the growing seasons (Chen, Wellings, Chen, & Liu, 2014). Recently, newly formed virulences of disease agent Pst have been reported, creating even more crop loss (Ali, Rodriguez-Algaba, Thach, & et al., 2017). Pst infects the green tissue of wheat by secreting proteins called effectors into the cytoplasm, apoplast, or organelles of the plant (Bozkurt, Schornack, Banfield, & Kamoun, 2012) (Dodds & Rathjen, 2010) (Kamoun, 2006), to mediate nutrient uptake and suppress host immunity, subsequently causing disease by forming yellow urediniospores appearing as stripes on the leaves. Although, genomic studies led prediction of many effector candidates (Cantu D., et al., 2013) (Wenming, Zeng. & et al., 2013) (Yin C., et al., 2009) molecular mechanisms that facilitate host or non-host resistance are not well characterized.

One of the Pst effector protein PstSCR1 was first detected as a small cysteine-rich candidate effector (Yin et al., 2009). Previously, we have reported (Dagvadorj, et al., 2017) that PstSCR1 as secreted by *Puccinia striformis* f. sp. *tritici* triggers PTI responses at the cell surface (apoplast). The expression of PstSCR1 showed that it was targeted to the extracellular space of *Nicotiana benthamina* and enhanced disease resistance activating immunity in non-host plant. In addition to investigate the mode of action of these effectors proteins in vivo, the structural analysis of the

pure protein may facilitate the understanding of its cellular function (Zhang X., et al., 2017) Thus, here in this study structure of the protein was identified by SAXS.

Apoplastic effector proteins may have an interaction with the proteins or extracellular molecules located on the surface of the host organism and may cause disturbance of the function of the related molecules (Dodds & Rathjen, 2010). Under the group of apoplastic effectors, there are toxins, ethylene inducing peptides, cell wall-degrading enzymes, and cysteine-rich small proteins which are predicted to be stronger and durable in severe apoplastic conditions by forming disulfide bonds (Dodds & Rathjen, 2010) (Saunders & et al., 2012). Some effectors can even trigger cell mortality via immune receptors which are localized on the surface of the host (Dodds & Rathjen, 2010) (Rivas & Thomas, 2005) (Ma & et al., 2015) (Postma & et al., 2016) (Tyler & et al., 2006).

1.2 Principles of plant immunity

Pathogens occupy extracellular niches to provide nutrients from host cells to enable growth. During an invasion, mainly cytoplasm and the organelles of an host organism are attacked by the pathogen as the main sites of interactions. In order to detect pathogen invasion, plants use two strategies. These are PAMP Triggered and Effector Triggered Immunities which are abbreviated as PTI and ETI respectively (Chisholm, Coaker, Day, & Staskawicz, 2006) (Jones & Dangl, 2006) (Figure 1).

1.2.1 PAMP Triggered Immunity (PTI)

One of the strategies applied by the plants against pathogen invasion is the recognition of the conserved microbial elicitors on the outer surface of the plant

cells. These molecular patterns are called as pathogen associated (or microbe associated) molecular patterns and abbreviated as PAMPs or MAMPs. These microbial elicitors are conserved molecules and their structures are common to microbial organisms. Receptor proteins of the host cell localized in the plasma membrane which detects PAMPs are called as pattern recognition receptors and are abbreviated as PRRs. Flagellin structure of bacteria and chitin structure of fungi are typical examples of PAMP, as these essential molecular patterns stimulate PRRs of the host cell and thereafter leading to PTI (Dodds & Rathjen, 2010).

Pattern recognition receptors (PRR) are inadequate for apparent signaling domain and they are classified in two different transmembrane receptor classes. These classes are either receptor like proteins or receptor kinases and they recognize the extracellular molecules of the pathogens (Zipfel, 2008). Plant PRRs recognize diverse group of PAMPs including proteins, lipids, carbohydrates and some other molecules including ATP. In contrary to expectations it is found out that all of the PAMPs can not be detected by all species (Boller & Felix, 2009).

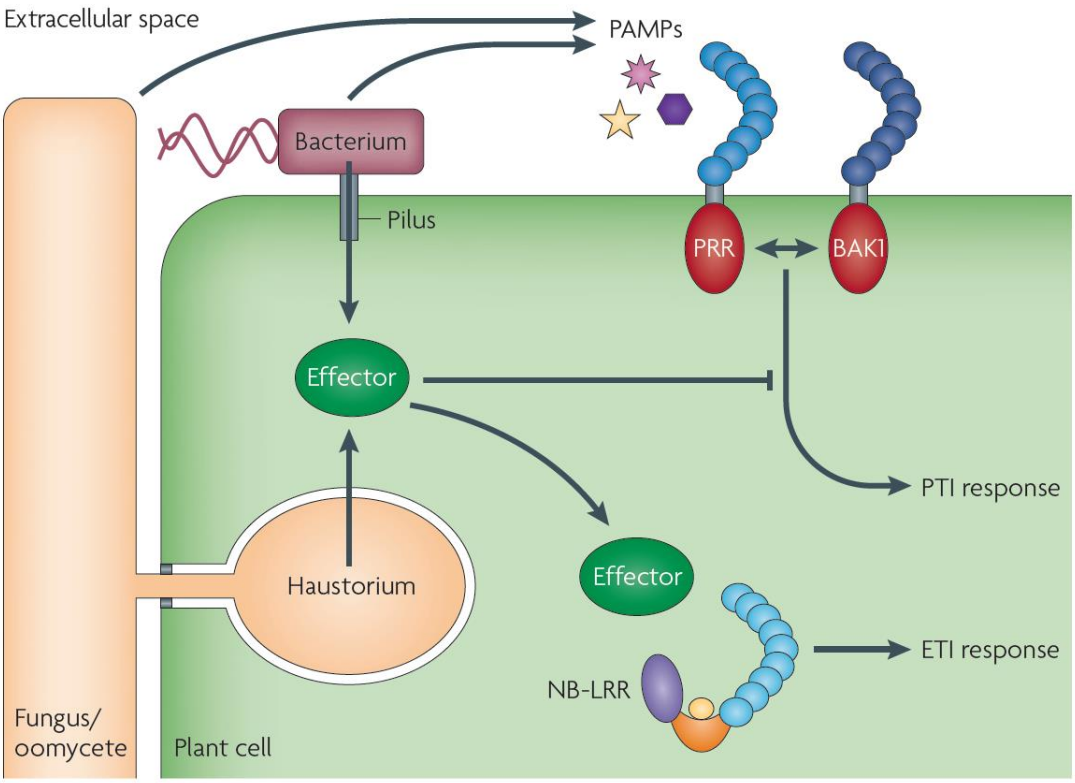


Figure 1. PTI and ETI type immunity strategies of the plants to detect pathogen invasion. *(Dodds & Rathjen, 2010).*

Main regulator of the PTI type plant immunity is BAK1 and it is the target of many pathogen effectors (Shan, et al., 2008). An LRR region is generally required for most of the recognized PRRs for their function (Chinchilla, et al., 2007) (Heese, et al., 2007) (see Figure 1) since they play an important role in recognizing other elicitors presumably by forming a heterodimer with PRRs (Dodds & Rathjen, 2010).

1.2.2 Effector Triggered Immunity (ETI)

Other strategy implemented by the plants for sensing the pathogen invasion is the recognition of the virulence molecules of the pathogen - that are proteins released by

pathogen organisms into the cytoplasm of the host cell to boost the infection - called effectors. Recognition of the effectors by the intracellular receptors of the plant host cell induces ETI. Diversification of the effectors secreted by pathogens and ETI receptors between and within the species is a norm and function of some PRR are conserved across families (Dodds & Rathjen, 2010).

Detection of a microbial attack to the plants by ETI mechanism is based upon the detection of the pathogen effectors secreted to the cytoplasmic environment of the host, by a class of sensory proteins having leucine rich repeats and nucleotide binding domains abbreviated as LRRs and NB respectively and these proteins are known as NB-LRR proteins (Jones & Dangl, 2006) (Chisholm, Coaker, Day, & Staskawicz, 2006). A widely distributed group of NB-LRR proteins also contains a coiled coil (CC) domain located at the N-terminal (Figure 1) (Dodds & Rathjen, 2010).

1.2.2.1 Direct and indirect recognition of the pathogen effectors

NB-LRR proteins grant plants resistance to various pathogen types containing oomycetes, viruses, bacteria, fungi, and insects. The NB-LRR multi domain play a role as a switch by translating various signals of the pathogen organisms into an immune response with its conserved form (Collier & Moffett, 2009). Many studies have revealed that the LRR domain is in charge of detection by intervening effector interaction (Ellis, Lawrence, Luck, & Dodds, 1999). In order to explain the mechanism of detecting pathogens, different models are proposed (Figure 2).

In direct recognition, immune signal was turned on by the effectors by means of physically binding of the effector to the receptor (Figure 2a). If there is no effector NB-LRR proteins kept in inactive form by intramolecular interactions until the introduction of the effectors to the environment. NB domain of the NB-LRR proteins

of the plants has a critical role in the proper functioning of the pathogen recognition through binding to nucleotides (Takken & Tameling, 2009). The process of turning on the signal might necessitate energy at the binding location (Tameling, 2006) .

Indirect recognition of an effector is postulated with either guard model or decoy model. Guard model assumes that a co-partner protein which might be a target point of virulence is transformed by the effector and then guard by NB-LRR protein. Recognition of the co-partner protein by NB-LRR can also be carried out by a constructional mimicking of such target which is proposed by decoy model (van der Hoorn & Kamoun, 2008) (Figure 2b). A variant of the decoy model named as bait model postulates a recognition mechanism comprised of two steps (Figure 2c). In the beginning, the effector interacts with the co-partner bait protein bound to the NB-LRR and later on NB-LRR protein directly recognises the effector protein leading to activation of the signalling instead of the recognition of the transformed target as postulated in guard model (Collier & Moffett, 2009).

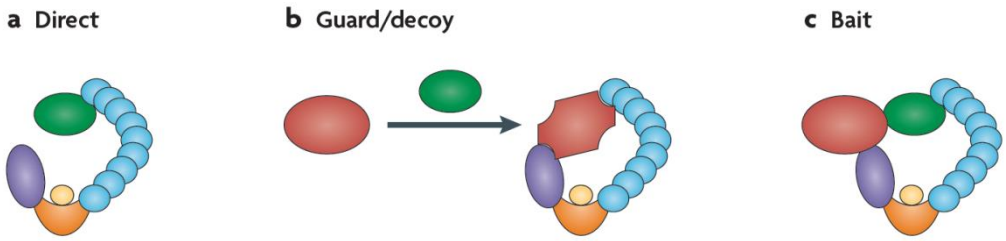


Figure 2. Recognition types of pathogen effectors by LRR and NB receptors of the plants. Effector: green; Receptors (purple, orange, yellow and blue) (Dodds & Rathjen, 2010)

LRR region of NB-LRR proteins are generally conserved, but which part of the LRR domain is not known. It is probable that direct and indirect recognition mechanisms

can drive dissimilar activation mechanism of NB-LRR proteins. Mentioned models are generalizations of some specific examples and they are beneficial conceptual tools to understand the recognition mechanism but they are not likely to describe all effector recognition events sufficiently (Dodds & Rathjen, 2010).

Innate immune responses of the animals also possess LRR and NB domains in their immune receptors (NLRs) that are take part in their PAMPs, in cell death protein 4 (CED4) and in the apoptotic factors of the animals (Ausubel, 2005) (Girardin, Philpott, & Lemaitre, 2003). NB-LRR proteins of the plants which possess an N-terminal TIR domain has a relation with the toll-like receptors of the animals which play a role in the intracellular signalling (Gay & Gangloff, 2007) .

1.2.3 Downstream responses of plant immunity

Important regulatory pathways of defence-gene expression are the salicylic acid and jasmonic acid – ethene pathways (Bari & Jones, 2009). Although both of the ETI and PTI type of immunities lead similar responses, generally PTI is effective against non-adapted pathogens (also known as known as non-host resistance) and ETI is effective against adapted pathogen organisms. Qualitatively ETI is more powerful and rapid than PTI and it generally causes local cell death at the infection sites which is known as hypersensitive response (Dodds & Rathjen, 2010).

Cellular events related with ETI and PTI are found as;

- reprogramming of gene expression,
- rapid inflow of calcium from exterior chambers,
- a flush of oxygen species such as H_2O_2,
- cell wall (callose) accumulation at the infection sites,
- initiation of mitogen-activated protein kinases (MAPKs),

and these cellular events generally leads to localized cell death (HR). ETI and PTI gene expression patterns mostly resemble to each other, supposing that the responses are the same in general but change in magnitude (Tao, 2003).

One of the mostly comprehended pathways is the kinase signalling (MAPK) pathway which is transfers signals from the outer surface receptors of the cell to the cellular responses. This signalling series is finalised as the expression of defence genes (Asai, 2002). One other pathway mostly understood is the ETI signalling. (Shirasu & Schulze-Lefert, 2003).

Bacterial pathogens targeting the plants proliferate solely in the external parts of the plant cells. Generally fungal pathogens elongate their invasive hyphae into the cytoplasmic region and generate a special feeding structure named as haustoria. Oomycetes also invade the plant cell in a similar fashion. Even though haustoria is generated from the host membrane they are not accepted as host intra cellular structure as they remain separated from the cytoplasm. Bacterial pathogens use type-3 secretion system (T3SS) pilus to distribute their effectors into the host cell mainly to suppress PTI in contrast to oomycetes and fungi which use haustoria to deliver effectors. However, several of them are detected by NB-LRR receptors found in cytoplasm and induce ETI (Dodds & Rathjen, 2010).

1.3 Importance of studying plant-pathogen interactions

In the history of humankind only 6 - 7 thousand of species have been cultivated out of 30 thousand edible plant species. Approximately 170 crops have commercial value which are worthwhile to grow in the fields in large scale and we the human being depend on only about 30 of these species to supply our daily nutrition and calories. More surprisingly 8 billion people in the world supply 40% of their calories

solely from three staples: rice, wheat and maize. In case of a global epidemic in one these staples, humankind would face a big scarcity. (FAO, 2018).

Food security is one of the biggest world issues on the rise which has also a global importance in the world economy because of the price spikes of some fundamental grains which has occurred in recent years. One of the main reasons of such price spikes are the intercontinental spread of plant diseases (Singh, 2008).

Such epidemics has promoted an enhanced interest on developing methods to protect crops by using resistance genes to take the plant diseases under control which has been applied for the last century by conventional methods (Flor, 1971).

These understandings suggest innovative biotechnological methods for protecting crops. In this manner focusing on the studies on plant–pathogen interactions can provide environment friendly and cost effective methods for crop protection and ensure food security.

1.4 Small angle X-ray scattering (SAXS)

SAXS is a structural analysis method that provides low resolution structural information of biological molecules and the complexes of these molecules in solutions and an established technique to study the kinetic processes of these molecules over different time ranges from μs to several hours. (Svergun, Koch, Timmins, & May, 2013). SAXS can be applied to an extensive range of MWs from a few kDa to the GDa range, which makes possible to study the structures and sizes of the proteins and complexes starting from peptides to large protein structures under many experimental environments changing from harsh conditions such as cryofrozen or high pressure to native environment (Kohn, 2004) (Wilkins, et al., 1999) (Stumpe & Grubmuller, 2007).

Specifically, SAXS is able to broadly characterize dynamic processes under variable experimental environments like pH, time, additives or pressure (Giehm, Svergun, Otzen, & Vestergaard, 2011) (Cordeiro, ve diğerleri, 2016) (Ryan, ve diğerleri, 2016) (Mojumdar, De March, Marino, & Onesti, 2017). But in general, it is used to characterize a pure monodisperse sample out of intermolecular interactions which is a necessity for the reconstruction of 3D models from SAXS data. No requirement of freezing and crystallization as in X-ray crystallography or other processes like chemical modification or labelling grants SAXS a general applicability. The method is especially convenient for studying disordered and flexible proteins which is difficult to study by other structural methods.

1.4.1 Working principle of SAXS

SAXS has a simple setup procedure in comparison with other structural experimental setups. Simply a solution of a macromolecular sample is placed in a quartz capillary tube and this capillary including the sample of interest is irradiated by a mono-chromatic X-ray beam. A concentration series of the sample is measured according to the solubility character and the MW of the particle which is usually from 0.1 or 0.5 to 5.0 or 10.0 mg/mL in order to account for probable concentration dependent attractive or repulsive intermolecular interactions (Mertens & Svergun., 2010). To prevent or reduce the radiation damage of the sample, flowing of the sample during data collection, solution additives like glycerol can be used (Jeffries, Graewert, Svergun, & C.I., 2015). Usually monodisperse solutions without aggregates are required for a meaningful analysis of the structure and the flexibility of the protein of interest. For a successful structural analysis, 95% or monodispersity is essential (Kikhney & Svergun, 2015).

A 2D SAXS detector is used to record the intensity of the scattered X-rays as a function of scattering vector (Figure 3). To obtain solely the signal of the

macromolecule of interest, the scattering pattern of the solvent is also measured and this pattern is subtracted from the pattern of the sample resulting the scattering profile of the macromolecule which is related to the complete size and low resolution structure of the macromolecule under examination (Mertens & Svergun., 2010) (Jacques & Trewhella, 2010).

The scattered radiation is found out as the function of scattering vector and defined with the equation $s = 4\pi\sin\theta/\lambda$, in which λ is the wavelength of the radiation and 2θ is the scattering angle (Figure 3). Pure signal coming from the macromolecules is obtained by deducting the solvent scattering from the solute scattering. The scattering profile of the macromolecule is isotropic and continuous as a result of spherical averaging, because the particles are tumbling freely in dilute solutions,

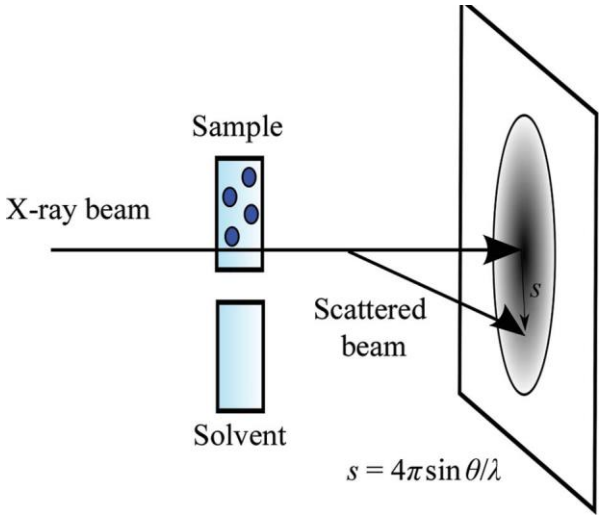


Figure 3. Simple schematic representation of a SAXS experiment. In SAXS analysis angular dependence of radiation scattered from the solution of a macromolecule is measured exposed to a monochromatic X-ray beam (Tuukkanen, Spilotros, & Svergun, 2017).

The intensity of the sample scattering has a relationship with the concentration of the sample and better s/n ratio can be observed if the concentration of the solvent-subtracted data is high. While at the same time together with the increasing

concentration the distances between individual particles of the sample come closer to the intra-particle distances which contributes to the scattering pattern (Franke, Kikhney, & D.I., 2012).

1.4.2 Information which can be provided by SAXS

From the experimental scattering pattern of the studied sample, numerous characteristic parameters can be obtained, including MW, Rg, excluded particle volume, folding state and maximum dimension D_{max}. SAXS can also determine flexible and unfolded proteins and structure of intrinsically disordered proteins - IDPs. SAXS results can be evaluated together with additional techniques such as X-ray crystallography, computational analysis of crystal packing, CD, FRET *in silico* sequence data, and especially for flexible systems, NMR which can complement SAXS analysis (Korasick & Tanner, 2018).

Notably, the SAXS results can be collected in publically available databases like SASBDB which is powered by EMBL, Hamburg (Valentini, Kikhney, Previtali, Jeffries, & Svergun, 2015) or Protein Ensemble Database (PED) for the collection of structural information on denatured protein and IDP based structures of NMR and SAXS data (Varadi et al., 2014). This creates a way for a more dynamic distribution of the results and for the usage of SAXS produced models and data by a wide-ranging biological society (Kikhney & Svergun, 2015)

The usual and the well-studied parameter mainly obtained from the SAXS data is Rg which supplies macromolecule's overall size. The R_g is defined as the average "root mean square" distanced to the center of density in the macromolecule weighted by the scattering length density. If two proteins with the same number of amino acids compared to each other; extended proteins have bigger R_g than the

smaller and compacter proteins. Guinier approximation (Guinier, 1939), states that for very small angles ($s < 1/R_g$) the intensity is depended only on two parameters:

$$I(s) = I(0) \exp(s^2 R_g^2/-3)$$

In practice, for a particle with any of a shape, the equation presented above shows that, the scattering intensity sketched as ln $I(s)$ vs s^2 will result in a linear curve and Rg can be calculated using the slope of this linear fit. The intercept shows the forward scattering I(0).

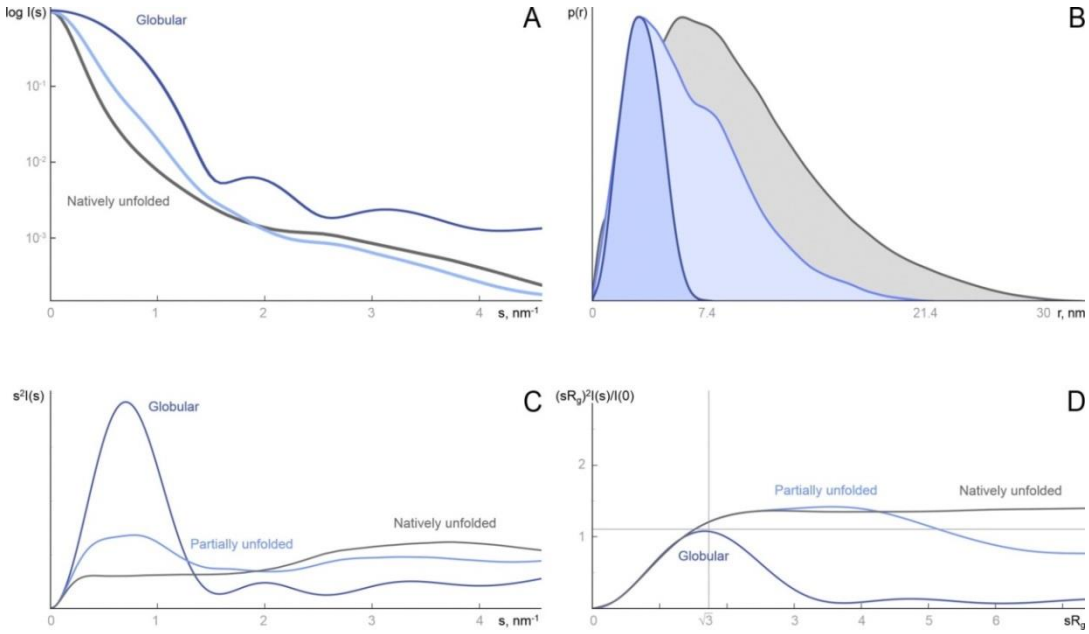


Figure 4. Model data simulation and comparison of three proteins with 60 kDa size.: Globular, 50% unfolded and fully disordered proteins depicted with dark blue, light blue, and grey colours respectively. A. Logarithmic graph of the scattering intensity I (s) vs. s., B. Distance distribution functions p(r) vs. r., C. Kratky plot $s^2I(s)$ vs. s., D. Normalized Kratky plot $(sRg)^2I(s)/I(0)$ (dimensionless) vs. sRg (Kikhney & Svergun, 2015).

1.4.3 Detection of protein flexibility by SAXS

Usually, the Kratky plot:

$$s^2I(s) \text{ (as a function of } s \text{ (Figure 4C))}$$

is used to identify the disordered state of a protein and to differentiate them from the globular ones. Using the Kratky plot it is also possible to observe specific properties of the scattering profiles such as the flexibility and folding state of the protein (Uversky & Dunker, 2010) (Bernado & M., 2009).

The scattering concentration from a solid body decays at high angles approximately as $I(s) \sim 1/s^4$ by exhibiting a bell-shaped Kratky plot. On the contrary, an ideal Gaussian chain has a $1/s^2$ asymptotic of $I(s)$ and exhibits a plateau at big s values. In case of an extended thin chain, the Kratky plot also exhibits a plateau over a specific variety of s, which is followed by a monotonic increment. Normally the second case is experimentally detected for unfolded proteins . In order to observe the relationship between the decline in the compactness of a protein and change in the Kratky plots, chemical or thermal unfolding experiments can be monitored by SAXS (Perez, Vachette, Russo, Desmadril, & Durand, 2001).

For the comparison of the folding states of dissimilar proteins, it is useful to normalize the data by taking $I(0) = 1$ and prior to drawing the plot by multiplying s with Rg. In this manner structural information is kept but the size information of the protein is dismissed. Kratky plot of these normalized data is named as dimensionless Kratky plot (Figure 4D).

In comparison with the globular proteins which have tightly packed core, unstructured proteins can be characterized by big average sizes due to the presence of extended conformations. To identify the unstructured nature of a protein experimentally measured R_g & D_{max} values can be compared with the predicted theoretical models (Kikhney & Svergun, 2015).

1.4.4 Structural reconstruction and modelling of proteins

SAXS enables to construct a low resolution model without any prior information. During the interpretation of the data of a monodisperse solution, pattern of the theoretical scattering which calculated from a single model has to fit the experimental data (Mertens & Svergun., 2010).

The method of rigid body modelling uses the domain scattering and subunit scattering to clarify including linkers and potentially flexible loops alternative modelling approaches would be required to explain the structure of disordered proteins. In the study of an intrinsically disordered or a flexible protein, layers of the models representing diverse states of a protein can be built and the scattering of the mixture of such models can be computed (Petoukhov & Svergun, 2005).

1.4.5 Relationship between radius-of-gyration & molecular-weight

Mylonas & Svergun studied the relationship between MW and Rg of several globular proteins, based on the light scattering and SAXS data (Mylonas & Svergun, 2007). For the globular proteins it is found out that the relevance between M_W–R_g is characterized by a power law with an exponent of 0.37 (Smilgies & Folta-Stogniew, 2015).

Figure 5 shows the relationship between the radius of gyration with molecular weight for many globular proteins. Light scattering data are shown in circles. The solid line demonstrates the fit to the data found out from light scattering studies, and the dashed lines shows the confidence range depending on the standard deviation of the residual. In the figure, SAXS data are shown with triangles. Diamond symbol depicts the light scattering data of the blood components for elongated and globular proteins. Within the limitations shown by the scatter in the

data, it can be detected whether a protein is non-globular if the aspect ratio is greater than $C/A > 1.25$. The method is sensitive for elongated protein structures however the effect is too small for flat structures. Combination of dynamic light scattering and static data can uncover if a protein in solution has a structure which is more complex than globular shape. (Smilgies & Folta-Stogniew, 2015).

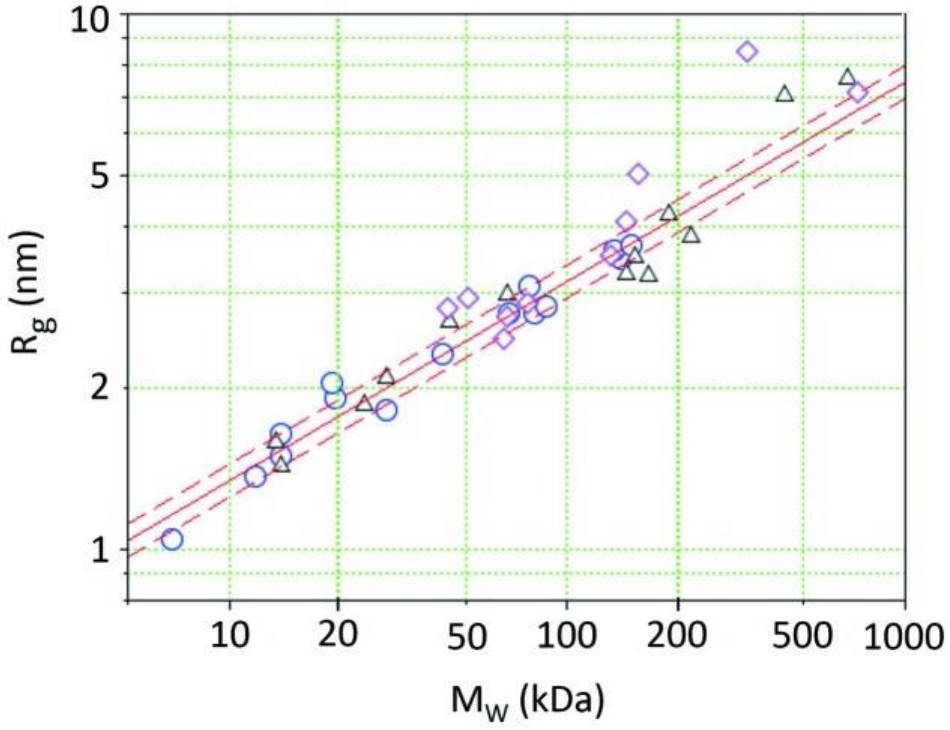


Figure 5. Radius of gyration versus molecular weight relation of several globular proteins. Solid line demonstrates the fit to the light scattering data; dashed lines shows the confidence range depending on the standard deviation of the residual. Circles: Light scattering data; Triangels: SAXS data; Diamonds: Light scattering data of the blood components for both globular and elongated proteins (Smilgies & Folta-Stogniew, 2015).

1.5 NMR spectroscopy of the biological molecules

NMR spectroscopy is not only a useful technique to study chemical molecules but it is also a powerful method for biologists to study dynamics, structure, and protein-protein and protein-ligand interactions of biological macromolecules. Using NMR spectral parameters, it is possible to obtain conformational data at the atomic level of the proteins. After determining the NMR spectra, distance or angular data can be used for computation of the structure a macromolecule. Using the same NMR spectral parameters, Molecular interactions of the proteins with another proteins, with a small cofactor or a RNA fragment can be projected (Marion, 2013).

NMR spectroscopy also enables scientists to study intrinsically disordered proteins which are unsuccessful to crystallize. By using RT-NMR and exchange spectroscopy or by doing relaxation measurements, NMR can even be used as a main method to study the protein dynamics. The main restriction of NMR is the size of the proteins which can be surpassed by studying the macromolecule partially (Marion, 2013).

As a complementary tool, solid-state NMR can be used for membrane bound macromolecules or large soluble multimeric proteins. The nature of crowded cellular environments on protein function and structure can be studied in-cell NMR spectroscopy. In these aspects NMR spectroscopy can provide complementary prospect in the identification of biological systems (Marion, 2013).

1.5.1 Macromolecular interaction studies by NMR spectroscopy

Protein–protein interactions have several important functions in several cellular processes. In case of the difficulties encountered during the co-crystallization of the complex with two partners because of some local disorder or low affinity, NMR can

complement these studies. The chemical shift perturbation (CSP) is a method based on the observing the spectrum of a molecule with ascending concentration of the partner. Heteronuclear Single Quantum Correlation (HSQC) spectrum of ^{15}N-^{1}H can be given as an example. Formation of the complex reveals that two molecules are in equilibrium (which is defined by the dissociation constant Kd) between their bound and free states. By the titration experiment; complex formation exchange rate and difference in the chemical shift between the bound and bound states can be observed. Perturbation analysis exhibits the amino acids which are located at the interface of the interaction (Chen Y, 1993) (Foster MP, 1998) (Marion, 2013).

1.5.2 Macromolecular structure studies by NMR spectroscopy

NMR Spectroscopy is frequently compared with X-ray crystallography for the purpose of determination the structural information of the biological molecules. As it is widely known that, structure of a biological molecule is studied by crystallography by the strike of the X-ray beams to the protein crystal which later on produce scattered beams. Diffraction pattern is measured from the scattered beams and converted into an e^- density map (which is an appearance of the protein) and used for the phase problem. By fitting the protein into the e^- density map, atomic model of the protein is obtained. But the growth of suitable crystals and the resolution of the phase problem (the electron-density map can be locally blurred as a result of local disorder) are the main bottlenecks of X-ray crystallography (Marion, 2013).

In contrast to crystallography, 3D structure of the protein derived by NMR spectroscopy is not an image of the genuine structure of the molecule but a model of that shape which is well-suited with the experimental data. The positional uncertainty in the molecular coordinates are given by the accuracy and the precision of the model (Chris A.E.M. Spronka, 2004).

Structural characterization and determination of a protein shape by NMR is carried out by using the resonance assignments. In order to do this, biggest possible number of spectral parameters like J-couplings, nOes, and residual dipolar couplings, have to be collected. Each of these information requires an amplitude and an assignment. Determination of a protein structure using NMR has also its own bottlenecks as in X-ray crystallography. Some of the distance restrictions may be hidden because of overlaps or during the interpretation of the data it can be misinterpreted (Marion, 2013).

1.5.3 Protein dynamics by NMR spectroscopy

During protein - protein interactions and enzymatic activities, fluctuations in the conformation of the protein can occur during protein folding, regulation as a function of time. The biological role of a protein is related to both dynamics and structure of the protein. NMR Spectroscopy is a sensitive method to study such fluctuations ranging from picoseconds to seconds. Resonance line-width effect and an array of NMR experiments would provide more data in details on the internal and global dynamic of the studied proteins. For example large proteins have broad signals but small molecules display narrow lines and fast internal fluctuations narrow the signals but slower motions in millisecond range act in the opposite direction. With a RT-NMR study, a series of NMR spectra can be recorded after starting the biological process like ligand binding, protein folding, pH changes, etc. using a rapid-mixing apparatus. Considering these applications NMR is the only experimental method to study protein dynamics at atomic resolution over a wide range of time scale (Marion, 2013) (Kleckner IR, 2011).

1.5.4 NMR study of intrinsically disordered proteins

In the beginning, NMR studies were carried out mostly on globular proteins which are composed of common secondary structure segments and the aim of the NMR study was determining the 3D shape of the proteins. However eukaryotic genome sequencing revealed that more than 30% of the proteins are composed of disordered regions like linkers loops or sequence termini and moreover these regions were performing important biological functions. Intrinsically disordered proteins (IDP) have an elevated content of Arg, Ala, Gln, Glu, Gly, Ser, Pro, and Lys residues in comparison with their globular counterparts and their intrinsic disorder character make them impossible, to crystallize. Therefore such studies must be subsidized either by NMR or SAXS) (Dunker AK, 2008) (Dyson HJ, 2004).

As a result of the flexibility property of the polypeptide chain of the IDPs, they display a more favourable line width in comparison with their counterparts; however they exhibit a restricted HN chemical shift dispersion. Also, a better resolution can be obtained by non-uniform sampling in all directions for a longer time period and secondly in the experiments with a higher dimensionality, the resonance can be extended in other dimensions. Except these major differences, the resonance assignment of the spectra of the IDP NMR pursues the same guidelines as for globular proteins. Hereafter it is possible to measure NMR parameters (since the both ensemble- and time-averaged resonances have been assigned) and using residual dipolar coupling, chemical shifts and nOe data, information at atomic resolution can be achieved (Pannetier N, 2007) (Narayanan RL, 2010) (Marion, 2013).

1.6 Comparison of SAXS and NMR in structural characterisation of biological molecules

SAXS and NMR both are well-known methods to study the structure and structural transitions of macromolecules in near native solution conditions. Both of the methods enable to study structural responses to chemical and physical changes or addition of ligands. NMR provides high resolution models at atomic level of the moderately sized macromolecules with its sensitivity to domain/subunit orientations and it is possibly the most robust method for the experimental analysis of the molecular dynamics. However SAXS method is not restricted by size with its sensitivity to domain/ subunit positions and used to reveal overall shape, flexibility and interactions of macromolecules. In tandem usage of these complementary methods provides a highly beneficial approach which enables the extensive characterization of bio-macromolecules. Joint application of SAXS and NMR are regarded as the most suitable tools for studying highly flexible and dynamic systems, which remain problematic studying with X-ray crystallography and modern electron microscopy (EM) (Haydyn & Dmitri, 2017).

1.7 Circular Dichroism Spectroscopy (CDS)

CDS is mainly used to study chiral molecules but it finds other important application areas such as secondary structure analysis and conformation of macromolecules. Determining the secondary structure of a new protein and evaluation of the folding and binding properties of this protein is one of the most important subject in structural genomics and proteomics studies. CDS is an exceptional technic which can be used to identify such properties of the proteins which are purified from tissues or expressed by recombinant techniques (Greenfield, 2006) .

1.7.1 Secondary structure analysis of macromolecules by CDS

Secondary structure of a protein is very sensitive to the medium of the protein therefore analysis of the structure changes of a protein with temperature or pH is required to determine in some cases. Advantage of the CDS can be taken in such situations to screen the alterations in the secondary structure of a protein with varying environmental conditions, mutations, denaturants or the binding interaction of a protein with other molecules. It is also possible to derive kinetic and thermodynamic information of the macromolecules from CDS spectroscopy. CDS also makes possible to determine if a newly expressed protein or a protein that is purified from a tissue is correctly folded, or to understand if a protein having a mutation has folded correctly in comparison with the wild-type, or can be used for the confirmation of the correctly folded active conformation of biopharmaceutical products in pharmaceutical industry (Greenfield, 2006) (Fishman, An introduction to circular dichroism spectroscopy, 2017).

As it is well known in chemistry, chiral molecules exist as pairs of mirror-image isomers which are non super-imposable known as enantiomers. Except the way that they interact with other chiral molecules and the way that they interact with polarised light, the physical and chemical features of a the enantiomers are alike. Most of the biological molecules are chiral molecules. For example, 19 of the 20 common amino acids except glycine are chiral. Since the proteins which are composed of these chiral aminoacids are host of other biologically important molecules, together with the proteins, DNA and RNA are chiral macromolecules. Notable chiral chemistry of biological molecules make them the main application field of the CDS technique. When a molecule includes one or more light-absorbing groups Circular Dichroism (CD) occurs, which are also known as chiral chromophores and it can be briefly defined as the variation in the absorption of right-handed circularly polarised light (R-CPL) and left-handed circularly polarised light

(L-CPL). During the absorption, the light is generally absorbed unequally by R-CPL and L-CPL. (Fishman, An introduction to circular dichroism spectroscopy, 2017).

1.7.2 Working principle of CDS

In CDS, CD of a molecule is measured over a range of wavelengths. Light beam has a time dependent electric and magnetic field and composed of two vectors of equal length at the x and y axis. The two circularly polarized waves are 90 degrees out of phase with each other and they can be separated by a broad array of prisms or electronic devices (Velluz, Legrand, & Grosjean, 1965).

In case of the interaction of asymmetric molecules with light, they absorb left and right handed circularly polarized light in variable amounts – circular dichroism is named after this phenomenon - and they refracts the two waves in different indices. As a result of the refraction of the light beam by the asymmetric molecule, the plane of the light wave is rotated and the vectors generates a new vector that traces out an ellipse which is expressed as the elliptical polarization of the light. It should be considered that CD spectrum of a DNA molecule or a protein is not the sum of the CD spectra of the individual residues or bases. However, the spectra are extensively influenced by the 3D structure of the macromolecule. Each three dimensional conformation has a unique CD motive. α-helix and β sheet secondary structural elements are the most extensively studied CD signatures of proteins (Fishman, An introduction to circular dichroism spectroscopy, 2017).

CD is reported in units of ΔE or in degrees ellipticity or the difference in absorbance of two vectors of equal length; ER and EL - one which rotates clockwise (ER) and the other one counter clockwise (EL) - by an asymmetric molecule (Fishman, An introduction to circular dichroism spectroscopy, 2017) .

1.7.3 CDS characteristics of structural shapes

Different structural shapes have their own characteristic CD spectra (Figure 6). These are basically;

- α-helical proteins have a positive band at 193 nm and negative bands at 222 nm and 208 nm (Holzwarth G, 1965).
- Proteins with well-defined antiparallel β-pleated sheets (β-helices) have positive bands at 195 nm and negative bands at 218 nm (Greenfield N., 1969),
- Disordered proteins have negative bands near 195 nm and they have very low ellipticity above 210 nm (Venyaminov S, 1993).

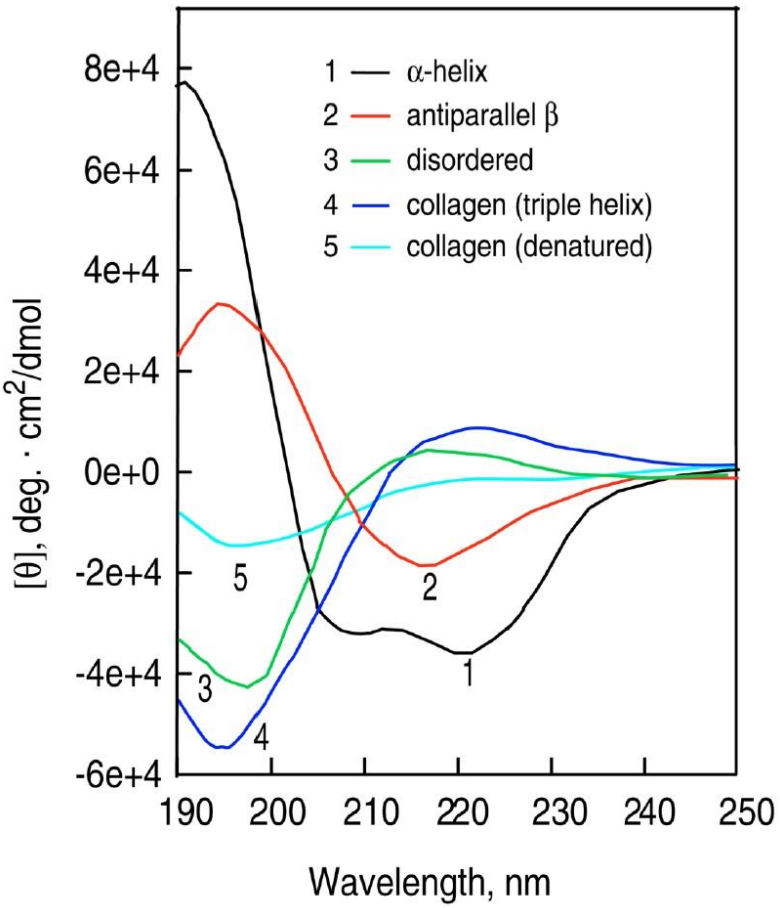


Figure 6. Characteristic CD spectra of some proteins and polypeptides and their characteristic secondary structures. CD spectra of poly-L-lysine: 1 (black) at pH: 11.1 in the α-helical conformation, 2 (red) at pH:11.1 in the antiparallel β-sheet conformations, 3 (green) at pH: 5.7 in the extended conformations (Greenfield N., 1969). CD-spectra-of placental collagen: 4 (blue) in its native triple-helical form, 5 (cyan) in denatured forms (Bentz H., 1978) (Greenfield, 2006).

1.7.4 Guide CDS of the varying conformations of the proteins

The spectra of some model proteins having extensively variable conformations, are presented in Figure 7. One advantage of CD is that; the data obtained from CD can easily be gathered and analysed in a short time using around 20 μg of sample protein solutions in aqueous buffers. But on the other hand it is not possible to detect the

secondary structure of specific residues by CD, as in X-ray crystallography and NMR structural determination studies.

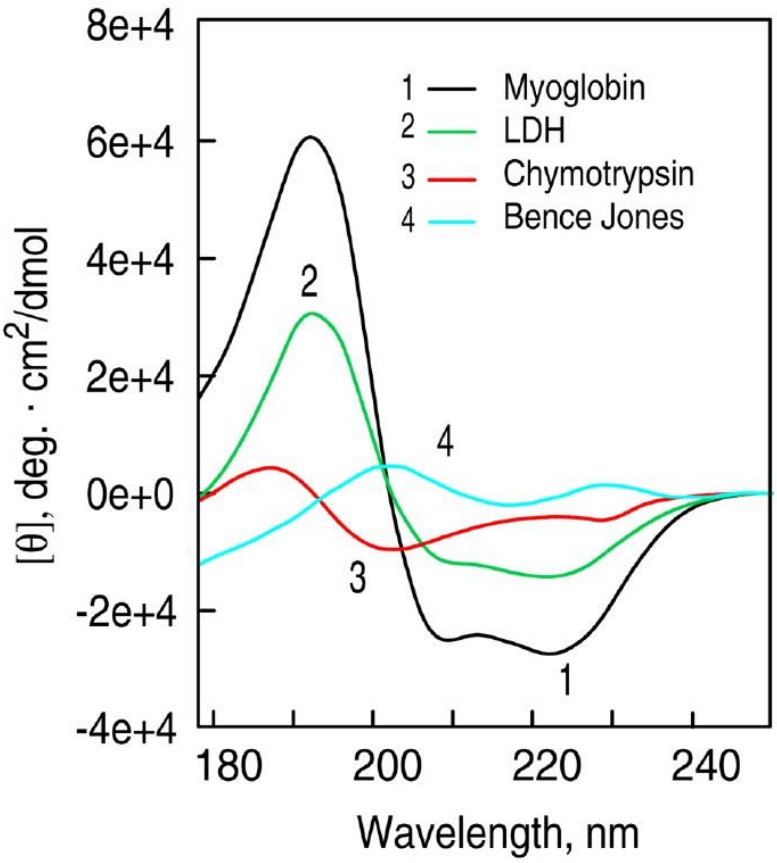


Figure 7. Model proteins having variable conformations and their CD spectra. (Greenfield, 2006). 1 sperm myoglobin of whale (black), 2 lactate dehydrogenase of chicken heart (green), 3 α-chymotrypsin of bovine (red), 4 Bence-Jones protein of human (cyan), REI light chain (light chain of κ type human immunoglobulin).

1.8 Differential Scanning Fluorimetry (DSF):

DSF is a fast and cost effective screening method which is mainly used for determining the Tm of a protein, screening its thermal stability and denaturation under various conditions or to identify the potential candidate low MW ligands

which bind and stabilize purified proteins (DeSantis & Reinking, 2016) (Niesen, Berglund, & Vedadi , 2007) (Senisterra & Finerty, 2009).

In DSF, a dielectric sensitive fluorescent dye that shows an affinity to the hydrophobic parts of the protein, is used and the unfolding temperature of a protein is identified by measuring the increment in the fluorescence of the dye as the hydrophobic parts of the protein is exposed with the raising temperature with a precision of 0.05 ^{o}C (Niesen, Berglund, & Vedadi , 2007) (Senisterra & Finerty, 2009).

DSF experiments can be performed by a RT-PCR instrument. To measure the melting point of a protein, purified protein in its buffer and dye is mixed and dispersed into the wells of the PCR plate using a micropipette. Intensity in the fluorescence is measured against the gradually raising temperature. The curve plotted fluorescence intensity against temperature shows the thermal degradation and Tm of the protein of interest (Niesen, Berglund, & Vedadi , 2007) (Senisterra & Finerty, 2009).

Solution conditions which affect protein stability and aggregation can also be studied by DSF, since specification of the optimum conditions and addition of small molecule compounds that stabilize proteins or removal of inhibitors that interfere with the purified proteins can simplify tough experimental researches such as protein purification and crystallization. From this aspect it might serve for pharmaceutics in drug discovery and drug formulation studies.

1.9 Stabilizing the conformation of proteins by ligand addition

Protein stability is influenced by varied buffer conditions like some additives, salt concentration or pH. Stabilization of the proteins by adding ligands into the mixture may extensively increase the amount of purified proteins and improve the quality of

them and also help the next step characterization processes such as crystallization (Vedadi et al, 2006).

Thermal stabilization of the proteins by introducing ligands such as; cofactors, substrates, metal ions, inhibitors, synthetic analogs of the natural ligands, and even other proteins, is a phenomenon which supplies enhanced protein thermal stability on binding (Schellman, 1997) (Privalov, 1979) (Brandts & Lin, 1990).

This phenomenon is referring to the protein unfolding and energetic pairing of ligand binding reactions. Such an energetic linkage leads to a ligand dependent conversion in the thermal stability of the complex of protein and ligand. There is a temperature range, for most of the proteins, where protein stability of is conversely related to temperature, meaning that the free energy of unfolding (DGu) decreases, as the temperature increases, reaching "0" at the equilibrium in which the concentrations of unfolded and folded protein are equal. At this equilibrium point, the temperature is considered as the Tm of the protein (Brandts & Lin, 1990). SYPRO Orange with its 465 nm excitation and 590 nm emission wavelengths has the most favorable properties to use in DSF studies, mainly because of its high signal to noise ratio (Niesen, Berglund, & Vedadi , 2007).

1.10 Protein-Metal interactions: Zinc binding assay

Zinc is one of the important trace metal which takes part in the proteins and it participates in many biological processes as an ion to carry out the functional roles. One of the primary and crucial steps during the characterization of a protein is the accurate determination of potential zinc binding sites in a protein. This finding would provide the understanding the functional role of zinc binding proteins. Taking the biological importance of zinc into consideration, several bioinformatics tools

have been developed such as ZincExplorer to predict the zinc-binding sites from a submitted protein sequences (Chen et al., 2013).

Using the ZincExplorer prediction tool, which incorporates the results of three different prediction tools, it is possible to correctly predict the potential zinc binding sites of a protein from the given aminoacid sequence of the protein. Four different types of zinc binding amino acids, HIS, CYS, GLU, and ASP can be predicted by the tool. Chen *et al.* stated that the co-dependent interactions of the predicted zinc binding sites, which are bound to the same zinc ion, could also be identified using the aforementioned bioinformatics tool (Chen et al., 2013). Using such bioinformatics tools to estimate the potential zinc binding properties and zinc binding sites of the protein of interest, before starting the wet lab experiments can be advantageous for time saving.

To determine the affinity of a metal or other type of ligands to a protein, DSF (see Part 1.6) can be used as a rapid and cost effective method. In the presence of a ligand, alteration in the thermal stability (or melting point) of a protein can be measured and suspected low-molecular-weight ligand or metal candidates can be determined using DSF. Transition midpoint that shows the change in the melting temperature of a protein in the presence and absence of a ligand demonstrates the binding affinity of the ligand. Since many proteins associated with metals serve several important biological functions, metal-protein interactions can be predicted first by bioinformatics tools and then confirmed by DSF (DeSantis & Reinking, 2016) (Niesen, Berglund, & Vedadi , 2007) (Senisterra & Finerty, 2009).

1.11 Challenges of studying protein purification and structural works

Each research area has its own difficulties faced during the implementation of the experiments. Cloning of a foreign gene into a microorganism, expression of this

gene in its native form and purifying the protein is always challenging because of the unique nature of the proteins. Even if a pure protein is obtained in the native structure, keeping this protein in its ideal condition throughout the research is a tough process.

Christendat *et al.* have shown that when using a single host vector and strain combination, fewer than 25% of recombinant proteins are produced in the cytoplasmic soluble fraction, and from these fractions only one part of fraction are purified up to higher purity levels. If the protein of interest to be overexpressed in *E. coli* belongs to a higher organism, given percentages are even fewer (Christendat et al., 2000) (Dobrovetsky et al., 2005).

Gene overexpression and protein purification researches are conducted generally on the new proteins which are not studied or expressed before, therefore there are not much information in the literature known on their physical and chemical properties. As a consequence, standard buffer conditions with variable salt and pH concentrations are used. However, complex properties of the proteins make their behaviour dependent on the media which means that the buffer chosen for the experimental studies and storage conditions can not necessarily be the ideal condition for purification and following structural studies such as crystallization (Senisterra & Finerty, 2009).

Before anything else, studies which aim for identifying the 3D structure of a protein firstly necessitate the presence of several other important components in the reaction mixture like chaperones or metals during expression, purification and subsequent processes in order to keep the protein in its correct folding conformation. Because of this situation, main purpose of a protein study (e.g: structural, functional studies etc.) principally require to carry out preliminary researches on the protein of interest itself and its optimum conditions (Copper, 2000).

CHAPTER 2

MATERIALS AND METHODS

2.1 PstSCR1 gene construct

PstSCR1 gene was made synthesized as in Dagvadorj, *et al.,* 2017 together with signal sequence and regained as pBSK/PstSCR1 (GeneScript). Plasmid information and sequence is presented in Appendix A.

2.2 Cloning of *PstSCR1* into pGEX-6P-1 vector

The gene was amplified by PCR using the following F and R primers respectively

PstSCR1_EcoRI_HindIII_F:CGGAAGCTTGAATTCTTCAAGTGTCCCGGTTTGC

PstSCR1_NoSS_XhoI_R: GGCCTCGAGCTAAGATGCTTTGGAGCAGTTG

to introduce *Eco*RI and *Xho*I enzyme sites for cloning into pGEX-6P-1 vector (GE Healthcare) presented in Figure 8 where GST-tag is located at the C-terminus of the PstSCR1, subsequently *E. coli* Top10 cells were transformed using the generated construct. A positive clone was selected after confirming by colony PCR and DNA sequencing. The *E. coli* cells of BL21 (DE3) (BNN93 *hflA150::Tn10* (Tet^R) (Studier, Rosenberg, Dunn, & Dubendorff, 1990)) were transformed by the plasmid isolated (pGEX-6P-1-PstSCR1) from the clone, which was validated by sequencing. Sequencing result was aligned with the sequence of the *PstSCR1* gene and it is

shown that cloning of *PstSCR1* into the vector is correct and in the frame (See Appendix B).

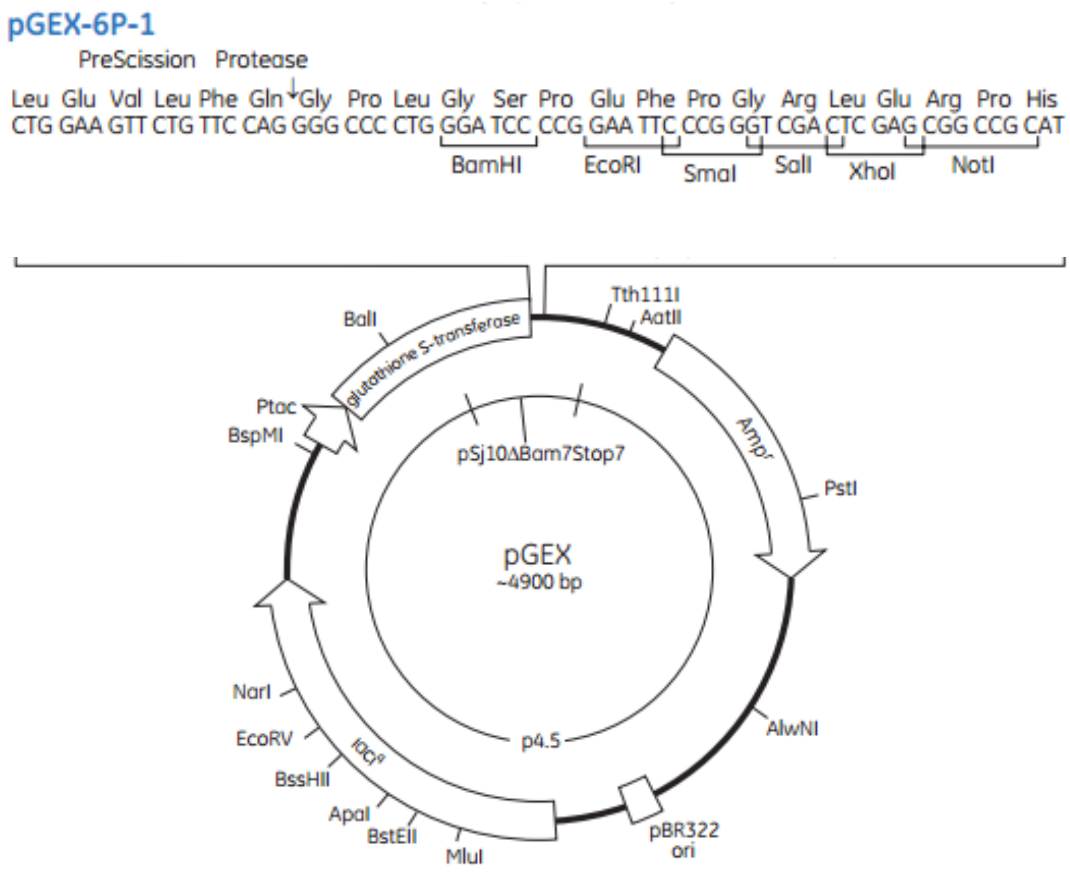


Figure 8. Plasmid map of pGEX-6P-1 (GE Healthcare). EcoRI and XhoI enzyme sites were added to PstSCR1gene by PCR and the insert was cloned into the MCS of the vector.

2.2.1 PCR amplification of the genes

PCR reaction ingredients and quantities are shown in Table 1 and 2. A template DNA was multiplied by either Taq DNA Polymerase (NEB) for standard reactions,

or Q5 DNA Polymerase (NEB) for clonings. To ensure the reproducibility of the reactions master mixes were prepared for multiple PCR reactions.

Table 1. *PCR components and quantities used in PCR reactions for Taq DNA polymerase.*

PCR Component	**Amount (µL)**
Template DNA	variable
10X PCR buffer	2.5
dNTPs (10mM)	0.5
F-primer (10 pmol/µL)	0.5
R-primer (10 pmol/µL)	0.5
Taq DNA polymerase (5 unit/µL) (NEB)	0.125
ddH_2O	variable
Total Volume	25

Thermo cycler conditions of Taq DNA polymerase are as follows: Initial denaturing at 95 °C for 30 seconds, 34 cycles: denaturing at 95 °C for 15 - 30 seconds, annealing at 57 °C for 20 - 60 seconds and extending at 68 °C for 1 min/kb, and final extension at 68 °C for 5 minutes. If the PCR reaction carried out overnight the mixture was kept at 4°C in the cycler.

Table 2. *PCR components and quantities used in PCR reactions for Q5 DNA polymerase*

PCR Component	**Amount (µL)**
Template DNA	1 (50 pg)
5X Q5 Reaction buffer	5
dNTPs (10mM)	0.5
F. primer (10 pmol/µL)	1.25
R. primer (10 pmol/µL)	1.25
Q5 DNA polymerase (2 unit/µL), (M0491, NEB)	0.25
5X Q5 High GC Enhancer	5
ddH_2O	variable
Total Volume	25

Thermo cycler conditions for Q5 DNA polymerase is as follows: Initial denaturing at 98 °C for 30 seconds, 34 cycles, denaturing at 98 °C for 5 - 10 seconds, annealing at 57 °C for 10 - 20 seconds and extending at 72 °C for 20 -30 sec/kb, and final extension at 72 °C for 2 minutes. If the PCR reaction carried out overnight the mixture was kept at 4°C in the cycler. The amplified PCR products were purified by QIAquick PCR Purification Kit and used for cloning.

2.2.2 Digestion of pGEX-6P-1 plasmid and amplified gene

Digestion of the amplified gene and pGEX-6P-1 vector to create sticky ends were carried out with *Eco*RI (NEB) and *Xho*I (NEB) restriction enzymes (RE). The digestion ingredients were prepared as presented in Table 3.

Table 3. *Reaction components and amounts of restriction enzyme digestion*

Ingredient	**Amount**
Insert DNA and Vector	500 ng
*Eco*RI (NEB)	0.25 µL
*Xho*I (NEB)	0.25 µL
Reaction Buffer 10X (NEB)	1.50 µL
ddH_2O	variable
Total	15 µL

Digestion reaction mix was incubated at 37 ^{o}C overnight. Digestion products were electrophoresed on 1% agarose gel for confirming the size and and extracting the DNA from the agarose gel.

2.2.3 DNA Purification

QIAprep^{R} Spin Miniprep Kit (Qiagen) was used for plasmid purification; QIAquick^{R} PCR purification Kit (Qiagen) was used for PCR purification and QIAquick^{R} Gel Extraction Kit (Qiagen) was used for extracting DNA from the agarose gel. For purification procedures manual of the manufacturer was used.

2.2.4 Ligation reaction of digested vector and insert

After linearization of the vector and digestion and purification of insert DNA, ligation reaction was performed by mixing the ingredients presented in Table 4.

Table 4. *Ingredients and quantities of the ligation reaction*

Ingredient	**Amount**
Vector	"5x" mol
Insert DNA	"x" mol
10X T4 ligase buffer	2 µL
T4 ligase	0.5 µL
ddH_2O	variable
Total	20 µL

The ligation reaction mixture was either incubated at RT for 15 min. or at 4 ^oC overnight. For transforming *E. coli*, 1 to 5 µL of incubated ligation mix was used.

2.3 Transformation of bacteria

In order to be able to amplify plasmids or overexpress the protein of interest in bacteria first of all *E. coli* competent cells were prepared and then transformed as explained in the following sections respectively.

2.3.1 Preparation of chemically competent *E. coli* cells

One *E. coli* Top10 colony was taken from the plate and inoculated in ~ 6 mL of LB medium without any antibiotics. Cells were grown in a shaker incubator at 37 ºC, 220 rpm ON. Next day 50 mL fresh LB medium in 100 mL flask was inoculated with 0.5 mL of ON bacterial culture. The culture was grown at 37 ºC, 220 rpm until OD_{600} absorbance was risen to ~0.4 (About 2-3 hours). Bacterial culture was placed

into a previously chilled sterile 50 mL falcon tube and palaced on ice for 10 minutes. The culture was centrifuged at 2500 rpm for 3 minutes at 4 ºC in the falcon tube and supernatant was decanted. Bacterial pellet was re-suspended in 10 mL cold 50 mM $CaCl_2$ solution by pipetting up and down or shaking gently. The tube was centrifuged at 2500 rpm for 3 min, at 4 ºC. Supernatant was decanted, and bacterial pellet was re-suspended again in 10 mL cold 50mM $CaCl_2$. The culture was placed on ice for 30 minutes. The competent cells were pelleted again carrying out the same steps explained above. After decanting the supernatant, bacterial pellet was resuspended in ~2 mL cold $CaCl_2$ solution. Fresh prepared *E. coli* Top10 competent cells were stored on ice or placed into the refrigerator until use.

2.3.2 Transformation of chemically competent *E. coli* cells

1 µL of a plasmid DNA or 5-10 µL of the ligation product was added to 50-100 µL of *E. coli* Top10 chemically competent cells in 1.5 mL tube. The tube was placed on ice and incubated for 20 minutes. The tube including the competent cell and DNA mixture was placed on a heat block at 42 °C to heat-shock the cells for ~ 45 seconds. The tube was placed on the ice immediately and incubated for 3-5 minutes. Then, 400 µL LB was put into the tube and incubated at 37 °C by shaking at 150 rpm for 1 to 2 hours. Transformed cells were spread on the LB agar plates with necessary antibiotics (Table 5). The plates were incubated at 37 °C and the cells were grown overnight. Several colonies were picked and used for colony PCR.

Table 5. *Plasmids, their antibiotic resistance and end concentration in the medium*

Vector	Antibiotic	Concentration (µM)
pGEX-6P-1	Ampicillin	100
pET28a+	Kanamycin	50
pET32a+	Ampicillin	100

2.4 Expression of PstSCR1-GST in *E. coli* to be used in SAXS studies

The cells in Ampicillin containing LB were grown at 37 ^{o}C for 16 hours by shaking at 180 rpm overnight as inoculate. 1 L Minimal Media (MM) (as in Marley et al., 2001) or LB were inoculated using 1 mL of the ON cell culture and grown at 37 ^{o}C by shaking at 220 rpm until OD_{600} is reached ~ 0.6. The cells were induced using 0.5 mM IPTG for 16 hours and harvested by centrifuging at 4 ^{o}C, 4000 rpm for 30 minutes.

2.5 Homogenization of the cells and purification of PstSCR1-GST by GST affinity column

Harvested cells were suspended in in the 20 mL Buffer 1 (20 mM Tris Base at pH: 8.0 including 300 mM NaCl and 10% Glycerol) and disrupted by passing through EmulsiFlex homogenizer three times at 10k psi. The protease inhibitor, PMSF to 1 mM concentration was added to the cell homogenate. Cell debris was pelleted by centrifuging at 10,000 rpm for 1 hour at 4 °C. (Cell debris was kept at -20 °C in case of inclusion body formation). The aqueous phase (50 mL) containing the PstSCR1-GST fusion protein was loaded onto 5 mL packed GST affinity column (Pierce™ GST Agarose). The recombinant protein was let bound to the colon material by

passing through the column at 2 mL/min rate and washed with Buffer 1 containing PMSF until no change observed in the absorbance at 280 nm of the flow-through. The protein bound to the GST-tag was eluted with Buffer 2 (Buffer 1 including 10 mM reduced Gluthatione). The samples of load, purification and washing steps were analysed on 12% SDS-PAGE for assessment of purification.

2.6 SDS-PAGE procedures

To analyze the proteins, fresh SDS-PAGE gels were prepared. The gels were prepared, load and run as explained in the following sections respectively.

2.6.1 Preparation of SDS-PAGE gel

In order to carry out SDS-PAGE analysis of either the crude cell extract or purified protein, following buffers were prepared: Separating-buffer (1.5-M Tris-HCl Buffer, at pH: 8.8), stacking buffer (0.5 M Tris-HCl Buffer, at pH: 6.8), 40% bisacrylamide – acrylamide solution (37.5 grams of acrylamide and 1 gram of bisacrylamide in 100 mL ddH_2O), 10% SDS, 10 % ammonium persulfate (APS), 1X SDS running-buffer (3.3 g Tris base including 1 g SDS detergent and 14.4 g glycine. All of them are dissolved in ddH_2O and total volume was adjusted to 1L). NNNN Tetramethylehylenediamine (TEMED), 5X Sample Buffer , loading dye (Thermoscientific), 1M-DTT and Precision Plus-Protein (Bio-Rad) was used for loading the gel., BioRAD electrophoresis system was used for SDS-PAGE experiments. For casting the gel the casting equipment of the manufacturer and for running the gels manual of the manufacturer was used. For separating gel, gel ingredients were mixed as explained in Table 6. APS and TEMED were added to the

mixture at last and the liquid gel mixture was poured into the cavity between the glasses immediately. Surface of the separating gel was-flattened using a few millilitres of isopropanol. The gel was let stand around 20 - 30 minutes for polymerization before the addition of stacking gel. Isopropanol was removed from the casting glasses and by mixing the gel ingredients as presented in Table 6 stacking gel was made ready. The stacking gel was poured onto the separating gel by using a pipette and the comb was placed immediately in the stacking gel. Prepared gel was rested around an hour for complete solidification.

Table 6. *SDS Separating and stacking gel contents*

Type of the gel and the gel content	**Amount**
Separating gel	
(4X) Separating buffer, 1.5 M Tris Base, pH: 8.8	2.5 mL
10 % SDS detergent (w/v)	100 µL
40 % Acrylamide (w/v)	3 mL
ddH_2O	4.35 mL
10 % APS (w/v)	50 µL
TEMED	6 µL
Stacking gel	
(4X) Stacking buffer 0.5 M Tris Base, pH: 8.8	1.25 mL
10 % SDS detergent (w/v)	40 µL
40 % Acrylamide (w/v)	0.45 mL
ddH_2O	3.25 mL
10 % APS (w/v)	20 µL
TEMED	8 µL

2.6.2 Loading and running SDS-PAGE gel

Prepared gels in the glass plates was put in BioRAD running chamber. Inner and outer chambers of the container was filled up with the running buffer (section 2.6.1) and the comb was gently removed out of the gel. The protein sample was mixed with 1X sample buffer and DTT (100mM) in a 50 µL PCR tube. The tube was placed on the heat block at 95 °C and incubated for 5 - 10 minutes. Protein samples

in the tubes were shortly spun down and loaded in the wells using a pipette and electrophoresis equipment was run at 100V until the protein samples were run into the separating gel border (approx. 15 - 20 minutes). Voltage of the electrophoresis equipment was increased to 220 V until the loading dye was hit to the bottom end of the separating gel (approx. 1.5 - 2 hours).

2.7 Western blot

After completion of the gel electrophoresis, the gel was taken out of the glass mould, rinsed in ddH_2O and for equilibration, incubated in Transfer Buffer (Thermoscientific) for 10-15 minutes by gently shaking. Four pieces of western blotting filter paper (Thermoscientific) and PVDF membrane (Thermoscientific, 0.2 μm pore size) were placed into the Transfer Buffer for 5 - 10 minutes for equilibrating the gel. The blotting was carried out using BioRAD Blotter at 1.3 A of constant current of for 15 minutes. The membrane was removed from the blotter and placed into the blocking buffer (3-5 % dried milk or BSA protein in TBS-T Buffer (Mixture of 0.137 M NaCl salt and 2.7 mM of KCl salt in 25mM Tris-HCl at pH:7.4 including 0.1% Tween 20) by shaking for 1- 1.5 hour at RT or in cold room overnight. Then, the PVDF membrane was placed into the primary antibody (AB) solution (primary AB was diluted at 1/4000 ratio in blocking buffer) and shaken gently overnight at 4 ^{o}C. Anti-His-Tag (His-Probe (H-3): sc-8036 Mouse Monoclonal IgG purchased from Santa Cruz Biotech. Inc.) and anti-GST (Santa Cruz) ABs were used in the experiments as primary ABs. After incubating the membrane in the primary AB, the PVDF membrane was washed 3 times using TBS-T for 5 minutes each time by gently shaking the membrane. Anti-mouse IgG, HRP-linked secondary antibody (Cell Signal. Tech.) was diluted at 1/10000 ratio in blocking buffer. And the PVDF membrane was placed in this solution and incubated at 4 ^{o}C for 1 hour. Then, the PVDF membrane was washed 3 times with TBS-T for

5 min each . The membrane was developped using Super-Signal West Pico Chemiluminescent Substrate (purchased from Thermo Sci.) according to the development protocol of the manufacturer. The PVDF membrane was sealed up in a plastic wrap and placed in a WB-cassette. The membrane was placed on the X-ray film (Thermo Sci.) and several films were exposed in a sealed dark room and placed directly into the film developer (Kodak). Several different exposure times were tried starting 1 second up to 30 minutes until desired colour was obtained.

2.8 Digestion of PstSCR1-GST, removal of GST-tag and ultra-purification of PstSCR1

To remove the GST-Tag from the PstSCR1 protein, eluted protein from GST column was incubated with 3C-His protease (1/100 (w/w) in 20 mM Tris base at pH:8.0 containing 20 mM beta-Mercaptoethanol (β-ME) at RT for 16 hours or at 4°C in the absence of 20 mM β-ME. (3C-His protease is expressed and purified as a recombinant protein in Sen Lab. at The University of Houston, TX, USA). The reaction was terminated by adding up 1mM PMSF.

Digested protein mixture was first concentrated using ultrafilter with a 3 kDa MWCO and 2 mL of the concentrated protein mixture was loaded directly on the size exclusion chromatography column (SEC) (Superdex-75, GE-Healthcare) which was equilibrated with PBS containing 20 mM β-ME, using Amersham Biosciences UPC-900 FPLC instrument at 0.5 mL/min. flow rate. The absorbance values of protein fractions were monitored at 280 nm wavelength and the fractions were collected per 500 µL volume.

3C-His-Protease which is used for proteolytic cleavage is expressed and purified in The University of Houston as a recombinant enzyme in *E. coli.* Since the protease contains histidine tag (His-tag) linked to the protease, it can easily be removed from

the digestion reaction medium by IMAC (Ni-Affinity Chromatography). Moreover if the size of the protein of interest (POI) is not close to the protease, IMAC step can be skipped and the protein mixture containing; POI, GST-tag and 3C-His protease can all be separated by Size Exclusion (Gel Filtration) Chromatography (SEC) and eluted into separate fractions. Generally 1/100 protease/protein ration is enough to chop the GST tag from the POI.

For truncation of the GST tag from the PstSCR1-GST recombinant protein, eluted protein was incubated at RT for 16 hours in the presence of the other reaction components shown in Table 7. (To keep the protein more stable, the reaction was carried out at 4 oC and without the presence of beta-Mercaptoethanol (β-ME), but the reaction could not be succeeded. Therefore, it was required to perform the reaction at RT with the addition of 20 mM β-ME).

Table 7. *3C-His Protease reaction components*

Component	**Concentration**
PstSCR1-GST recombinant protein	
3C-His Protease	(used at a ratio of 1/100 by weight)
Tris pH: 8.0	20 mM
beta-Mercaptoethanol (β-ME)	20 mM

2.9 Expression of PstSCR1-GST in *E. coli* to be used in NMR studies

E. coli BL21 competent cells, transformed with pGEX-6P-1/ PstSCR1 were grown in minimal media (MM). For 1D NMR studies, 5% D_2O was put into the protein solution (~400 µl) prior to the NMR measurement. Although this extra 5% D_2O leads to partial loss of resonances for amide protons, it is absolutely required to lock

the spectral frequcncy. For 2D NMR studies ^{15}N isotop labelling was carried out during the preperation of the MM by using ^{15}N labelled NH_4Cl and $(NH_4)_2SO_4$ (Marley, Lu, & Bracken, 2001).

Cell growth, induction of the gene, homogenization of the cells, GST purification of the Pst2a5, cleavage of GST tag by 3c-His protease, SEC purification of the protein and concentration of the protein with centrifugal ultrafilter (3kDa MWCO) were carried out as explained previously.

2.10 Preperation of minimal media (MM) and isotope labelling for NMR studies

For 1 liter of the 5X M9 salt solution; the reagents shown in Table 8 were dissolved in distilled water in a bottle, pH value was adjusted to 7.4 using NaOH. The bottle was filled with water up to 1L and autoclaved.

Table 8. *Composition of 5X M9 salt solution*

Component	**Amount (g)**
$Na_2HPO_4 \cdot 7H_2O$	64
OR	
Na_2HPO_4 (anhydrous)	56.8
KH_2PO_4	15
NaCl	2.5

Reagents shown in Table 9 were prepared, filter sterilized and used at the specified amount for the preparation of the minimal media.

Table 9. *Composition of minimal media*

Component	Amount
5X M9 salt solution	200 mL
1 M $MgSO_4$	2 mL
1 M $CaCl_2$	100 µl
100X Vitamin Solution	10 mL
Solution Q	4 mL
*NH_4Cl (0.1 g/mL)	10 mL
*$(NH_4)_2SO_4$ (0.1 g/mL)	10 mL
20% (w/v) D-Glucose	20 mL
Antibiotic (100 mg/mL)	1 mL
Distilled water	up to 1000 mL

* If the protein is to be used for 2D NMR studies, N^{15} isotope labelled chemicals were used.

Table 10. *Composition of 100 mL 100x vitamin solution (sterile filtered, kept chilled)*

Amount (g)	Component
0.01	riboflavin
0.01	niacinamides
0.01	pyridoxine monohydrate
0.01	thiamine

Table 11 shows the Solution Q components and their final concentration in the minimal media.

Table 11. *Composition of solution Q*

Micronutrient	Final Concentration (M)
$(NH_4)_6Mo_7O_{24}\cdot 4H_2O$	$3x10^{-9}$
H_3BO_3	$4x10^{-7}$
$CoCl_2 \cdot 6H_2O$	$3x10^{-8}$
$CuSO_4 \cdot 5H_2O$	$1x10^{-8}$
$MnCl_2 \cdot 4H_2O$	$8x10^{-8}$
$ZnSO_4 \cdot 7H_2O$	$1x10^{-8}$
$FeSO_4 \cdot 7H_2O$	$1x10^{-6}$

2.11 SAXS experimental methods

SAXS studies of the protein were carried out in National Synchrotron Radiation Research Center, NSRRC- SR- 23 A SAXS Beam Line. Three different concentrations of the PstSCR1 samples were prepared (%9, #7; %7 #9 and %5 #1) to study the protein. Experimental parameters of Taiwan NSRRC- SR- 23 A SAXS Beam Line is presented in Table 12.

Table 12. *Experimental Parameters of NSRRC- SR- 23 A SAXS Beam Line, Taiwan.*

Measurement Parameters	**10% (S#7)**
Photon energy (keV) and resolution	15, $2x10^{-4}$
Monochromator	DCM, Si(100)
Flux (0.5x0.6 mm^2)	10^9 foton/s
Scale factor	994.32
Data collection mode	multi
Sample and empty cell	0.64338
Sample cavity (Kapton inner cell) (mm)	2.784
Detector distance (mm)	2599.217
Pinhole Size (mm)	0.50x0.50
Total Scattering Data	1024 (along one axis)

2.12 NMR experimental methods

NMR experiments were carried out on a 800 MHz spectrometers TXI cryoprobe optimized for proton detection. The data acquired was converted from XWINNMR format to NMRVIEW format with NMRpipe and analyzed via NMRVIEW version 5.3 (Johnson, 2004) (Delaglio *et al.*, 1995).

2D $^{1}H/^{15}N$ HSQC spectra were acquired utilizing INEPT magnetization transfer (Morris and Freeman, 1979) and echo/anti-echo gradient selection. Spectra were acquired with 2,048 points (9615.385 Hz) in t2 dimension, and 256 complex points (2,594 Hz) in t1 dimension. States-TPPI quadrature detection (Marion *et al.*, 1989b) was performed using decoupling pulse sequence of GARP (Shaka *et al.*, 1985). The center of the nitrogen dimension was set to 118 ppm and number of scans of the spectrum was 228.

2.13 CDS experimental methods

~ 350 µL of SEC purified PstSCR1 protein in its buffer was placed in the cuvette and CD of the protein is measured between 200-250 nm wavelengths. Measurements were repeated in three different sets. ΔE values of each wavelength between 200-250 nm are plotted against wavelength.

2.14 Differential Scanning Fluorimetry (DSF)

Bio-Rad CFX Connect Real Time System was used for DSF measurements of the protein. Supplied custom plates and custom seal were used and the experiments were conducted according to the manual of the producer. Plate wells were filled with ~ 5 µM protein by pipetting with care to avoid bubbles. Sypro Orange dilution ration in the reaction mixture was 1:~2000. 25 µL recommended total well volume was not exceeded. A plate-spinner was used prior to the run. The data was processed by Bio-Rad CFX Manager 3.1 software. Processed data was exported in Microsoft Excel format and used for plotting the curve.

2.15 Zinc binding assay

In order to find out the binding property of PstSCR1 to zinc which was predicted by bioinformatics tools (see Section 3.10), zinc binding assay was carried out. Nickel content of 5 mL Ni-NTA chromatography cartridge was stripped off and the column was recharged with $ZnSO_4$. To carry out recharging process; Nickel content of 5 mL HisPur Ni-NTA chromatography cartridge (Thermo Fisher) was stripped off by stripping buffer (20 mM sodium phosphate buffer at pH 7.4, including 0.5 M sodium

chloride, 50 mM EDTA). Ni-NTA column was washed using 5 CV of binding buffer (20 mM sodium phosphate buffer at pH 7.4 including 0.5 M sodium chloride and 10 mM imidazole) and 5 CV of ddH_2O. The column was recharged with Zinc using ZnSO4 by 3.0 ml of 0.1 M $ZnSO_4$ in ddH_2O. After recharging process, the column was washed with 5 CV of ddH_2O and 5 CV of binding buffer to adjust the pH of the column and stored in 20% ethyl alcohol until use. To test the binding of PstSCR1 to zinc charged column, PstSCR1 was loaded on the column and washed with washing buffer and eluted with the stripping buffer in fractions. Analysis of the fractions were carried out by SDS-PAGE gel analysis.

2.16 TaYr10 gene construct

TaYr10 gene was made synthesized as pUK/TaYr10 (GeneScript). *Triticum aestivum* (wheat) stripe rust resistance protein Yr10 mRNA, complete cds information and sequence is presented in Appendix E.

2.17 Separation of TaYr10 protein into its domains

TaYr10 gene could not be expressed as a whole protein in pEt28a+ expression vector. Therefore it was decided to express the protein in different expression plasmids and additionally express it as independent domains by separating it into its domains. Since TaYr10 is composed of coiled coil domain (CC), nucleotide binding domain (NB) and leucine rich repeat domain (LRR), the whole protein sequence was separated into these domains to be expressed separately. NB domain was additionally separated into two sub-domains as N-terminal and C-terminal NB domains and named as N-NB and C-NB respectively. Independent domains with their initial and end residues are shown in Table 13, domain images predicted by

RaptorX (99% of residues modelled at >90% confidence) (Ma et al., 2013) (Källberg et al., 2012) (Peng & Xu, 2011a) (Peng & Xu, 2011b) is shown in Figure 9 and secondary structure prediction of TaYr10 predicted by Phyre2 (Kelley et al., 2015) (Wass, Kelley, & Sternberg, 2010) is shown in Appendix D.

Table 13. *Separated TaYr10 Domains to be expressed independently*

1- 512S-824R (LRR Domain)
2- 171G-512S (NB Domain)
3- 171G-824R (NB-LRR Domains together)
4- 171G-323G (N-NB Domain)
5- 323G-512S (C-NB Domain)
6- 1M-144S (CC Domain)
7- 1M-824R (Yr10 whole protein (CC-NB-LRR Domains))

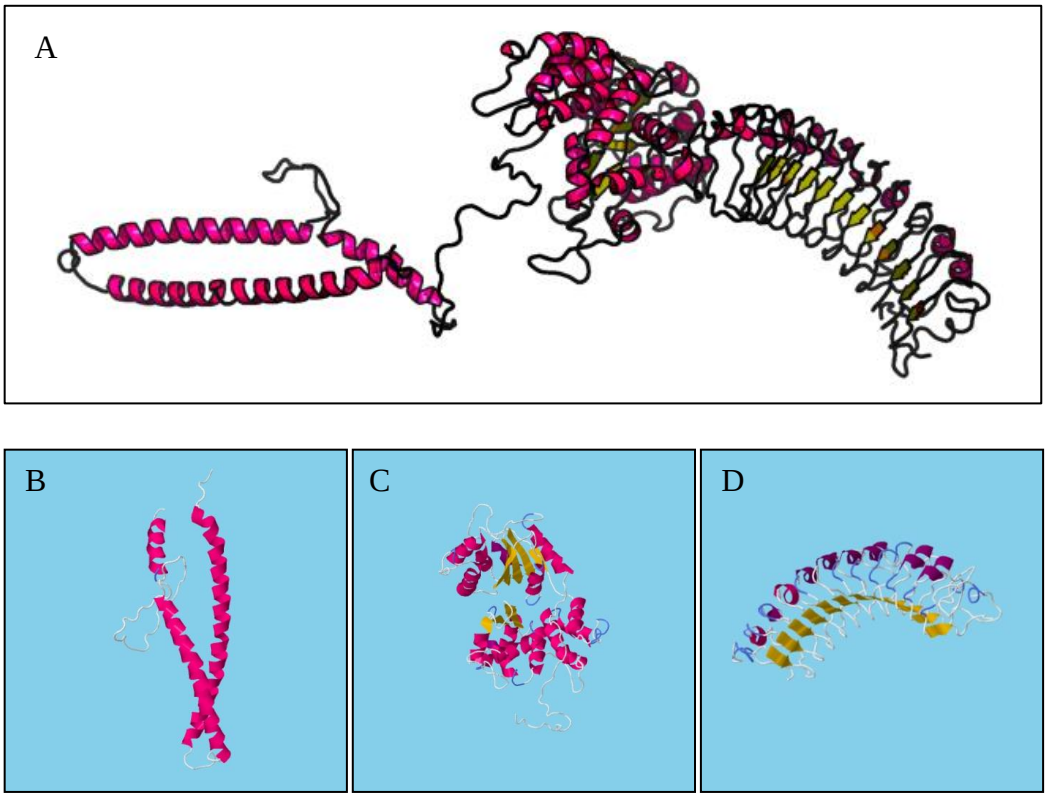


Figure 9. 3D structure-prediction of TaYr10 protein and-its domains by RaptorX. A. 1-139aa coiled coil domain (CC), B. 140-517aa nucleotide binding domain (NB) (including two sub-domains), C. 518-824aa leucine rich repeat domain (LRR).

Forward and reverse primers shown in Table 14 were designed and ordered to clone the relevant domain into different expression vectors. Since most of the vectors we have in the lab. have generally the *Eco*RI and *Xho*I sites, these restriction sites were decided to integrate into the F- and R- primers respectively. Additionally *Hind*III enzyme site was also integrated into the downstream of *Eco*RI forward primer in case of requirement.

Table 14. *Primers designed for cloning the respective TaYr10 domain into expression plasmid. All of the sequences are 5' to 3' direction. Enzyme sites are shown with colour codes.*

Yr10_512_EcoRI_HindIII_F:	GCCGAAGCTTGAATTCTCAAATGAGGAGAACTTTCT
Yr10_824_XhoI_R:	GGCCTCGAGTCAGCGTGGAGTTACCTTC
Yr10_171_EcoRI_HindIII_F:	GCCGAAGCTTGAATTCGGCACCAAAGAAAAGAGCCTTGAC
Yr10_512_XhoI_STOP_R:	GGCCTCGAGTCATGACAAGGAAGTGATAAGG
Yr10_323_XhoI_STOP_R:	GGCCTCGAGTCAACCACCAACTTCTTTGGCAACATC
Yr10_323_EcoRI_HindIII_F:	GCCGAAGCTTGAATTCGGTGTTTATCAGCTTAAACCTC
Yr10_1_EcoRI_HindIII_F:	GCCGAAGCTTGAATTCATGGAGGTCGTGACCGGG
Yr10_144_XhoI_STOP_R:	GGCCTCGAGTCAGCTATCAACCTTGTACCTTTC

2.18 Cloning of *TaYr10* domains into pET32a+ vector

The gene domains were amplified by PCR using F- and R- primers specified for each domain (Section 2.17). *Eco*RI and *Xho*I enzyme sites were introduced by PCR for cloning into pET32a+ vector (Figure 10) where S-tag, His-tag and Trx-tag (thioredoxin) are merged to the N-terminus of the selected *TaYr10* domains. The vector adds a 189 amino acid long peptide tags to the protein of interest which is approximately around 20 kDA. The vector encodes an enterokinase cleavage site which removes all of the tags and a thrombin cleavage site which removes only thioredoxin and His tag but keeping the S-tag bound to the protein of interest. S-tag is composed of 15 amino acids (KETAAAKFERQHMDS) which is derived from N terminus of RNase A. Subsequently *E. coli* Top10 competent cells were transformed using the recombinant construct. A positive clone was selected after confirming by colony PCR and DNA sequencing. The *E. coli* cells of BL21 (DE3) (BNN93 *hflA150::Tn10* (Tet^R) (Studier, Rosenberg, Dunn, & Dubendorff, 1990)) were transformed by the plasmid isolated (pET32a+/ selected TaYr10 domain) from the clone, which was validated by sequencing. Sequencing result was aligned with the sequence of the selected *TaYr10* domain (See Appendix E).

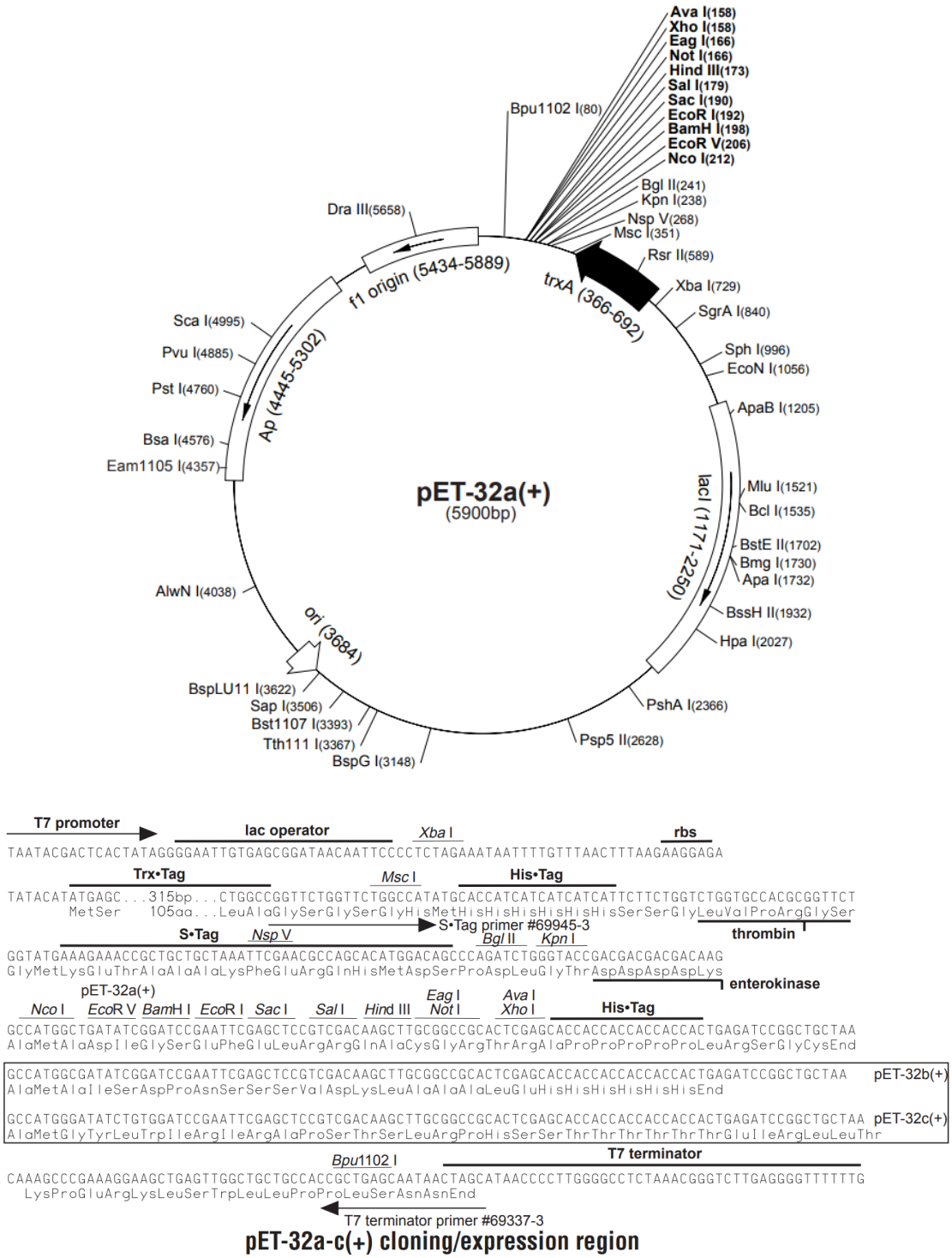


Figure 10. Plasmid map of pET32a+ (Novagen). EcoRI and XhoI enzyme sites were added to the selected TaYr10 domains by PCR and the DNA inserts were integrated into the MCS of the vector.

2.19 Limited proteolysis (partial digestion) of pET32 expressed N- and C-nucleotide binding (NB) domains of TaYr10.

Partial digestion of the proteins is generally used to find out the general properties of a protein such as the domain structure and looser parts of the proteins and to study the changes in their conformation after this process. It is also used to remove the tags bound to the recombinant proteins by finding out the required enzyme/protein (e/p) ratio, duration of the proteolysis and other required proteolysis conditions (Hubbard, 1998) (Fontana et al., 1997).

To find out the required e/p ratio and the digestion time for the digestion of the tag and remove from the protein of interest, a sequence of sample digestion experiments is carried out using a very low concentration of the protease. The protease will cut the protein of interest at its recognition sites in the whole protein starting from the exposed and flexible regions such as linkers or loops. These digestion series are analyzed on SDS gel to determine the cleavage products. The appearance of protein bands which have lower MWs than the protein of interest reveals the formation of the digestion process. In the SDS analysis of the partial digestion experiments, protein bands can disappear or new bands can appear and the band intensities can change due to the proteolysis process. All these changes such as temperature, dilution rates, e/p ratio and incubation time etc. can be monitored and these variables can be used for the further studies or for a big scale digestion process (Quevillon-Cheruel et al., 2007) (Moldoveanu et al., 2001).

The minimal e/p ratio and duration of the proteolysis observed on the SDS gel can be used to digest the related recombinant protein and get rid of the tag from the POI. Since proteases are expensive research materials used in the protein studies and each commercial expression vectors use their own proteolytic cleavage sites, it can be cost effective to apply limited proteolysis method instead of purchasing the related commercial protease to remove the tag (Quevillon-Cheruel et al., 2007).

CHAPTER 3

RESULTS AND DISCUSSION

3.1 PCR amplification of *PstSCR1* and cloning of *PstSCR1* into pGEX-6P-1 vector

PstSCR1 gene which was made synthesized as in Dagvadorj, et al., 2017 was re-amplified using primers as explained previously in section 2.2. and *Eco*RI and *Xho*I enzyme sites were introduced. Amplified gene product is shown below in Figure 11.

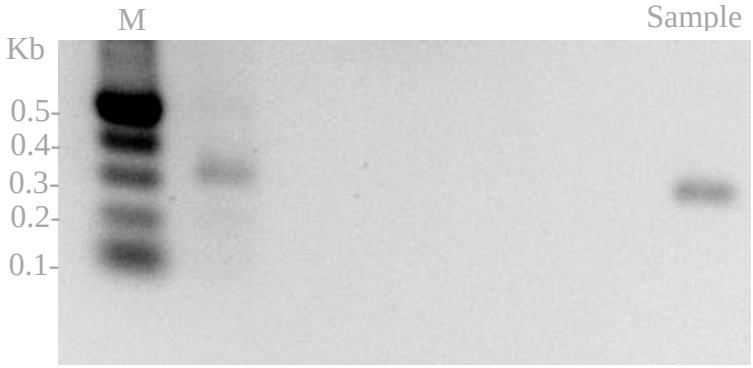


Figure 11. 0.8% agarose gel electrophoresis of PCR amplified PstSCR1 gene. M: 2-log DNA Ladder, NEB.

Cloning of the *PstSCR1* gene into pGEX-6P-1 vector was performed as explained previously in section 2.2 and pGEX-6P-1/PstSCR1 plasmid was obtained. This plasmid was used for the expression of the *PstSCR1* gene.

3.2 SDS PAGE analysis of the GST-tagged PstSCR1 protein

Recombinant proteins having a size of about 36.2 kDa, purified from the cells grown in both media are shown in Figure 12. In all cases, 26 kDa GST protein bands were also observed which was either expressed independently from the recombinant protein or broke off the recombinant protein. The amount of purified protein obtained from the cells grown in MM medium was found to be higher than the ones grown in LB medium.

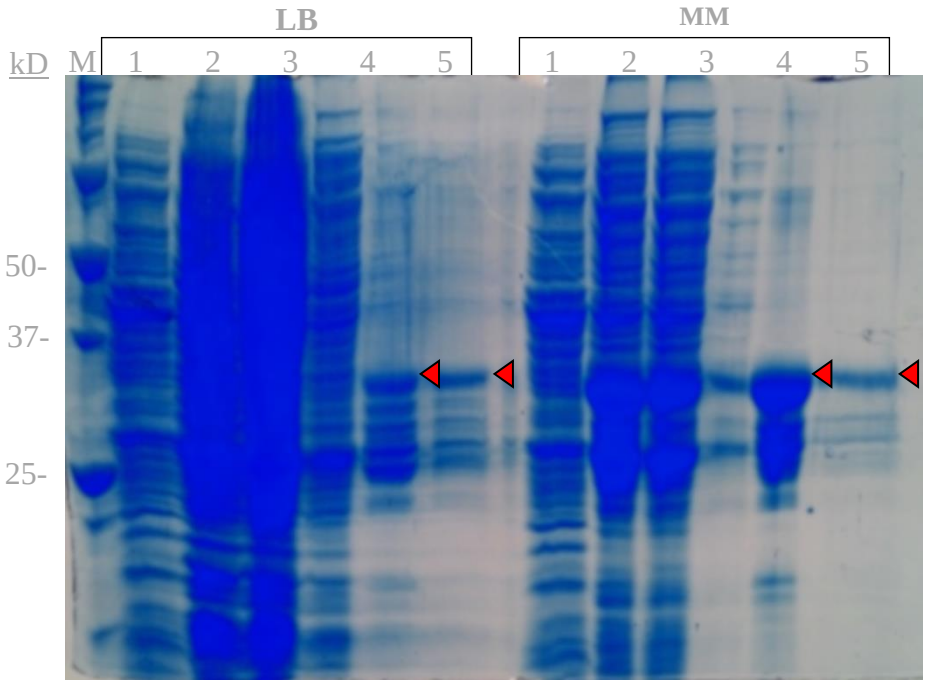


Figure 12. 12% SDS PAGE analysis of the GST-tagged PstSCR1 proteins in different purification steps. PstSCR1/pGEX-6p-1 plasmid was transferred into BL21 (DE3) competent cells and GST-tagged PstSCR1 was expressed in BL21 (DE3) cells by growing in LB and MM media induced with 0.5 mM IPTG for 16 hr at 37 ^{o}C. Lanes: 1. Un-induced cell extract, 2. Induced cell extract, 3. Flow through, 4. Washing fraction, 5. and 6. Elution fractions.

Since the expression vector fuses a GST tag protein of about 26 kDa to the protein of interest, the PstSCR1 protein which has a theoretical size of about 10.2 kDa became a recombinant protein with an expected total size of about 36.2 kDa. Supernatant of the homogenized cell was loaded into the GST column containing 5 mL column material for purification of the recombinant protein. Fractions of the purification steps were loaded onto the 12% SDS-PAGE gel and samples were run on the gel for ~2 hours at 120 volts. The gel was incubated in coomassie blue solution to stain the proteins overnight and destained by washing in hot water until obtaining clear bands. As it can be seen in SDS PAGE analysis, expressed protein is not pure enough to use for structural studies. Therefore additional purification methods have to be applied. Before proceeding to the next purification step it is decided to remove the GST tag which is linked to the PstSCR1 protein.

3.3 Western blot analysis results of the His-tagged PstSCR1 protein

Western blot analysis of the Ni-NTA purified PstSCR1 expressed in pET28a+ expression plasmid is carried out as explained in section 2.7. Membrane image is shown in Figure 13.

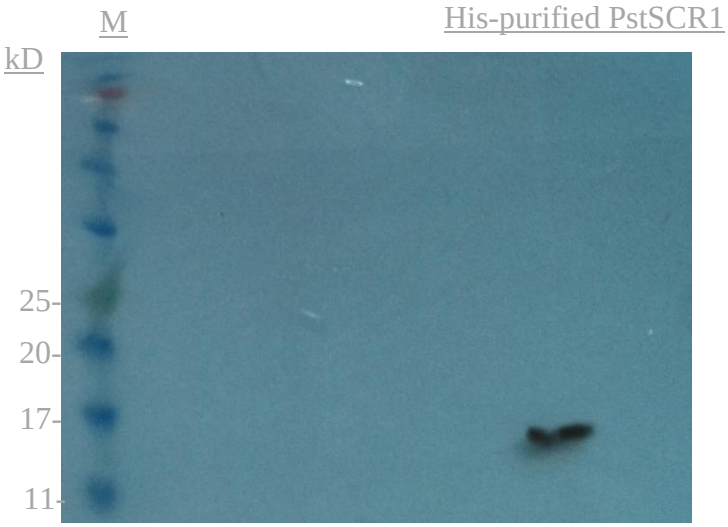


Figure 13. Western Blot analysis of Ni-NTA purified PstSCR1 protein expressed in pET28a+ expression plasmid using His-Probe (H-3): sc-8036 Mouse Monoclonal IgG 1/4000 as a primary AB (Santa Cruz Biotech. Inc.) and Anti - mouse IgG, HRP - linked secondary AB 1/10000 (Cell Signaling Technologies). (Developed with: Super Signal West Pico Chemilumoinescent Substrate from Thermo Scientific) 30 min exposure. M: Marker; Precision Plus Protein of Bio-Rad Inc.)

Since the expression level of the protein using pET28a+ expression plasmid was very low and only detectable by WB, different expression plasmids such as pGEX-6P-1 was tried to boost up the overexpression level of PstSCR1 to reach the sufficient amount in order to further characterize the protein.

3.4 Size Exclusion Chromatography (SEC) purification of PstSCR1 protein

Purification of PstSCR1 expressed in pGEX-6P-1 was carried out by Size Exclusion Chromatography after digesting the protein with 3C-His protease in order to remove GST tag. SEC chromatogram of the purification of PstSCR1 is presented in Figure 14. According to the manufacturer manual of the column (S75-GE) and the experiments carried out by protein standards, the proteins with the sizes of around 10

kDa is expected to leave the column before 12 mL after the sample injection. Therefore the peak at 11.76 mL is expected to belong to the PstSCR1 protein.

GST tag has a 26 kDa size and it is well known that it forms a dimer after cleavage which will have a 52 kDa size during the purification by SEC. Since the larger molecules leave the column earlier, the peak at 9.06 mL should belong to GST tag.

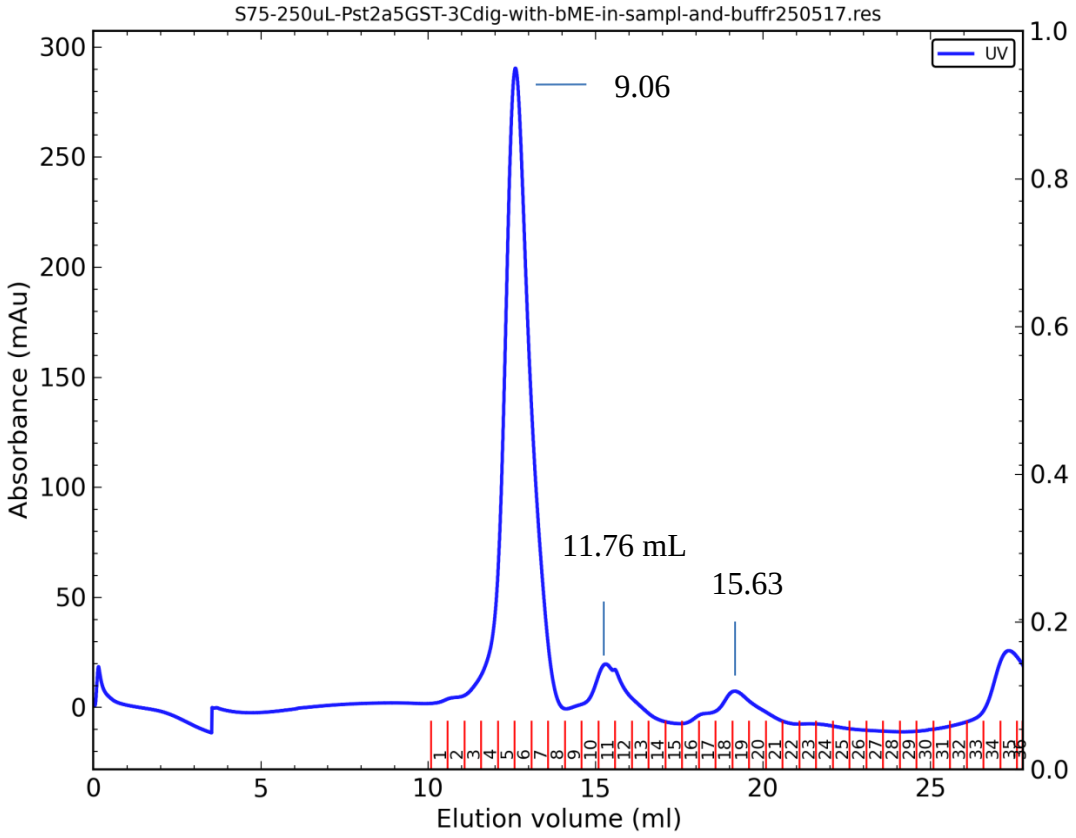


Figure 14. SEC chromatogram of the purification of PstSCR1. The peak at 11.76 mL belongs to the PstSCR1 protein. The peak at 9.06 mL belongs to GST tag which has a bigger MW than PstSCR1 and therefore eluted from the column

The 3C-His protease with its 22 kDa size, was expected to be separated by a distinct peak from the PstSCR1 protein which has a 10 kDa size however, elution of PstSCR1 protein and 3C-His protease together as an almost single peak gives the

impression that the PstSCR1 protein was actually dimerized and eluted from the colony as a dimerized 20 kDa size protein which gave rise to unification of the 20 and 22 kDa peaks of 3C-His protease and dimerized PstSCR1 protein respectively. But since protease/protein ratio was only 1/100 and the fractions were collected in 500 µL volumes, this could cause a small impurity in the former fractions but it was estimated that as the fraction number increases, the impurities could be reduced and purer PstSCR1 samples would be obtained in the latter elution fractions of PstSCR1.

In order to confirm this argument, the fractions eluted from the SEC column were collected in 500 µL volumes and 20 µL samples were loaded on 12% SDS-PAGE gel and run on the gel for ~2 hours at 120 volts. The gel was stained in the coomassie solution overnight and de-stained by washing in warm water until obtaining clear bands. SDS-PAGE gel analysis of the SEC elution fractions is presented in Figure 15.

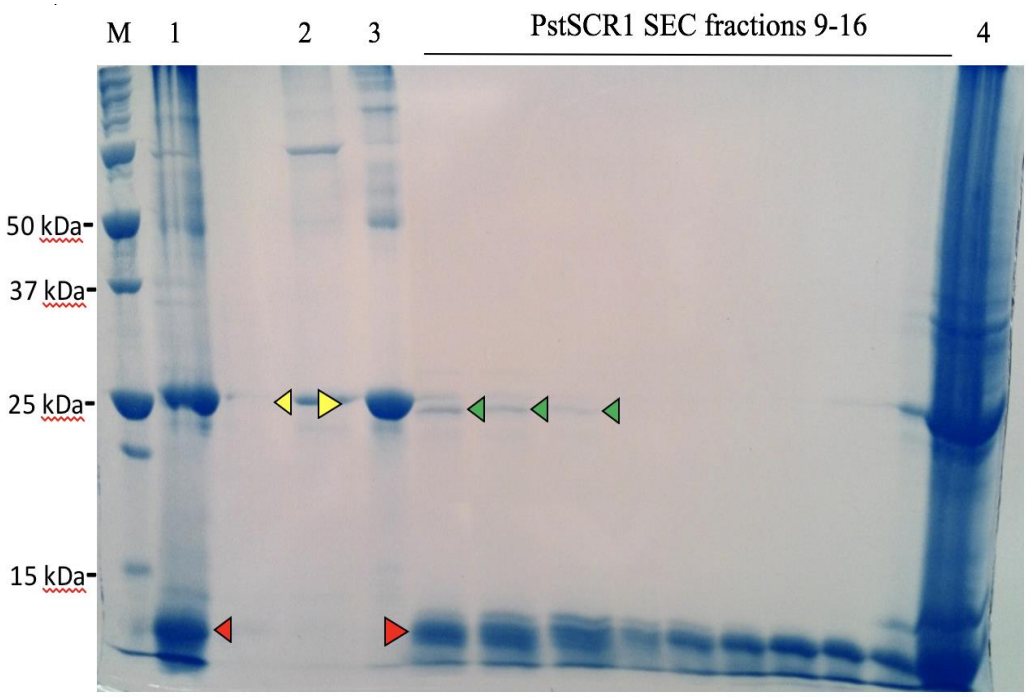


Figure 15. 12% SDS-PAGE gel analysis of PstSCR1 SEC fractions. The bands indicated by red arrow are the PstSCR1 protein bands (~10 kDa). The bands indicated by the yellow arrow are the GST bands (~26 kDa) and the weaker bands indicated by green arrow are the 3C-His protease bands (~22 kDa). Lane numbers: 1. recombinant protein mixture digested with 3C-His protease, 2. 2nd fraction shown in the chromatogram, 3. 6th fraction shown in the chromatogram, PstSCR1 SEC fractions 9-16. elution fractions of PstSCR1 (9 to 16 shown in the chromatogram), 4. Proteins that are precipitated during concentration.

12% SDS PAGE analysis revealed that; digested recombinant protein mixture with 3C-His protease have mainly two thick bands (lane number 1) at around 26 kDa and 10 kDa which belong to the cleaved GST and PstSCR1 protein respectively. This analysis shows that a complete digestion was obtained. Lane number 2 and 3 belong to the SEC fractions 2 and 6 which are taken from the ~7mL and ~9 mL elution fractions of the GST peak. Since there is mainly only one peak at 26 kDa, it can be suggested that these fractions are composed of mainly GST and small amount of other larger size impurities. Examining the SEC fractions of PstSCR1 peak eluted between 11-14 mL (fractions 9-16) reveals that there are mainly strong PstSCR1 protein bands at 10 kDa together with weak 3C-His protease bands at around 22 kDa

only in preceding fractions (lane numbers 9,10,11). But in the following fractions (12-16) pure PstSCR1 protein bands at 10 kDa are observed without any impurities which can be used for further characterization experiments as a pure protein sample.

On the last lane of the 12% SDS PAGE gel (lane number 4), white precipitates obtained during the ultrafiltration of the protein was loaded on the gel to find out the composition of the precipitate. In this lane mainly two strong bands are observed at around 26 kDa and 10 kDa which belongs to GST and PstSCR1 protein respectively suggesting that significant amount of PstSCR1 protein was lost together with GST tag during the concentration process which reduced the amount of pure protein and yield.

It was required to further increase the overexpression yield of PstSCR1 protein either by optimizing the expression conditions or reducing the loss of protein during the purification or concentration process so that the required quantity of the protein for studying the crystal structure could be achieved and used for X-ray crystallography studies. Unfortunately the yield of the protein was not sufficient for crystallography studies. Therefore it was decided to continue characterization of the protein with NMR and SAXS analysis.

3.5 SAXS analysis results of PstSCR1 protein

Structural characterization of the Pst effector protein PstSCR1 was performed by SAXS. 3D structure of the *E. coli* expressed pure recombinant PstSCR1 was identified by SAXS in solution and presented in different angles in Figure 16. According to the data that it was found out that PstSCR1 has a 23.1 Å radius of gyration and the nano formations are arranged in the crystal configuration which has a distance of 39.3 Å.

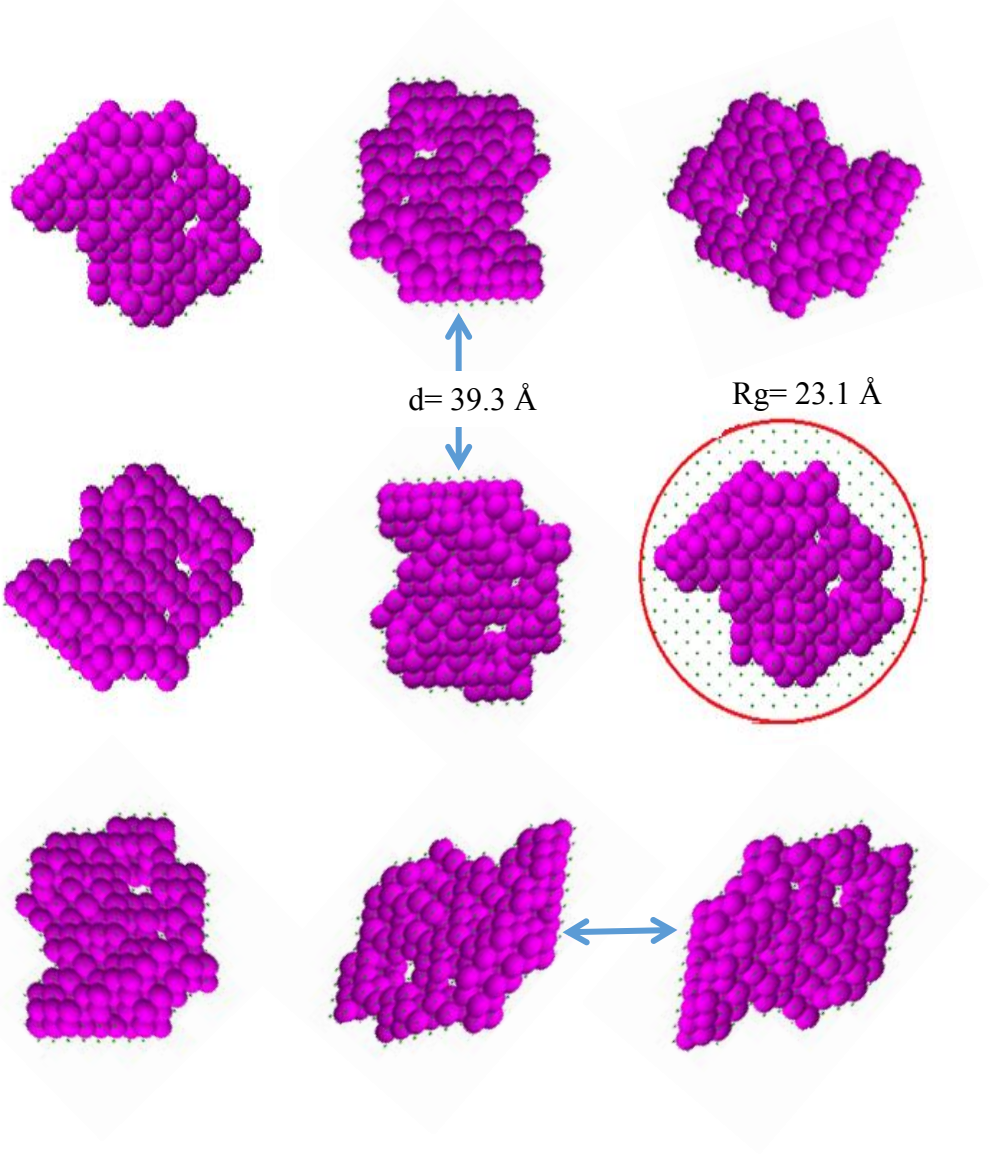


Figure 16. 3D structure of pure recombinant PstSCR1 in different angles identified by SAXS. PstSCR1 has a 23.1 Å radius of gyration and arranged nano-formation distance in the crystal configuration is 39.3 Å.

There is also a correlation between MW and Rg of several globular proteins based on the light scattering and SAXS data according to the study of (Mylonas & Svergun, 2007). The relevance between MW–Rg is represented by a power law with

an exponent of 0.37 (Smilgies & Folta-Stogniew, 2015). According to these data MW of PstSCR1 was found as ~20 kDa.

2D Charge Coupled Device (CCD) detector; Synchrotron Radiation (SR) SAXS profile; 1D data obtained from CCD; comparison of the SAXS profiles of the samples in three different concentrations; fit curve and Pair Distribution Function vs. Distance (Å) curve are presented in Appendix C.

3.6 1D-NMR analysis result of PstSCR1 protein

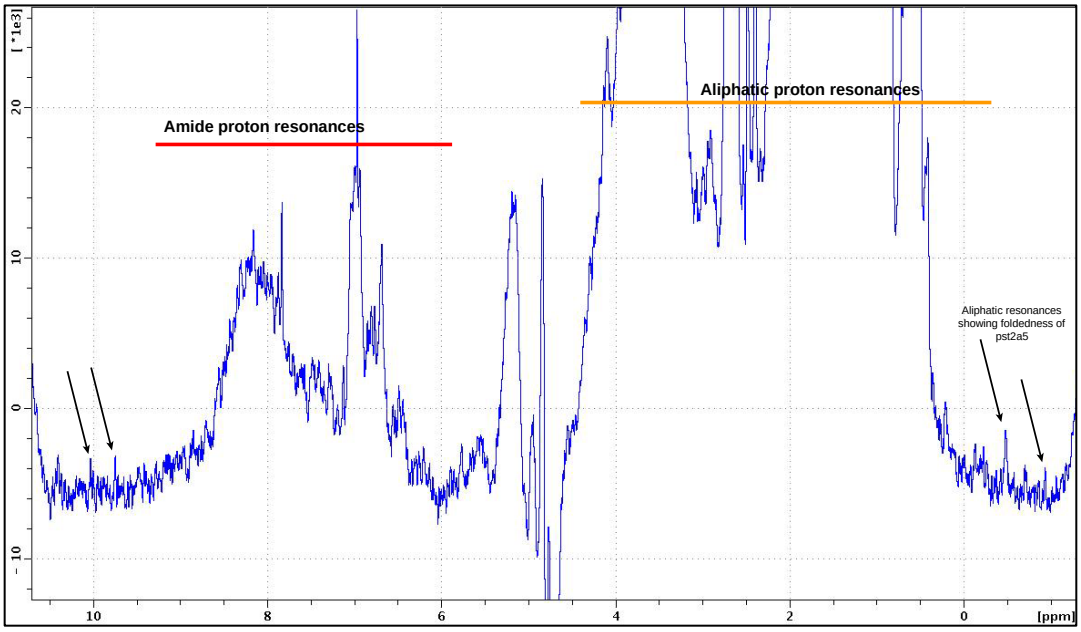


Figure 17. 1D NMR water suppression spectra of PstSCR1 protein.

1-Dimensional (1D) water suppression NMR experiment is a benchmark for checking if the recombinantly expressed proteins are folded and globular. Here, PstSCR1 clearly showed unclustered amide proton resonances (Figure 17). For

example, tyrtophan nε1 proton due to resonance moves significantly downfield in the NMR timescale in the folded protein structure. Additionally, well-dispersed peaks of amide proton resonance from 6 ppm to 9 ppm is a good indication of foldedness of PstSCR1. Aliphatic residues, Valine, Isoleucine, Leucine when packed against polar residue or water molecules moves significantly upfield when protein is folded. In our 1D-water suppression spectra indeed contains more than two resonance despite finding such resonance is rare. In short, PstSCR1 recombinantly expressed in pGEX-6P-1 system, is well-folded and globular.

3.7 2D NMR result of PstSCR1 protein

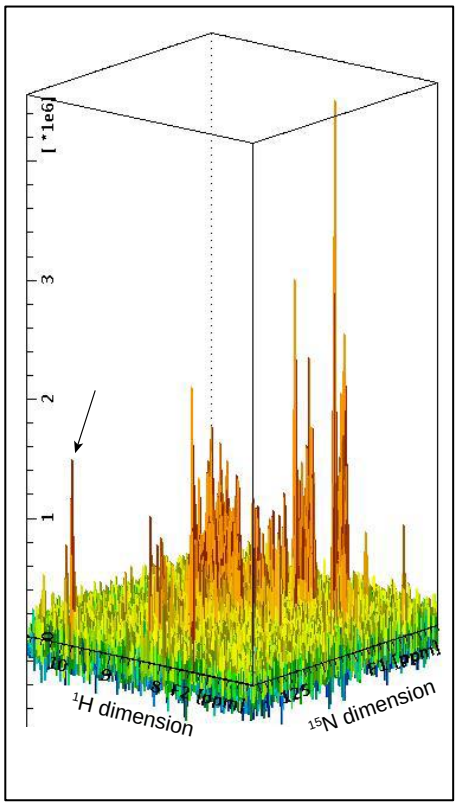


Figure 18. 2 Dimensional $^{1}H/^{15}N$ HSQC spectra of the PstSCR1.

2D $^{1}H/^{15}N$ HSQC of the PstSCR1 domain multiple intense amide peaks. As isolated PstSCR1 domain has 92 amino acids, most of the backbone amide proton resonances were detected. $^{1}H/^{15}N$ HSQC GAP domain spectrum is shown in Figure 18. Specifically, we detected number of tyrtophan nε1 peaks (shown in black arrow).

3.8 Secondary structural analysis of PstSCR1 by Circular Dichroism (CD)

CD of the SEC purified PstSCR1 protein is measured between 200-250 nm wavelengths. Measurements were repeated in three different sets. ΔE values of each wavelength between 200-250 nm are plotted against wavelength. Resulting CD spectrum of each set is shown in Figure 19.

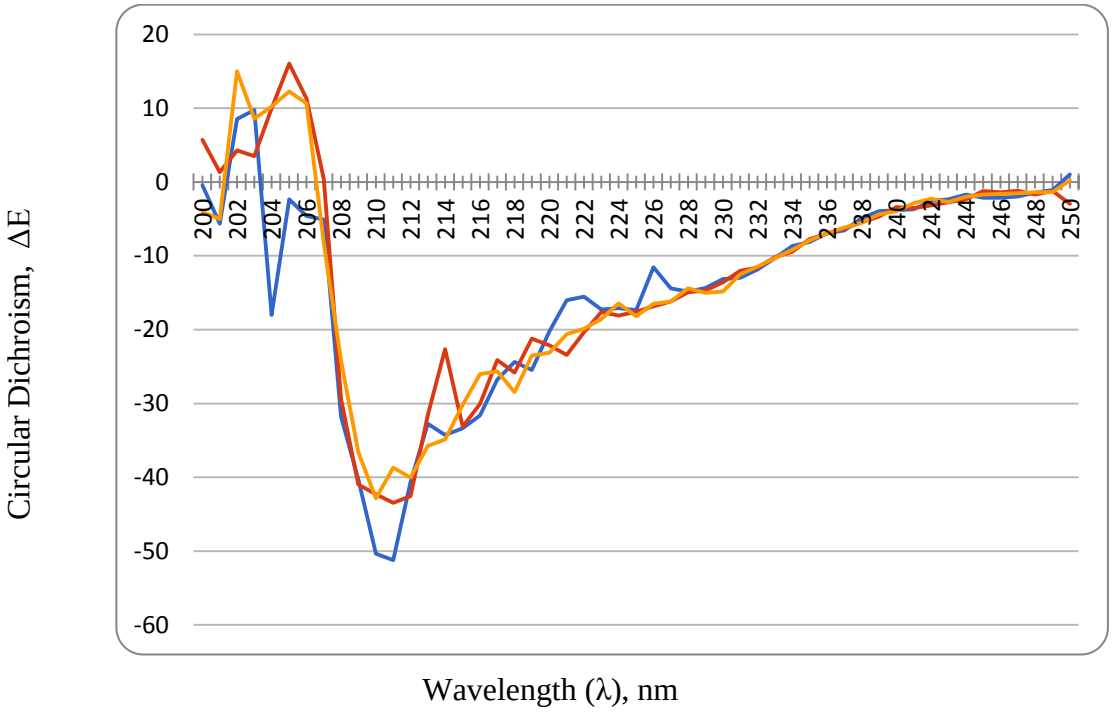


Figure 19. CD Spectrum of PstSCR1 protein. Measurements were repeated in three different sets shown in the spectra in different colors.

CD Spectrum of PstSCR1 is revealed that the protein has negative bands between 207-250 nm and positive bands between 200-207 nm. The spectrum of PstSCR1 has similar spectrum pattern with the β-sheet and β-turn conformations when compared with the reference spectra shown previously in Figure 4 as explained in (Greenfield, 2006). In the same reference spectra model, it can be seen that a protein with an irregular (disordered) conformation should have negative bands between 0-220 nm in its spectra but no positive bands at all. Considering this fact it can be concluded that *E. coli* expressed, and SEC purified PstSCR1 is in a folded state and has β-sheet and β-turn conformations.

3.9 DSF analysis of PstSCR1 and determination of melting point (Tm)

DSF Analysis of PstSCR1 was performed as presented in section 2.14. According to the DSF data, melting point (Tm) of PstSCR1 protein was found as ~35 ^{o}C (Figure 20b).

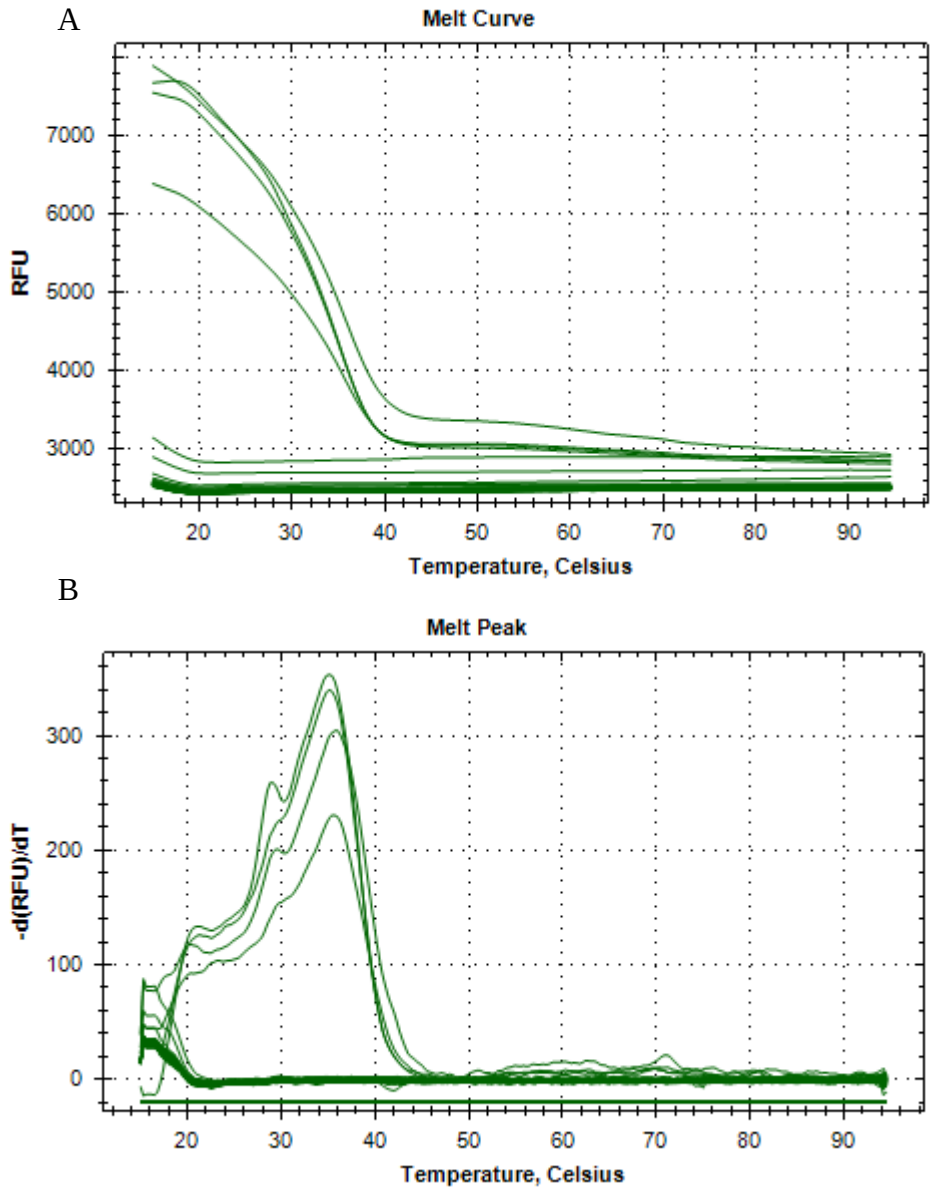


Figure 20. DSF Analysis of PstSCR1 protein. A. Melt curve B. Melt peak.

3.10 Zinc binding prediction

Zinc binding prediction of PstSCR1 was carried out by ZincExplorer bioinformatics tool (Chen et al., 2013). According to the prediction tool there are three potential

residues which may bind zinc ion (Figure 21). However this prediction should still be confirmed by zinc binding assays.

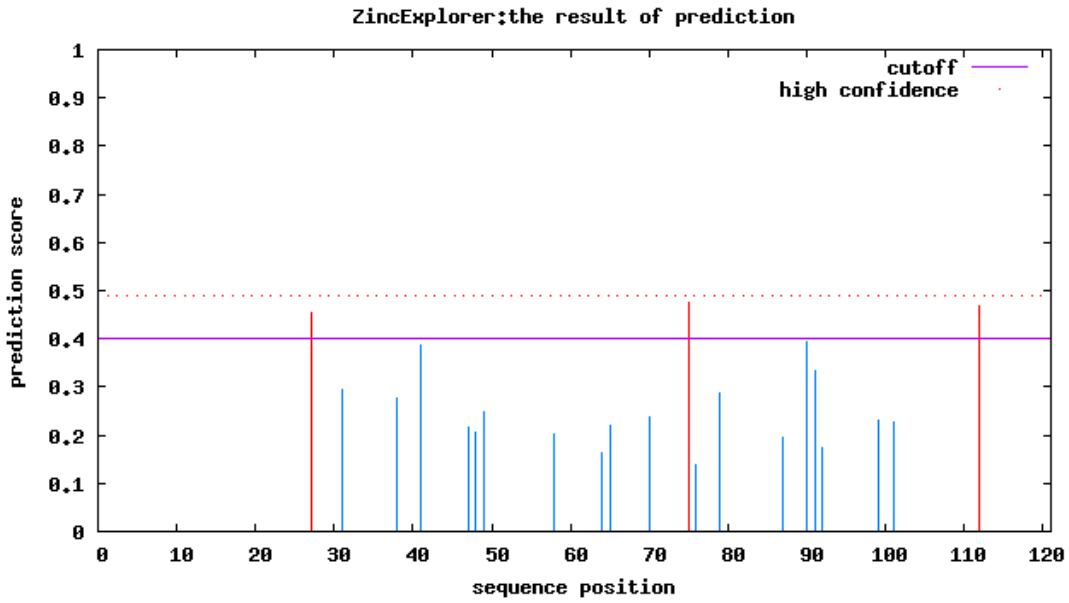


Figure 21. Zinc binding prediction of PstSCR1 by ZincExplorer (Chen et al., 2013).

3.11 DSF analysis of PstSCR1 in the presence of zinc

To confirm the binding of PstSCR1 protein to the zinc ion by observing the changes in the Tm of PstSCR1 protein in the presence of zinc, a separate DSF Analysis was carried out. An increase in the Tm of PstSCR1 protein was expected with the increasing concentrations of zinc ion due to the binding of the protein to the zinc and having a stronger structure. However, in the DSF Analysis of the PstSCR1 protein in the presence of zinc, significant peaks could not be observed (Figure 22 a and b and Appendix F). Absence of the regular increase in the Tm of the protein with the increasing zinc concentrations suggesting that, in fact, PstSCR1 protein does not bind to zinc.

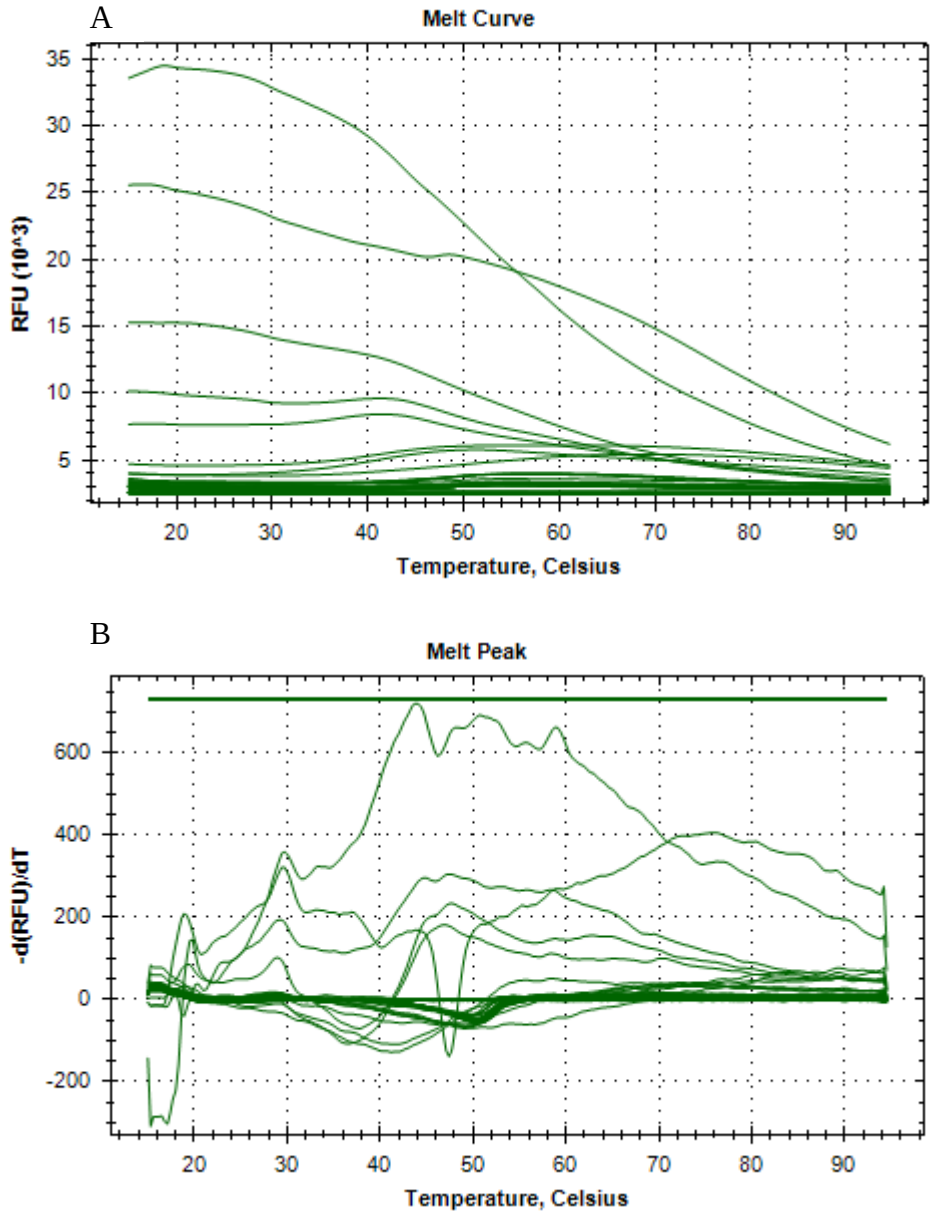


Figure 22. DSF Analysis of PstSCR1 in the existence of Zinc ion A. Melt curve B. Melt peak.

3.12 Zinc binding assay

Zinc binding prediction of PstSCR1 (see section 3.10), carried out by ZincExplorer bioinformatics tool (Chen et al., 2013) estimated that there are three potential

residues which may bind zinc ion. To confirm this prediction zinc binding assay was carried out by stripping the Nickel content of Ni-NTA chromatography cartridge and recharging the column by zinc and loading the protein to zinc charged column as explained in section 2.15. The fractions were analysed by SDS-PAGE gel (Figure 23).

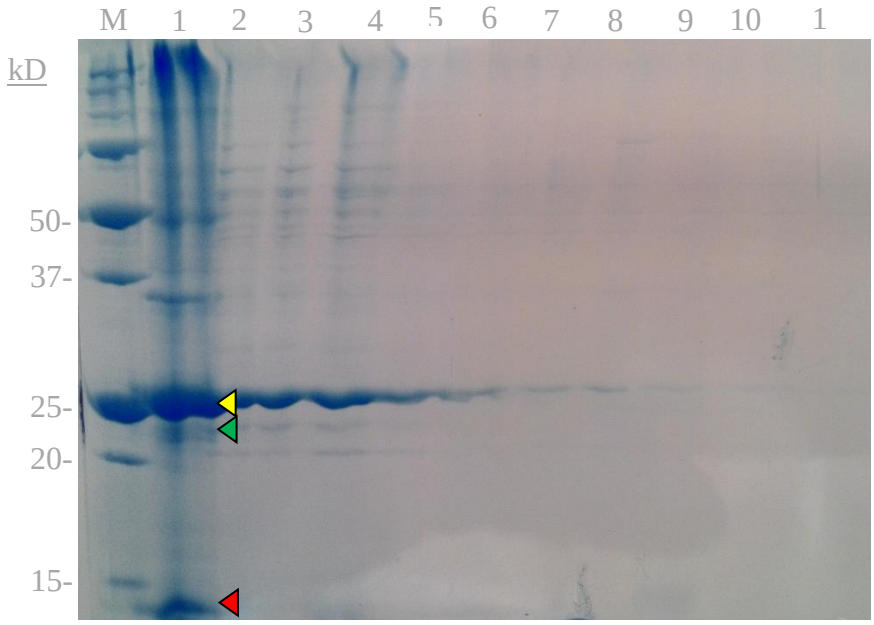


Figure 23. SDS-PAGE gel analysis of zinc binding assay. Red arrow shows the PstSCR1 protein band (~10 kDa). Yellow arrow shows the GST band (~26 kDa) and the weaker green arrow shows the 3C-His protease band (~22 kDa). 1. 3C-His protease digested PstSCR1-GST protein, 2.Flow through 3. Wash fraction 4-7: Elution fractions of 3C-His protease by elution buffer containing imidazole, 8-11: Elution fractions of PstSCR1 by stripping buffer containing EDTA.

PstSCR1-GST protein was digested with 3C-His protease prior to loading to the zinc charged column. Digested protein was loaded in the first lane of the gel (Figure 23). PstSCR1, GST and 3C-His protease bands are shown with colour coded arrows. Zinc charged column material was first washed using washing buffer (section 2.15) and loaded on the 3rd lane. 3C-His protease was eluted in four fractions by elution buffer

containing imidazole (loaded on the lanes 4-7) and then eluted by stripping buffer containing EDTA (loaded on the lanes 8-11).

PstSCR1 protein was expected to be eluted by stripping buffer from the zinc charged column and pure protein bands were expected to be seen on the SDS-PAGE gel (lanes 8-11) if it was binding to the zinc as predicted by bioinformatics tools. However no protein bands could be observed on the lanes 8-11 at all. This result is suggested that, in contrast to the bioinformatics prediction tools, PstSCR1 protein does not bind to zinc therefore flowed through the zinc charged column.

3.13 Characterization of PstSCR1 by bioinformatics tools

Secondary structure of PstSCR1 and prediction of the model of PstSCR1 was studied using bioinformatics tools presented in Section 3.13.1 and 3.13.2.

3.13.1 Secondary structure of PstSCR1

Secondary structure prediction of PstSCR1 was carried out by Phyre2 (Figure 24) (Kelley et al., 2015) (Wass, Kelley, & Sternberg, 2010). In the figure, the sequence shows the amino acid sequence of PstSCR1. Secondary structure (SS) confidence is presented with the colours, the highest confidence with red and lowest confidence with violet colour. Disordered residues are presented with a question mark (?). Alpha helix is presented with green helix image and beta strand is presented with a blue arrow (Kelley et al., 2015) (Wass, Kelley, & Sternberg, 2010).

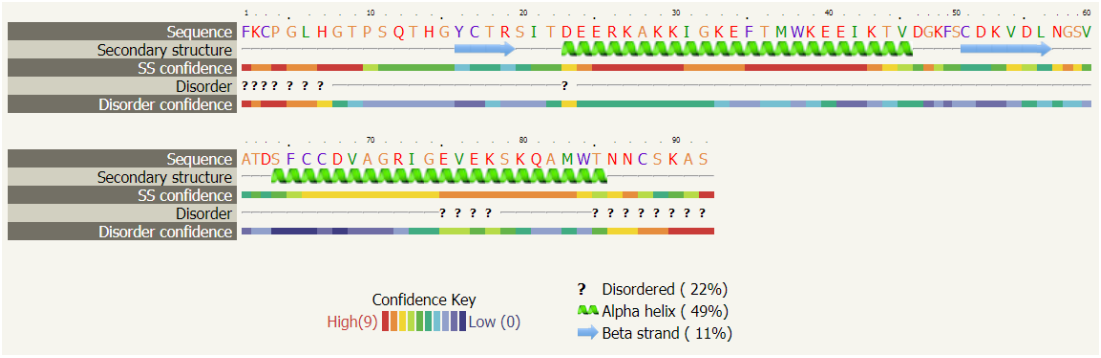


Figure 24. Secondary structure prediction and disorder of PstSCR1 predicted by Phyre2. (Wass, Kelley, & Sternberg, 2010) (Kelley et al., 2015).

3.13.2 Prediction of the model of PstSCR1

Model of PstSCR1 was predicted using Phyre2 bioinformatics modelling tool shown in Figure 25a (Kelley et al., 2015) (Wass, Kelley, & Sternberg, 2010). Predicted image was coloured by rainbow N to C terminus. 17 residues (18% of the sequence) have been modelled with 17.0% confidence. PstSCR1 amino acid sequence was also submitted to RaptorX to make an estimation on the model of the protein. The prediction revealed the model presented in Figure 25b. In this model 92 (100%) residues are modelled and 5 (5%) positions were predicted as disordered (Källberg et al., 2012). Binding sites of PstSCR1 predicted by RaptorX shows that P4, L6 and K48 residues and V76 are predicted to bind Calcium ion; A82 residue is predicted to bind Glycerol and I21, D23 and K30 residues are predicted to bind Chloride ion (Källberg et al., 2012).

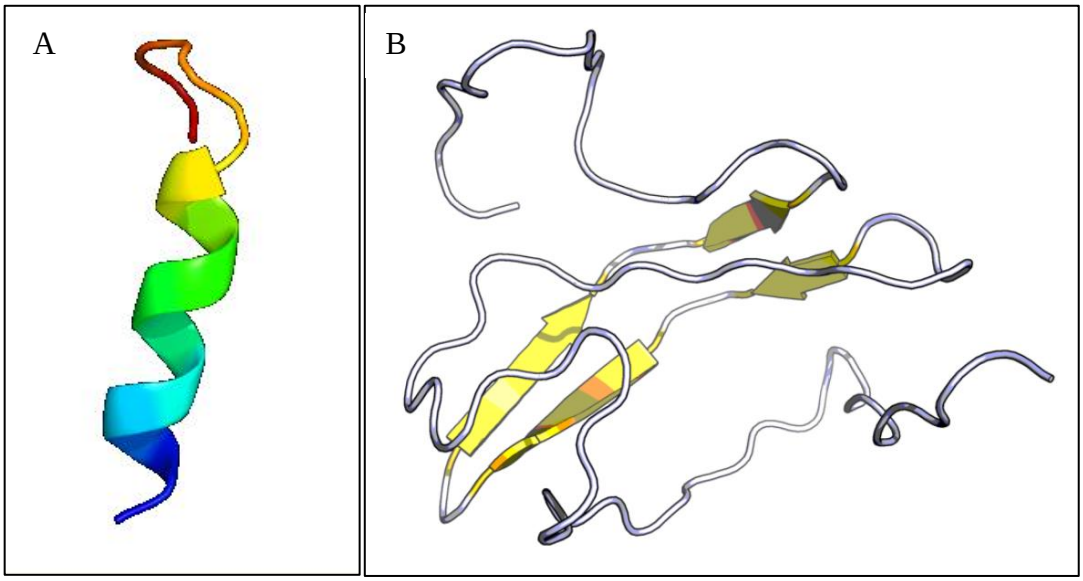


Figure 25. Model of PstSCR1 predicted by: A. Phyre2 (Kelley et al., 2015) (Wass, Kelley, & Sternberg, 2010) and B. RaptorX (Källberg et al., 2012).

3.14 Cloning of TaYr10 domains into pET32a+ vector

Two Nucleotide Binding sub-domains of TaYr10; N-NB and C-NB were cloned into pET32a+ vector as explained in section 2.18 and overexpressed in *E. coli* as explained previously in section 2.4. However the expression level was not enough to further characterize the expressed proteins by X-ray crystallography.

3.14.1 PCR amplification of the TaYr10 domain coding regions

TaYr10 protein domain coding parts were amplified by PCR and loaded on the 1% agarose gel. Amplified bands were observed at the expected sizes (Figure 26 and 27). The bands were excised from the 1% agarose gel and PCR amplified DNA were extracted from the gel for further cloning steps.

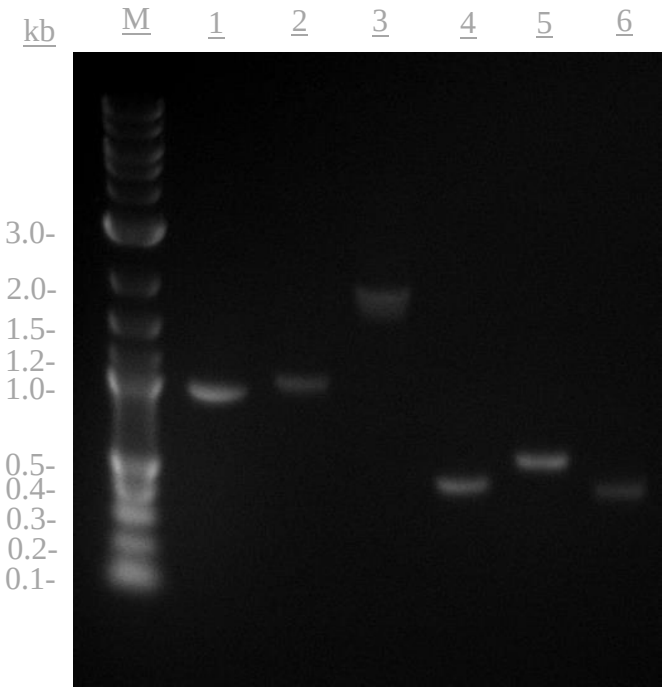


Figure 26. PCR amplified bands of the coding regions of the separated domains of TaYr10. M: 2-log DNA Ladder, NEB. 1. 512S-824R (LRR Domain), 936 kb; 2. 171G-512S (NB Domain), 1023 kb; 3. 171G-824R (NB-LRR Domains together), 1959 kb; 4. 171G-323G (N-NB Domain), 456 kb; 5. 323G-512S (C-NB Domain), 567 kb; 6. 1M-144S (CC Domain), 429 kb.

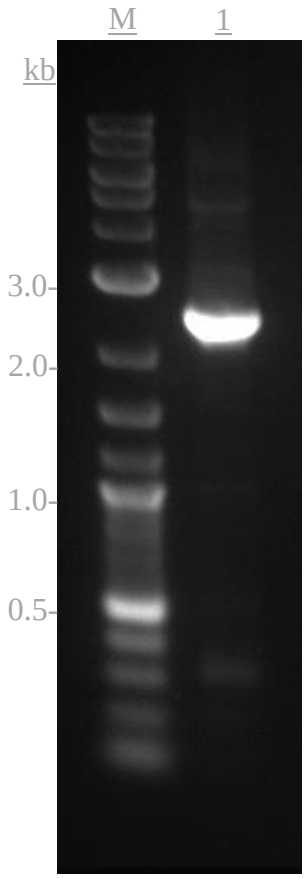


Figure 27. PCR amplified band of the whole TaYr10 gene (coding region of CC-NB-LRR Domains) which has a 2472 kb size shown in the lane 1; M: 2-log DNA Ladder, NEB.

3.15 IMAC purification of N-NB domain of TaYr10

N terminus Nucleotide Binding (NB) domain of TaYr10 protein was purified using a 5 mL HisPur Ni-NTA chromatography cartridge (Thermo Fisher) using the purification protocol of the manufacturer. Elution of the protein was carried out by SEC detector to obtain the highest protein concentration by taking the protein in the lowest amount of elution buffer. SEC Elution chromatogram of N-NB Domain of TaYr10 protein is shown in Figure 28. Protein elution fractions eluted between 20-80 mL were collected to further purify by SEC.

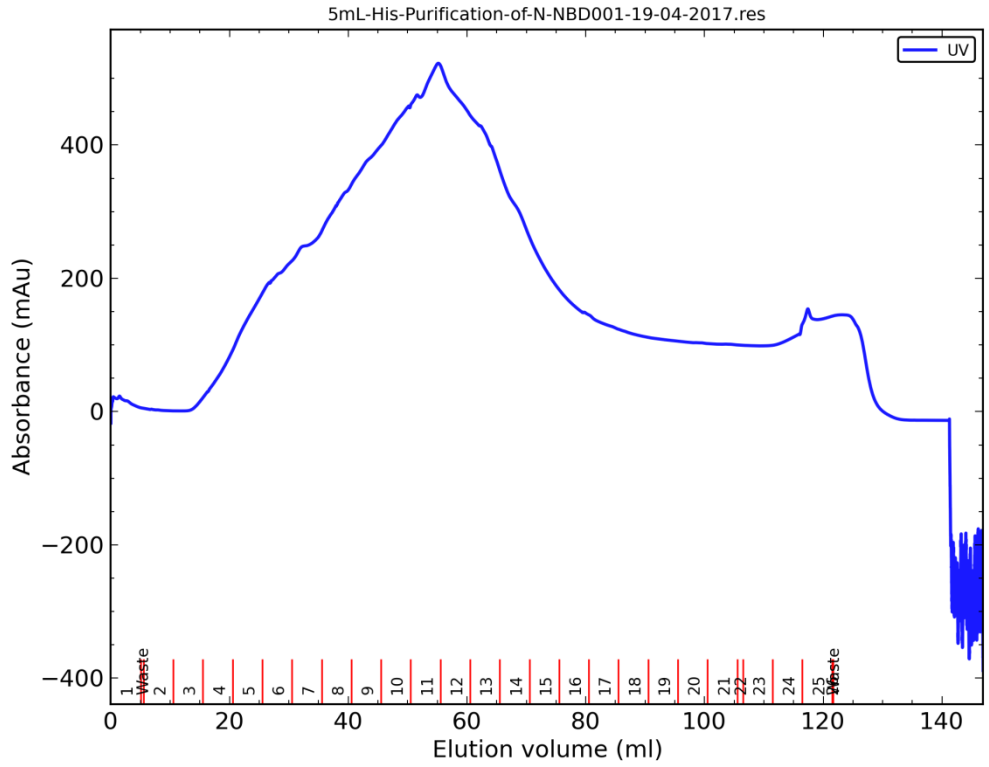


Figure 28. SEC Elution chromatogram of the N-NB domain of TaYr10 protein. Fractions eluted between 20-80 mL volumes belong the proteins eluted from the Ni-NTA cartridge.

3.16 SEC purification of N-NB domain of TaYr10

Purification of the N-NB domain of TaYr10 protein expressed in pET32 was carried out by SEC using S75 column (GE) in order to get rid of the unspecific proteins which were also eluted from Ni-NTA cartridge. SEC chromatogram of the purification of N-NB protein is presented in Figure 29. According to the manual of the column (S75-GE) and the experiments carried out by protein standards, the sharp peak eluted at 10.9 mL belongs to the pure recombinant N-NB including the tags (S-tag, His-tag and Trx-tag (thioredoxin)) coming from the vector. Elution fractions eluted between 10-13 mL were collected to use in the subsequent experiments.

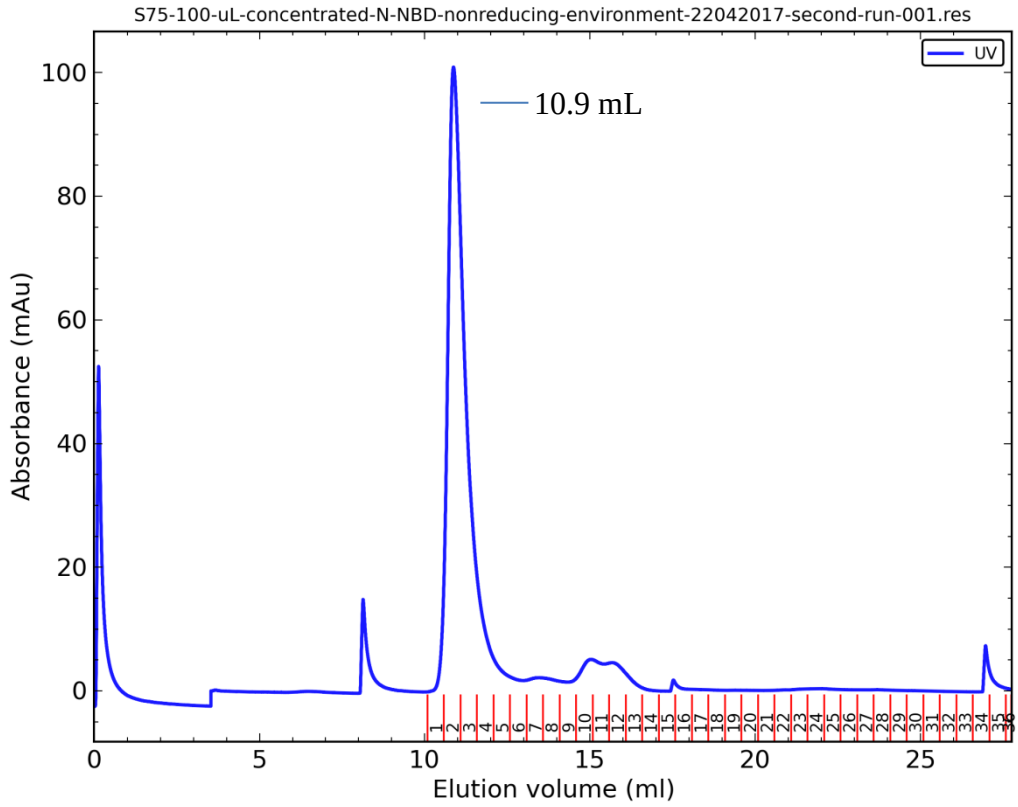


Figure 29. SEC chromatogram of the purification of N-NB protein. The peak eluted at 10.9 mL belongs to the recombinant N-NB protein including the tags (S-tag, His-tag and Trx-tag (thioredoxin)) coming from the vector.

3.17 Thrombin digestion and SEC purification of N-NB domain of TaYr10

pET32 expressed N-NB domain of TaYr10 protein was digested with thrombin in order to cut the protein from the thrombin cleavage site and remove thioredoxin and His tags from the N-NB protein (which keeps the S-tag bound to the N-NB protein). Digestion product was primarily flowed through a Ni-NTA chromatography column to get rid of the thioredoxin and His tags from the N-NB protein (including the S-tag) and then further purified by SEC. SEC chromatogram of the purification of the digested protein is presented in Figure 30. After digesting the protein and separating the tags, the N-NB protein (bound to the S-tag) was eluted by SEC at 14.3 mL with

its smaller size. The peak observed at 11.1 mL suggested that thrombin digestion was not completed although the digestion was carried out overnight.

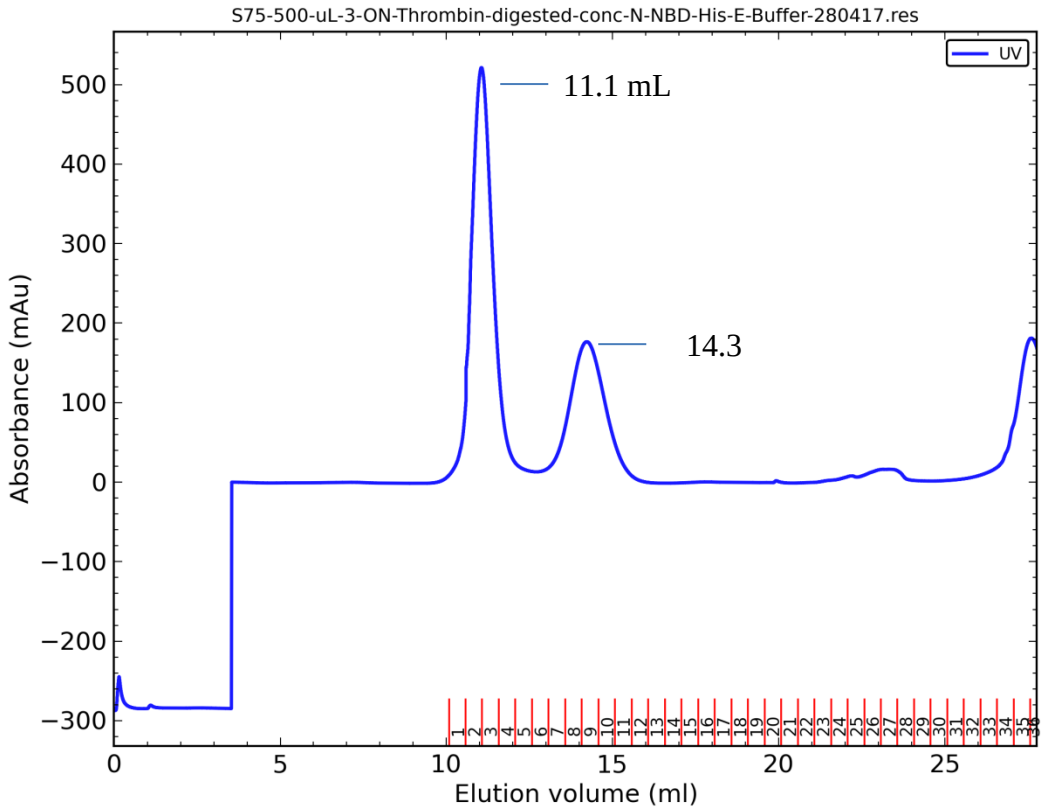


Figure 30. SEC Purification of pET32 expressed N-NB domain of TaYr10 after thrombin digestion. N-NB protein (bound to the S-tag) was eluted at 14.3 mL.

Digesting the N-NB protein bound to the tags using thrombin protease only removes the thioredoxin and His tags but not the S-tag. In order to remove all three tags from N-NB protein, the protein with the tags had to be digested using enterokinase which is an expensive protease that we did not have in the laboratory. Since the S-tag bound to the N-NB protein could highly affect the crystallization process and damage the structure of the N-NB protein during X-ray crystallization studies, it was

decided to remove all of the tags by partial digestion of the protein instead of purchasing enterokinase.

3.18 Limited proteolysis (partial digestion) of pET32 expressed N-NB domain of TaYr10.

pET32 expressed N-NB domain of TaYr10 protein was partially digested by trypsin. SDS - Gel Analysis (see Figure 31) of the partial digestion of N-NB protein by trypsin protease showed that 1/800 ratio (Lane 8 in Figure 30) is the optimum e/p ratio to partially digest the N-NB protein and remove the tags. The band at ~18.2 kDa is the N-NB protein and intensity of the protein is the highest at the mentioned e/p ratio.

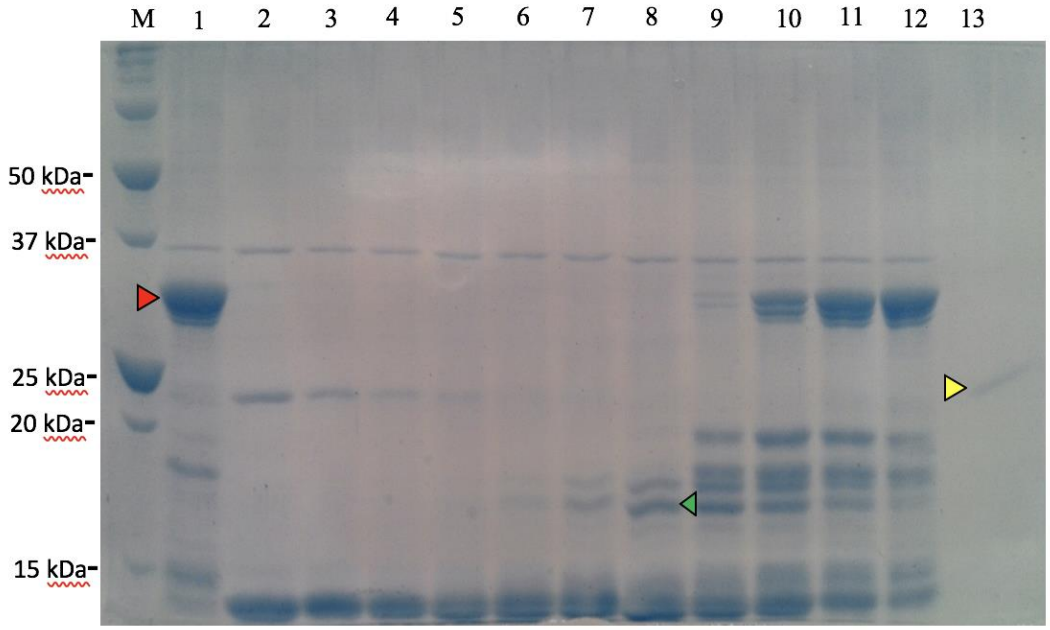


Figure 31. SDS-Gel Analysis of trypsin digested N-NB protein and the tags expressed in pET32. The band shown in lane 8 is expected to be the N-NB protein (~18.2 kDA) yielded after partial digestion (shown in green arrow). M. Marker (Precision-Plus-Protein, Bio-Rad), 1. Undigested N-NB protein and the tags (shown with a red arrow), 2-12. e/p ratios of the trypsin digestion from 1/12.5 to 1/12800, two fold decrease in each lane respectively. 13. Trypsin as a marker (shown in yellow).

Although the partial digestion method can be used to digest and remove the tags in order to obtain N-NB separated from its tags, it would still be very challenging to separate the pure N-NB by SEC from the other proteins due to their closer size.

3.19 SEC purification of the C-NB domain of TaYr10

Ni-NTA purified C-NB domain of TaYr10 protein expressed in pET32 was additionally purified by using SEC-S75 column (GE) in order to remove the unspecific proteins eluted from the Ni-NTA column. SEC chromatogram of the further purification of C-NB protein is presented in Figure 32. According to the manual of the column (S75-GE) and the experiments carried out by protein standards, the sharp peak eluted at 11.2 mL belongs to the pure recombinant C-NB protein including the tags coming from the vector (S-tag, His-tag and Trx-tag (thioredoxin)). Other two sharp peaks observed at 14.0 mL and 15.3 mL are supposed to belong either unspecific proteins eluted from the Ni-NTA column or a tag or the tags break off the C-NB protein. Elution fractions eluted between 10-12 mL were collected to use in the subsequent experiments.

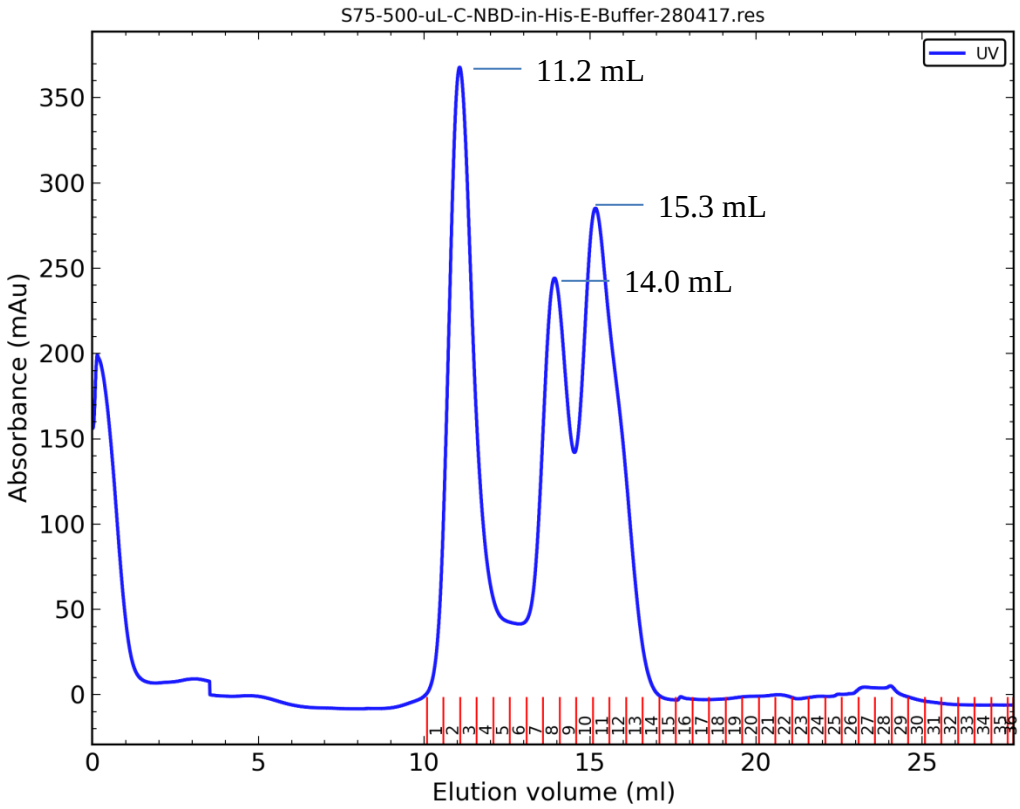


Figure 32. SEC chromatogram of the purification of C-NB protein. The fraction eluted at 11.2 mL belongs to the recombinant C-NB protein including the tags coming from the vector (S-tag, His-tag and Trx-tag (thioredoxin)). Other peaks observed in the chromatogram belong to the unspecific proteins eluted from the Ni-NTA column.

3.20 Thrombin digestion and SEC purification of the C-NB domain of TaYr10

pET32 expressed C-NB domain of TaYr10 protein was digested with thrombin in order to cut the expressed protein from the thrombin cleavage site and remove thioredoxin and His tags from the C-NB protein (digestion with thrombin keeps the S-tag bound to the C-NB protein). Thrombin digested C-NB protein was purified by SEC and thioredoxin and His tags were separated from the C-NB protein (bound to the S-tag). SEC chromatogram of the purification of the digested protein is presented in Figure 33. After digesting the recombinant protein and cleaving the tags

(thioredoxin and His), C-NB protein together with the S-tag was eluted by SEC at 13.37 mL observed as a small peak. Although the digestion process was carried out overnight, the peak observed at 11.2 mL suggested that thrombin digestion was not completed. The peak observed at 15.9 is supposed to belong to the tags (thioredoxin and His) cleaved by thrombin digestion.

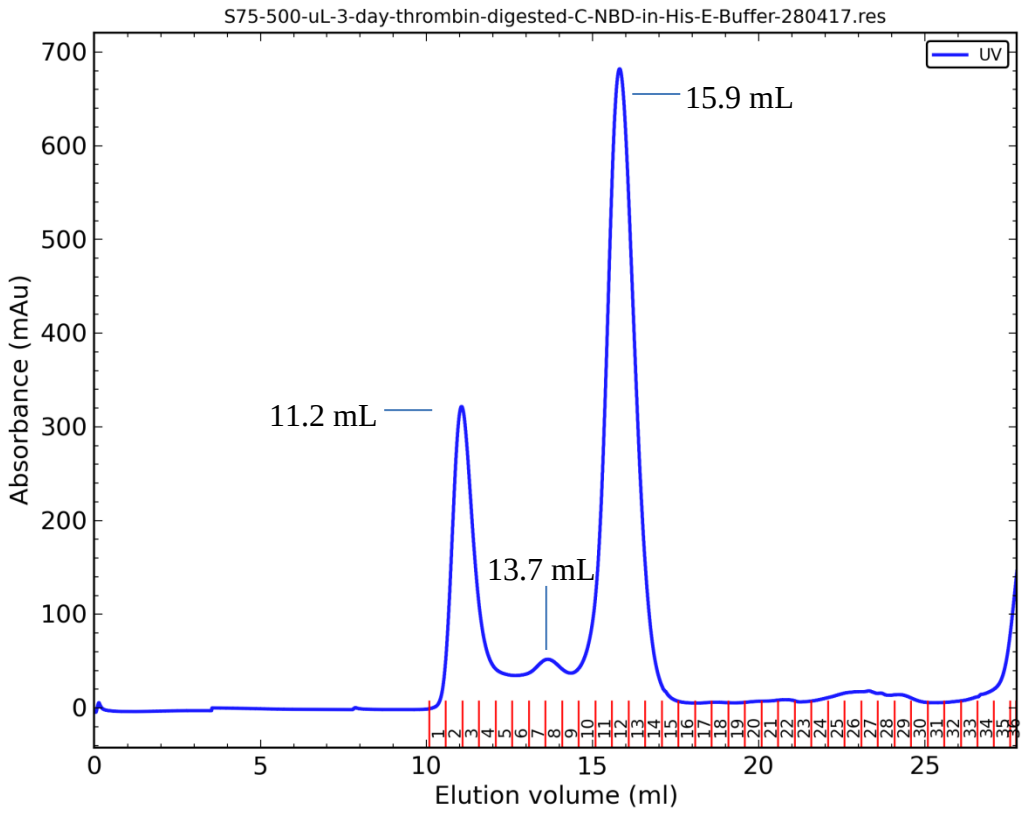


Figure 33. SEC Purification of pET32 expressed C-NB domain of TaYr10 after thrombin digestion. C-NB protein (bound to the S-tag) was eluted at 13.7 mL. The peak observed at 15.9 belongs to the thioredoxin and His tags cleaved by thrombin digestion.

As previously explained in part 4.13.4 digesting the C-NB protein bound to the tags using thrombin protease only removes the thioredoxin and His tags but not the S-tag. In order to remove all three tags from C-NB protein, the protein with the tags had to be digested using enterokinase which is an expensive protease that we did not have

in the laboratory. Since the S-tag bound to the C-NB protein could highly affect the crystallization process and damage the structure of the C-NB protein during X-ray crystallization studies, it was decided to remove all of the tags by partial digestion procedure.

3.21 Limited proteolysis (partial digestion) of pET32 expressed C-NB Domain of TaYr10.

pET32 expressed C-NB domain of TaYr10 protein was partially digested by trypsin in order to identify the optimum digestion conditions to remove all of the tags from C-NB protein. SDS - Gel Analysis (see Figure 34) of the partial digestion of C-NB protein by trypsin protease showed that digestion of C-NB protein with its tags yielded a strong band lower than 15 kDa in all digestion samples and several tiny bands at lower enzyme concentration. Considering the expected size of C-NB protein as ~21.6 kDA suggests that partial trypsin digestion method yields either very low amount of pure C-NB (see Figure 34 green band) or it digests the expressed protein into smaller unspecific digestion products. Even if the partial digestion method achieved the low yield of pure C-NB; it would be very challenging to separate the pure C-NB by SEC from the other proteins which have a very close size. In the best case it was not sufficient to continue for X-ray crystallization studies.

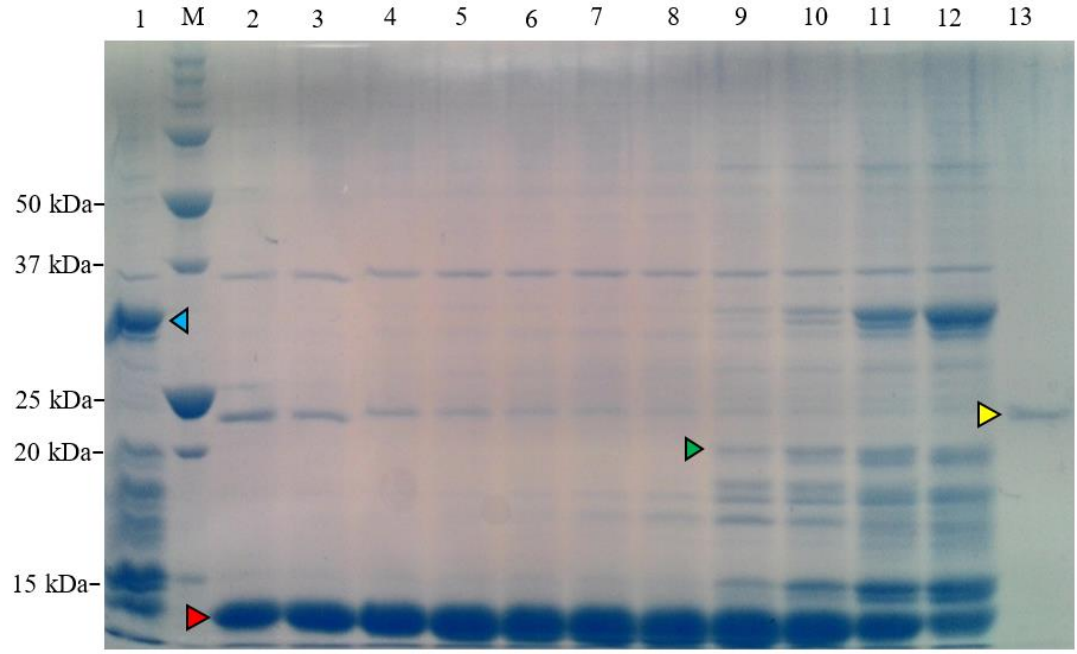


Figure 34. SDS-Gel Analysis of trypsin digested C-NB protein and the tags expressed in pET32. Partial digestion of C-NB protein and the tags mostly yielded the smaller protein fragments below 15 kDA (shown with a red arrow). M: Marker (Precision Plus Protein, Bio-Rad), 1. Undigested C-NB protein and the tags (shown with a blue arrow), 2-12. e/p ratios of the trypsin digestion from 1/12.5 to 1/12800, two-fold decrease in each lane respectively. Green arrow indicates the low yield, potential C-NB protein band, free from the tags. 13. Trypsin as a marker (shown with a yellow arrow)

CHAPTER 4

CONCLUSION

In this thesis one of the Pst effector proteins, PstSCR1 was successfully cloned into bacterial expression plasmids, expressed in *E. coli*, purified using several different purification techniques and subsequently structurally characterized by SAXS, protein NMR and other theoretical and practical protein characterization methods.

PstSCR1 was first detected as a small cysteine-rich candidate effector first identified by Yin et al., 2009. In our laboratory, it was discovered that it triggers PTI responses at the cell surface (apoplast) of the plants as secreted by *Puccinia striformis f. sp. tritici* (Dagvadorj, et al., 2017). Since apoplastic effectors are able to interact with the surface proteins or other outer surface molecules of the host organism and cause disturbance of the function of the related molecules (Dodds & Rathjen, 2010), structural characterization of such effectors present valuable information to the literature for the ongoing or future works in the field of protein–protein interactions in order to identify the possible peers of the disease agent of Pst. Characterization of effectors and effector like proteins can also provide valuable understandings on their cellular functions on the purpose of reducing or totally preventing crop loss caused by biological agents by means of iRNA or transgenic studies.

In this study PstSCR1 was cloned into pET28a+ and pGEX-6P-1 bacterial expression vectors and expressed in *E. coli*. *PstSCR1* expressing cells were homogenized and the proteins were purified from the crude cell extract by Ni-NTA and GST Affinity Column respectively. Additional purification procedures were applied such as ultrafiltration and size exclusion chromatography. Overexpression

and purification of the gene was verified by SDS-PAGE and western blot analysis. Tags were digested and removed from the protein and as a result pure protein was obtained for further characterization.

In fact, the primary aim of this thesis was identifying the crystal structure of PstSCR1 by X-ray crystallography. However, cloning of a foreign gene into a microorganism and expression of this foreign gene in the native form and obtaining the pure protein in its folding state has always been a challenging task due to the unique nature of each protein. Even if a pure protein is obtained in its native form, keeping this protein in its ideal condition throughout the research and performing characterization studies using this sensitive biological molecule is a tough process. Although several optimization studies were carried out, it was not possible to obtain a high yield which is essential for crystallization. Therefore, the protein was characterized using other valuable protein characterization methods such as SAXS and NMR.

Small Angle X-ray Scattering (SAXS) studies of the protein of interest was carried out in Taiwan National Synchrotron Radiation Research Center, NSRRC- SR- 23 A, SAXS Beam Line. 3D structure of the pure recombinant PstSCR1 effector was resolved using SAXS in solution. Beside the structure, Rg of the studied protein was also found by SAXS as 23.1 Å. It is found out that the nano formations are arranged in the crystal configuration which has a distance of 39.3 Å. Melting point of PstSCR1 was found out as ~35 ^{o}C by DSF.

Since the relationship between MW and Rg of several globular proteins based on the light scattering and SAXS data was studied (Mylonas & Svergun, 2007) and for the globular proteins it is found out that the relevance between MW–Rg is represented by a power law with an exponent of 0.37 (Smilgies & Folta-Stogniew, 2015), MW of PstSCR1 could be find out using the plot "radius of gyration *versus* molecular weight relation" prepared by using several globular proteins. According to the plot, MW of PstSCR1 was roughly found out as ~20 kDa which is

~2 times more than the size of PstSCR1 suggesting that PstSCR1 forms dimer. This finding is also in parallel with the SEC results.

The protein mixture had to be concentrated by ultra-filtration prior to loading on the SEC column. It was found out that significant amount of protein was precipitated as white precipitates at the bottom of the tube during concentration process. SDS PAGE analysis of this white precipitates revealed that they are composed of mainly PstSCR1 and GST proteins. Unfortunately loss of protein during concentration of the protein or in the other steps such as digestion, purification etc. caused substantial amount of decrease in the yield and amount of protein. This reason was one of the main obstacles not to be able to continue to carry out crystallography studies.

NMR experiments the protein was carried out in University of Houston, Texas, USA on a 800 MHz spectrometers TXI cryoprobe, optimized for proton detection. It is well known that 1-Dimensional (1D) water suppression NMR experiment is a benchmark for controlling if the recombinantly expressed proteins are folded and globular. From 1D NMR studies it is found out that PstSCR1 protein clearly showed unclustered amide proton resonances. Furthermore, well-dispersed peaks of amide proton resonance from 6 ppm to 9 ppm was observed which is a good indication of foldedness of PstSCR1. Aliphatic residues, Valine, Isoleucine, Leucine when packed against polar residue or water molecules moves significantly upfield when protein is folded. In our 1D-water supression spectra indeed contains more than two resonance despite finding such resonance is rare. Therefore recombinantly expressed PstSCR1 in pGEX-6P-1 expression system, is well-folded and globular.

Circular Dichroism and DSF data were also supported the findings of folded state of the protein.

Characterization of PstSCR1 by bioinformatics tools was studied. Based on the secondary structure prediction studies, model of PstSCR1 was predicted. Zinc binding prediction of PstSCR1 was also carried out by ZincExplorer bioinformatics tool. Aforementioned prediction tool is estimated that there are three potential

residues which may bind zinc ion. To confirm this prediction zinc binding assay was carried out by stripping the Nickel content of Ni-NTA chromatography column and recharging the column by zinc. The protein was loaded to the zinc charged column and eluted with stripping buffer containing EDTA as explained in details in section 2.15. Surprisingly SDS-PAGE gel analysis showed that, in fact, PstSCR1-GST protein does not bind to zinc. If it was so, PstSCR1 protein was expected to be eluted by stripping buffer from the zinc charged column and pure protein bands were expected to be seen on the SDS-PAGE gel.

A separate DSF study on the PstSCR1 protein was also performed in the existence of zinc ion and probable shift in the Mp of PstSCR1 protein in the presence of zinc was observed. An increase in the Tm of PstSCR1 protein was expected with the increasing concentrations of zinc ion based on the affinity of the protein to the zinc and consequently having a stronger structure. An increase in the melting point of the protein with the increasing zinc concentration was not observed suggesting that PstSCR1 protein does not bind to zinc indeed which also supports the findings found out from zinc binding assay.

Since PstSCR1 effector protein was successfully overexpressed and purified in the framework of this thesis, this pure protein can be used in the future studies to find out the interacting proteins of the PstSCR1 effector during the infection process by carrying out protein-protein interaction experiments. Such studies can open up new ways to understand the disease mechanism of the pathogen and fight against the plant pests. Such findings may also lead to the development of alternative plant protection products and methods which might be used in the agriculture industry in favour of food safety and food security.

Apart from the characterization of Pst effector protein PstSCR1, TaYr10 which is a resistance protein of wheat against Avr containing pathogen races was also studied. Since it was not possible to express this large eukaryotic molecule in *E. coli* at the beginning, TaYr10 was separated into its sub-domains and the sub-domains; LLR,

CC, N-NB and C-NB were cloned into bacterial expression plasmids. N-NB and C-NB domains of TaYr10 were overexpressed in *E. coli* and purified by IMAC, ultrafiltration and Size Exclusion Chromatography. Despite several optimization attempts LLR domain of TaYr10 protein was failed to express in bacteria. If it was succeeded to express and obtain the pure LRR protein, the aim was also studying the possible interaction between the LRR domain of the resistance protein of wheat and the PstSCR1 effector protein of Pst by DSF and other methods.

N-NB and C-NB domains of TaYr10 protein were integrated into pET32 and overexpressed in *E. coli.* Purification of these domains were carried out first by Ni-NTA Chromatography and ultrafiltration and then by SEC using S75 column (GE). pET32 plasmid adds three tags (S-tag, His-tag and Trx-tag (thioredoxin)) to the protein of interest. Therefore pET32 expressed N-NB and C-NB domains of TaYr10 protein was digested with thrombin protease and thioredoxin and His tags were removed from the protein of interests.

Digesting the pET32 expressed N-NB and C-NB domains of TaYr10 using thrombin protease only removes the thioredoxin and His tags but not the S-tag. In order to remove all three tags from N-NB and C-NB protein, pET32 expressed protein domains with their tags had to be digested using enterokinase which is an expensive protease that we did not have in the laboratory. Since the S-tag bound to the N-NB and C-NB proteins could highly affect the crystallization process and damage the structure of the N-NB and C-NB domains during X-ray crystallization studies, it was decided to remove all of the tags by partial digestion of the protein instead of purchasing enterokinase.

pET32 expressed N-NB domain of TaYr10 protein was partially digested with trypsin in order to separate N-NB from its tags. By this method, pure N-NB protein was obtained without the tags and the conditions required for the partial digestion were identified. Although the partial digestion method could be used to digest and remove the tags, it could still be very challenging to separate the pure N-NB by SEC

from the other proteins existing after partial digestion method due to their closer sizes.

pET32 expressed C-NB domain of TaYr10 protein was also partially digested with trypsin in order to separate C-NB from its tags. Partial digestion studies carried out with pET32 expressed C-NB domain revealed that this method yields either very low amount of C-NB or it digests the expressed protein into smaller unspecific digestion products. Even if the partial digestion method achieved the low yield of pure C-NB it could still be very challenging to separate the pure C-NB by SEC from the other digestion products as similarly faced in the separation of N-NB by SEC.

Although N-NB and C-NB subdomains of TaYr10 protein were successfully cloned, expressed and purified from the bacterial proteins, the purest form of these proteins without tags could not be achieved by SEC. Therefore, structures of the N-NB and C-NB subdomains of TaYr10 protein could not be further characterized by X-ray crystallography.

Ion exchange columns can still be a good solution for the future studies to separate and purify the N-NB and C-NB subdomains from other unspecific proteins which have closer sizes to the POI. Also, another future prospect for these proteins can be the usage of fermenter, instead of using shaker incubator, to boost up the overexpression yield of these proteins. Application of these future prospects may allow the production of N-NB and C-NB subdomain proteins with a higher yield and usage of ion exchange chromatography in addition to SEC would allow obtaining these proteins with higher purity and studying the high resolution X-ray structure of the proteins. Production of these nucleotide binding proteins in higher yield and purity would also allow to find out interacting molecules during the resistance mechanism by studying protein-molecule interaction studies.

Studying plant-pathogen interactions is not only important for science but it is also very important in terms of food security. Food security has a significant importance on feeding the world population. It has also a global importance in the world

economy in order to prevent the price spikes of fundamental grains such as rice and wheat. According to FAO the world population supply 40% of their calories solely from three staples. These are rice, wheat and maize. In case of a global epidemic in just one these staples, world population would face a devastating scarcity and huge price fluctuations.

Therefore such studies like this thesis on plant-pathogen interactions and characterization of pathogen effectors and host resistance proteins would supply us very valuable information. Results of these studies can be used to prevent diseases in crop plants either by conventional breeding strategies using these resistance genes as a molecular marker or by transgenic means to generate disease resistant varieties. Transfer of plant immunity receptors among the plant species would have a big potential to provide resistance against pathogens when we especially take the highly conserved recognition component nature of the pathogens into consideration. Additionally effector proteins of plant pathogens can be screened against wild relatives of the crops in order to discover new resistance sources.

REFERENCES

Ali, S., Rodriguez-Algaba, J., Thach, T., & et al. (2017). Yellow Rust Epidemics Worldwide Were Caused by Pathogen Races from Divergent Genetic Lineages. *Plant Disease*, 8: 1057.

Asai, T. a. (2002). MAP kinase signalling cascade in Arabidopsis innate immunity. *Nature*, 415, 977–983.

Ausubel, F. M. (2005). Are innate immune signaling pathwaysin plants and animals conserved? *Nature Immunology*, 973–979.

Bari, R., & Jones, J. D. (2009). Role of plant hormones in plant defence responses. . *Plant Mol. Biol.* , 69, 473–488.

Bentz H., B. H. (1978). Physical evidence for the assembly of A and B chains of human placental collagen in a single triple helix. *Eur. J. Biochem* (PubMed:738278), s. 92:563–567.

Bernado, P., & M., B. (2009). A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophys. J.*, 97: 2839-2845.

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

Bozkurt, T. O., Schornack, S., Banfield, M. J., & Kamoun, S. (2012). Oomycetes, effectors, and all that jazz. *Current Opinion in Plant Biology*, 15, 483–492.

Brandts, J. F., & Lin, L. N. (1990). Study of strong to ultratight protein interactions using differential scanning calorimetry. *Biochemistry*, 29(29), 6927–6940.

Cantu D., Segovia V., MacLean D., Bayles R., Chen X., Kamoun S., Uauy C. (2013). Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics.*, 14:270.

Chen et al., Z. (2013). ZincExplorer: an accurate hybrid method to improve the prediction of zinc-binding sites from protein sequences. *Molecular Biosystems*, 9(9):2213-22.

Chen Y, R. J. (1993). Mapping of the binding interfaces of the proteins of the bacterial phosphotransferase system, HPr and IIAglc. *Biochemistry*, 32(1):32-7.

Chen, W., Wellings, C., Chen, X. K., & Liu, T. (2014). Wheat stripe (yellow) rust caused by *Puccinia striiformis* f. sp. *tritici*. *Molecular Plant Pathology*, 15(5):433-46.

Chinchilla, et al., D. (2007). A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature*, 448, 497–500.

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

Chris A.E.M. Spronka, S. B. (2004). Validation of protein structures derived by NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 45, 315–337.

Christendat et al., D. (2000). Structural proteomics of an archaeon. *Nat. Struct. Biol.*, 7(10), 903–909.

Collier, S. M., & Moffett, P. (2009). NB-LRRs work a 'bait and switch' on pathogens. *Trends Plant Sci.*, 14, 521–529.

Copper, G. (2000). Protein Folding and Processing. G. Copper içinde, *The Cell: A Molecular Approach. 2nd edition.* Oxford University Press.

Cordeiro, T. N., Chen, P., De Biasio, A., Sibille, N., Blanco, F. J., Hub, J. S., Bernado, P. (2016). Disentangling polydispersity in the PCNA-p15PAF complex, a disordered, transient and multivalent macromolecular assembly. *Nucleic Acids Res.*, 45, 1362–4962.

Dagvadorj, B., Ahmet Caglar, O., Ayse, A., Cian, D., Tolga Osman, B., & Mahinur S., A. (2017). A *Puccinia striiformis* f. sp. *tritici* secreted protein activates plant immunity at the cell surface. *Nature/ Scientific Reports*, 7: 1141.

DeSantis, K. A., & Reinking, J. L. (2016). Use of Differential Scanning Fluorimetry to Identify Nuclear Receptor Ligands. *Methods in Molecular Biology. The Nuclear Receptor Superfamily.*, Vol 1443; pp 21-30.

Dobrovetsky et al., E. (2005). High-throughput production of prokaryotic membrane proteins. *J. Struct. Funct. Genomics*, 6(1), 33–50.

Dodds, N. P., & Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics*, 539.

Dodds, P. N., & Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics*, 11,539–548.

Dunker AK, S. I. (2008). Function and structure of inherently disordered proteins. *Current Opinion in Structural Biology*, 18(6):756-64.

Dyson HJ, W. P. (2004). Unfolded proteins and protein folding studied by NMR. *Chemical Reviewes*, 104(8):3607-22.

Ellis, J. G., Lawrence, G. J., Luck, J. E., & Dodds, P. N. (1999). Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. *Plant Cell*, 11, 495–506.

FAO. (2018). *www.fao.org*. November 21, 2018, http://www.fao.org/fao-stories/article/en/c/1154584/

Fishman, D. A. (2017). An introduction to circular dichroism spectroscopy. *www.chem.uci.edu*. October 18, 2017, University of California, Department of Chemistry: www.chem.uci.edu

Fishman, D. A. An introduction to circular dichroism spectroscopy. October 18, 2017, University of California, Department of Chemistry: www.chem.uci.edu

Flor, H. H. (1971). Current status of gene-for-gene concept. *Annu. Rev. Phytopathol*, 9, 275–296.

Fontana et al., A. (1997). Probing the partly folded states of proteins by limited proteolysis. *Folding and Design*, Volume 2, Issue 2, Pages R17-R26.

Foster MP, W. D. (1998). Chemical shift as a probe of molecular interfaces: NMR studies of DNA binding by the three amino-terminal zinc finger domains from transcription factor IIIA. *J Biomol NMR*, 12(1):51-71.

Franke, D., Kikhney, A., & D.I., S. (2012). Automated acquisition and analysis of small angle X-ray scattering data. *Nuclear Instrum. Methods Phys. Res. A*, 689:52-59.

Gay, N. J., & Gangloff, M. (2007). Structure and function of Toll receptors and their ligands. *Annu. Rev. Biochem.*, 76, 141–165.

Giehm, L., Svergun, D., Otzen, D., & Vestergaard, B. (2011). Low-resolution structure of a vesicle disrupting α-synuclein oligomer that accumulates during fibrillation. *Proc Natl Acad Sci USA.*, 108(8):3246-51.

Girardin, S. E., Philpott, D. J., & Lemaitre, B. (2003). Sensing microbes by diverse hosts. Workshop on pattern recognition proteins and receptors. *EMBO Rep.*, 4, 932–936.

Greenfield N., F. G. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry*, pp. 8:4108–4116.

Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, s. 1(6): 2876–2890.

Guinier, A. (1939). La diffraction des rayons X aux tr‘es petits angles: application ‘a l’étude de phénom‘enes ultramicroscopiquesed. *Les Ulis*, 11. Sér. 12:161-237.

Haydyn, M. D., & Dmitri, S. I. (2017). Combining NMR and small angle X-ray scattering for the study of biomolecular structure and dynamics. *Archives of Biochemistry and Biophysics*, 628: 33-41.

Heese, et al., A. (2007). The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. *Proc. Natl Acad. Sci. USA* , 104, 12217–12222.

Holzwarth G, D. P. (1965). The Ultraviolet Circular Dichroism of Polypeptides. *J. Am. Chem. Soc.*, pp. 87:218–228.

Hubbard, S. J. (1998). The structural aspects of limited proteolysis of native proteins. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, Volume 1382, Issue 2, Pages 191-206.

Jones, J. D., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444, 323–329.

Källberg et al., M. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature Protocols*, 7, 1511-1522.

Kamoun, S. (2006). A catalogue of the effector secretome of plant pathogenic oomycetes. *Annual Review of Phytopathology*, 44, 41–60.

Kelley et al., L. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10, 845-858 .

Kikhney, A. G., & Svergun, D. I. (2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters*, Vol:589, 2570-2577.

Kleckner IR, F. M. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta.*, 1814(8):942-68.

Korasick, D. A., & Tanner, J. J. (2018). Determination of protein oligomeric structure from small-angle X-ray scattering. *Protein Science*, VOL 00:00—00.

Ma et al., J. (2013). Protein threading using context-specific alignment potential. *Bioinformatics (Proceedings of ISMB)*, Vol. 29, Issue 13, pp. i257-i265.

Ma, Z., & et al. (2015). A Phytophthora sojae glycoside hydrolase 12 protein is a major virulence factor during soybean infection and is recognized as a PAMP. *Plant Cell*, 27, 2057–2072.

Marion, D. (2013). An Introduction to Biological NMR Spectroscopy. *Molecular & Cellular Proteomics* , 12.11 3006-3025.

Marley, J., Lu, M., & Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR*, 20: 71–75.

Mertens, H., & Svergun., D. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.*, 128-141.

Mojumdar, A., De March, M., Marino, F., & Onesti, S. (2017). The Human RecQ4 Helicase Contains a Functional RecQ C-terminal Region (RQC) That Is Essential for Activity. *J Biol Chem.* , 292(10):4176-4184.

Moldoveanu et al., T. (2001). Ca(2+)-induced structural changes in rat m-calpain revealed by partial proteolysis. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, Volume 1545, Issues 1–2, Pages 245-254.

Mylonas, E., & Svergun, D. I. (2007). Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Journal of Applied Crystallography*, 40, s245–s249.

Narayanan RL, D. U. (2010). Automatic assignment of the intrinsically disordered protein Tau with 441-residues. *Journal of the American Chemical Society*, 132(34):11906-7.

Niesen, F. H., Berglund, H., & Vedadi , M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*, 2(9), 2212-2221.

Pannetier N, H. K. (2007). Optimized 3D-NMR sampling for resonance assignment of partially unfolded proteins. *Journal of Magnetic Resonance*, 186(1):142-9.

Peng, J., & Xu, J. (2011a). A multiple-template approach to protein threading. *Proteins*, Jun;79(6):1930-9.

Peng, J., & Xu, J. (2011b). RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins*, Vol 79, Issue S10, pp. 161-171.

Perez, J., Vachette, P., Russo, D., Desmadril, M., & Durand, D. (2001). Heat-induced unfolding of neocarzinostatin, a small all-beta protein investigated by small-angle X-ray scattering. *J. Mol. Biol.*, 308:721-743.

Petoukhov, M., & Svergun, D. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.*, 89:1237-1250.

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 Phytologist*, 210, 627–42.

Privalov, P. L. (1979). Stability of proteins: small globular proteins. *Adv. Protein Chem.*, 33, 167–241.

Quevillon-Cheruel et al., S. (2007). Cloning, Production, and Purification of Proteins for a Medium-Scale Structural Genomics Project. *Macromolecular Crystallography Protocols*, 21-37.

Rivas, S., & Thomas, C. M. (2005). Molecular interactions between tomato and the leaf mold pathogen Cladosporium fulvum. *Annual Review of Phytopathology*, 43, 395–436.

Ryan, T., Xun, Y., Cowieson, N., Mata, J., Jackson, A., Pauw, B., McGillivray, D. (2016). Combined pressure and temperature denaturation of ribonuclease A

produces alternate denatured states. *Biochem Biophys Res Commun*, 473(4):834-839.

Saunders, D., & et al. (2012). Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS One*, 7, e29847.

Schellman, J. (1997). Temperature, stability, and the hydrophobic interaction. *Biophys. J.*, 73(6), 2960–2964.

Senisterra, G. A., & Finerty, P. J. (2009). High throughput methods of assessing protein stability and aggregation. *Molecular BioSystems*, 5, 217–223 |.

Shan, et al., L. (2008). Bacterial effectors target the common signaling partner BAK1 to disrupt multiple MAMP receptor-signaling complexes and impede plant immunity. *Cell Host Microbe*, 4, 17–27.

Shirasu, K., & Schulze-Lefert, P. (2003). Complex formation, promiscuity and multi-functionality: protein interactions in disease-resistance pathways. *Trends Plant Sci.* , 8, 252–258 .

Singh, R. P. (2008). *Advances in Agronomy* (Vol. 98).

Smilgies, D. M., & Folta-Stogniew, E. (2015). Molecular weight–gyration radius relation of globular proteins: a comparison of light scattering, small-angle X-ray scattering and structure-based data. *Journal of Applied Crystallography*, 48(Pt 5): 1604–1606.

Studier, F., Rosenberg, A., Dunn, J., & Dubendorff, J. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology*, 185:60-89.

Svergun, D. I., Koch, M. H., Timmins, P. A., & May, R. P. (2013). Small Angle X-ray and Neutron Scattering from Solutions of Biological Macromolecules. *Oxford University Press.*

Takken, F. L., & Tameling, W. I. (2009). To nibble at plant resistance proteins. *Science*, 324, 744–746.

Tameling, W. I. (2006). Mutations in the NB-ARC domain of I-2 that impair ATP hydrolysis cause autoactivation. *Plant Physiol.*, 140, 1233–1245.

Tao, Y. (2003). Quantitative nature of Arabidopsis responses during compatible and incompatible interactions with the bacterial pathogen Pseudomonas syringae. *Plant Cell* , 15, 317–330.

Tuukkanen, A. T., Spilotros, A., & Svergun, D. I. (2017). Progress in small-angle scattering from biological solutions at high-brilliance synchrotrons. *IUCrJ*, 4, 518–528.

Tuukkanen, A. T., Spilotros, A., & Sverguna, D. I. (2017). Progress in small-angle scattering from biological solutions at high-brilliance synchrotrons. *International Union of Crystallography Journal*, 4 (Pt 5): 518–528.

Tyler, B., & et al. (2006). Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science*, 313, 1261–6.

Uversky, V., & Dunker, A. (2010). Understanding protein non-folding. *Biochim. Biophys. Acta,*, 1804:1231-1264.

Valentini, E., Kikhney, A. G., Previtali, G., Jeffries, C., & Svergun, D. I. (2015). SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Research*, Volume 43, Issue D1, Pages D357–D363.

van der Hoorn, R. A., & Kamoun, S. (2008). From guard to decoy: a new model for perception of plant pathogen effectors. *Plant Cell*, 20, 2009–2017.

Varadi et al., M. (2014). pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Res.*, (Database issue); Epub 2013 Oct 29.

Vedadi et al, M. (2006). Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc. Natl. Acad. Sci.*, 103(43), 15835–15840.

Velluz, L., Legrand, M., & Grosjean, M. (1965). Optical Circular Dichroism: Principles, Masurements, and Applications. *Verlag Chemie Academic Press Inc*.

Venyaminov S, B. I. (1993). Circular dichroic analysis of denatured proteins: inclusion of denatured proteins in the reference set. *Anal. Biochem*, s. 214:17–24.

Wass, M., Kelley, L., & Sternberg, M. (2010). 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Research*, 38, W469-73.

Wenming, Zeng., & et al. (2013). High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nature Communications*, 4:2673.

Yin C., Chen X., Wang X., Han Q., Kang Z., & Hulbert SH. (2009). Generation and analysis of expression sequence tags from haustoria of the wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici*. *BMC Genomics.*, 10:626.

Zhang X., Farah N., Rolston L., Ericsson DJ., Dodds PN., & et al. (2017). Crystal structure of the Melampsora lini effector AvrP reveals insights into a possible nuclear function and recognition by the flax disease resistance protein P. *Molecular Plant Pathology*, 10.1111/mpp.12597.

Zipfel, C. (2008). Pattern-recognition receptors in plant innate immunity. *Curr. Opin. Immunol.* , 20, 10–16.

APPENDICES

A. pBSK/PstSCR1 PLASMID INFORMATION

PstSCR1 gene was made synthesized in pBSK vector for our MSA Plant Pathogen Interactions Laboratory as previously described in Dagvadorj, et al., 2017.

```
LOCUS GS50144_pBSK-PTSha2a5 3292 bp DNA circular SYN 18- OCT-2010
DEFINITION Ligation of inverted PTSha2a5 into modified pBluescript
ACCESSION GS50144_pBSK-PTSha2a5
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown
        Unclassified.
REFERENCE 1 (bases 1 to 3292)
        AUTHORS Self
        JOURNAL Unpublished.
COMMENT SECID/File created by SciEd Central, Scientific & Educational Software
COMMENT SECNOTES|Vector molecule: modified pBluescript
 Fragment ends: EcoRV
 Fragment size: 2857
 Insert molecule: PTSha2a5
 Fragment ends: blunt
 Fragment size: 435
FEATURES Location/Qualifiers
 CDS 21..327
 /gene="f1 origin"
 /SECDrawAs="Gene"
 misc_feature 600..643
 /gene="T7/M13F"
 /SECDrawAs="Region"
 CDS complement (460..615)
 /gene="LacZ"
 /SECDrawAs="Gene"
 CDS complement (656..1088)
 /gene="PTSha2a5"
 /SECDrawAs="Gene"
 misc_feature complement (1103..1159)
 /gene="T3/M13R"
 /product="T3 primer"
 /SECDrawAs="Region"
 misc_feature 1489..2156
/gene="pUC origin"
 /SECDrawAs="Region"
 CDS complement (2307..3164)
 /gene="Ampcillin"
 /SECDrawAs="Gene"
```

ORIGIN

ctgacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcg
cccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctcccttagggttccgat
ttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccc
tttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagg
gattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgcttacaatttgcc
attcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgc
tgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgaattgtaatacgactcactata
gggcgacccggggatattgcggccgcctaagatgctttggagcagttgtttgtccacatagcttgtttacttttctcaacttcaccaattcta
cctgcaacgtcacaacagaagctatctgtggcaaccgacccattcaagtccactttatcacacgagaatttcccgtcgactgtcttgatttc
ttccttccacatggtgaactccttgccaatctttttgcctttcgttcttcatcggtgattgatctggtgcaataaccatgtgtttggcttggcgtt
ccatgcaaaccgggacacttgaacgacttcacagaaatgaattgagtgttgatcaacaacactgcgaatacgatgaagaagttgaagct
ttgcttgtcgtcgtcgtcggaattctcgagaagcttgactttgtcatcgtcgtccttgtagtccatttaattaatcctaatcctcgaggttccctt
tagtgagggttaattgcgagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacga
gccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcg
ggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgct
cactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggg
gataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccatag
gctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgt
ttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggc
gctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagccc
gaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacag
gattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtat
ctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttt
gtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaa
aactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaa
gtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgc
ctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctca
ccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagt
ctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtca
cgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggtta
gctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcat
gccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccg
gcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaag
gatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtga
gcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttttcaatattat
tgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccc
cgaaaagtgccac

B. SEQUENCING RESULT OF pGEX-6P-1/ PstSCR1 PLASMID

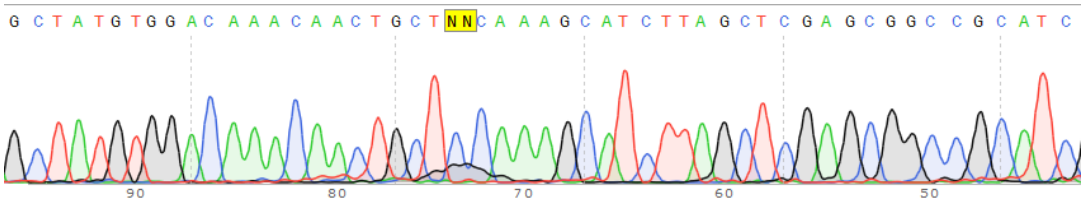


Figure B1. The chromatogram of the sequenced plasmid DNA pGEX-6P-1/ PstSCR1 The sequencing was carried out with pGEX forward primer.

```
i-P4_PGEX3.s    601 tttatacatggacccaatgtgcctggatgcgttcccaaaattagtttgttttaaaaaacgtattgaagctatcccacaaattgataagtacttgaaatcc
Pst2a5 no SS        ----------------------------------------------------------------------------------------------------

i-P4_PGEX3.s    701 agcaagtatatagcatggcctttgcagggctggcaagccacgtttggtggtggcgaccatcctccaaaatcggatctggaagttctgttccaggggcccc
Pst2a5 no SS        ----------------------------------------------------------------------------------------------------

i-P4_PGEX3.s    801 tgggatccccggaattcttcaagtgtcccggtttgcatggaacgccaagccaaacacatggttattgcaccagatcaatcaccgatgaagaacgaaaggc
Pst2a5 no SS      1 ----------------ttcaagtgtcccggtttgcatggaacgccaagccaaacacatggttattgcaccagatcaatcaccgatgaagaacgaaaggc

i-P4_PGEX3.s    901 aaaaaagattggcaaggagttcaccatgtggaaggaagaaatcaagacagtcgacgggaaatttctcgtgtgataaagtggacttgaatgggtcggttgcc
Pst2a5 no SS     84 aaaaaagattggcaaggagttcaccatgtggaaggaagaaatcaagacagtcgacgggaaatttctcgtgtgataaagtggacttgaatgggtcggttgcc

i-P4_PGEX3.s   1001 acagatagcttctgttgtgacgttgcaggtagaattggtgaagttgagaaaagtaaacaagctatgtggacaaacaactgctnncaaagcatcttagctc
Pst2a5 no SS    184 acagatagcttctgttgtgacgttgcaggtagaattggtgaagttgagaaaagtaaacaagctatgtggacaaacaactgct-ccaaagcatcttag---
```

Figure B2. Alignment of the sequencing result with the sequence of the *PstSCR1* gene. It is shown that cloning of *PstSCR1* into the pGEX-6P-1 vector is correct and in the frame (The C base in the non-readable NN region during sequencing can be seen in the following chromatogram.)

C. SAXS DATA

2D Charge Coupled Device (CCD) detector, Synchrotron Radiation (SR) SAXS profile and 1D data obtained from CCD are shown in Figure C1. Comparison of the SAXS profiles of the samples in three different concentartions are shown in Figure C2. Fit curve is shown in Figure C3 and Pair Distribution Function vs. Distance (Å) curve is shown in Figure C4.

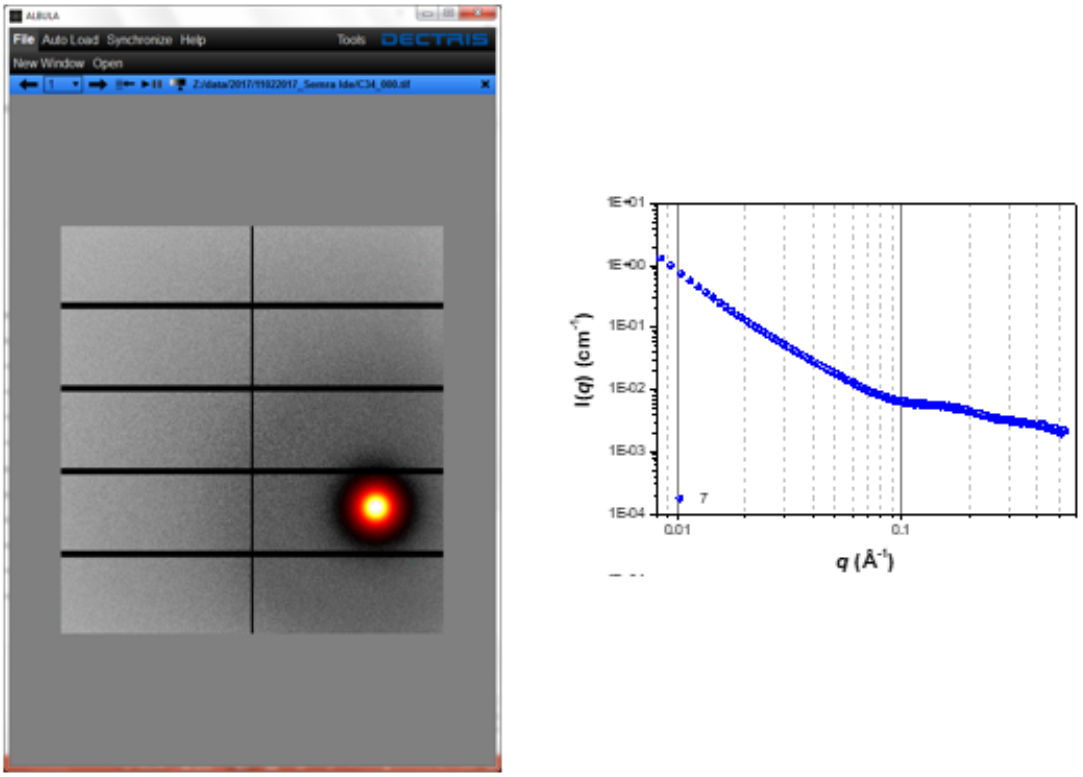


Figure C1. 2D CCD detector SR-SAXS profile and 1D data obtained from CCD.

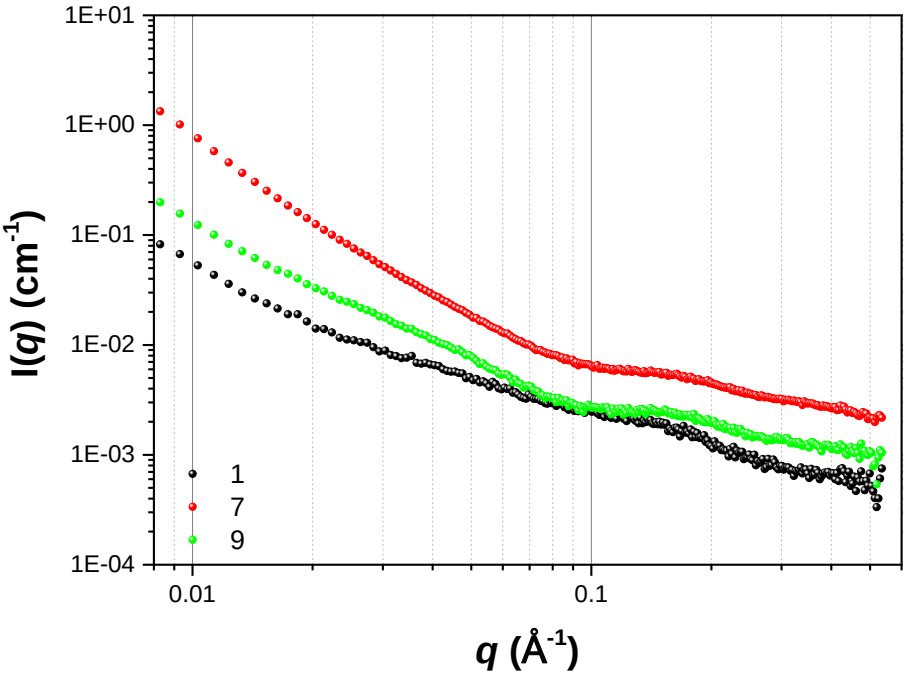


Figure C2. Comparison of the SAXS profiles of the samples in three different concentrations.

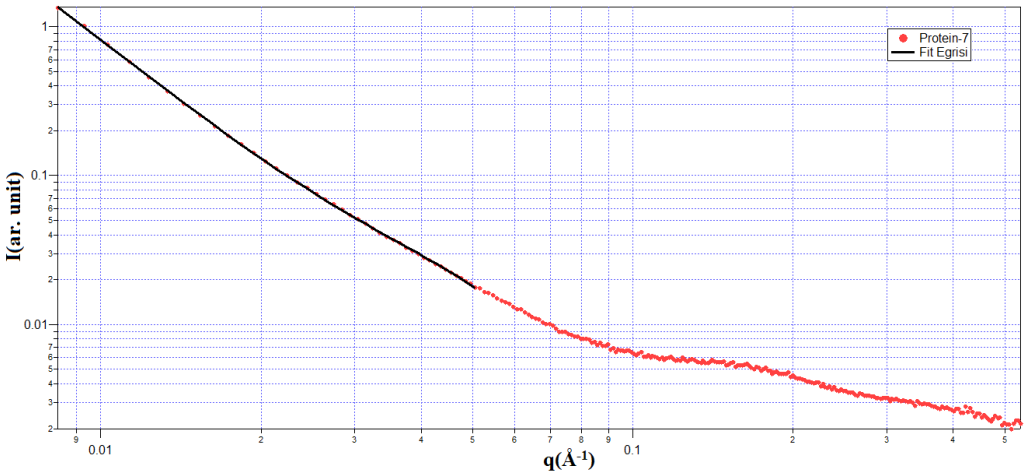


Figure C3. Fit curve

Fractal structure model
Maximum Extent = 686,11 Angstrom
Radius of Gyration = 23,158+- 1 Angstrom
Reduced Chi Squared = 3,34
The nano formations are arranged in the crystal configuration which has a distance d = 39.3 Å

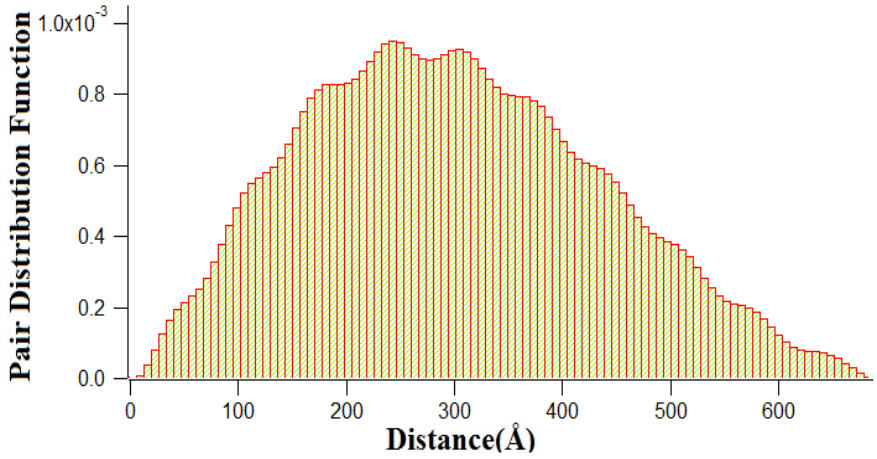
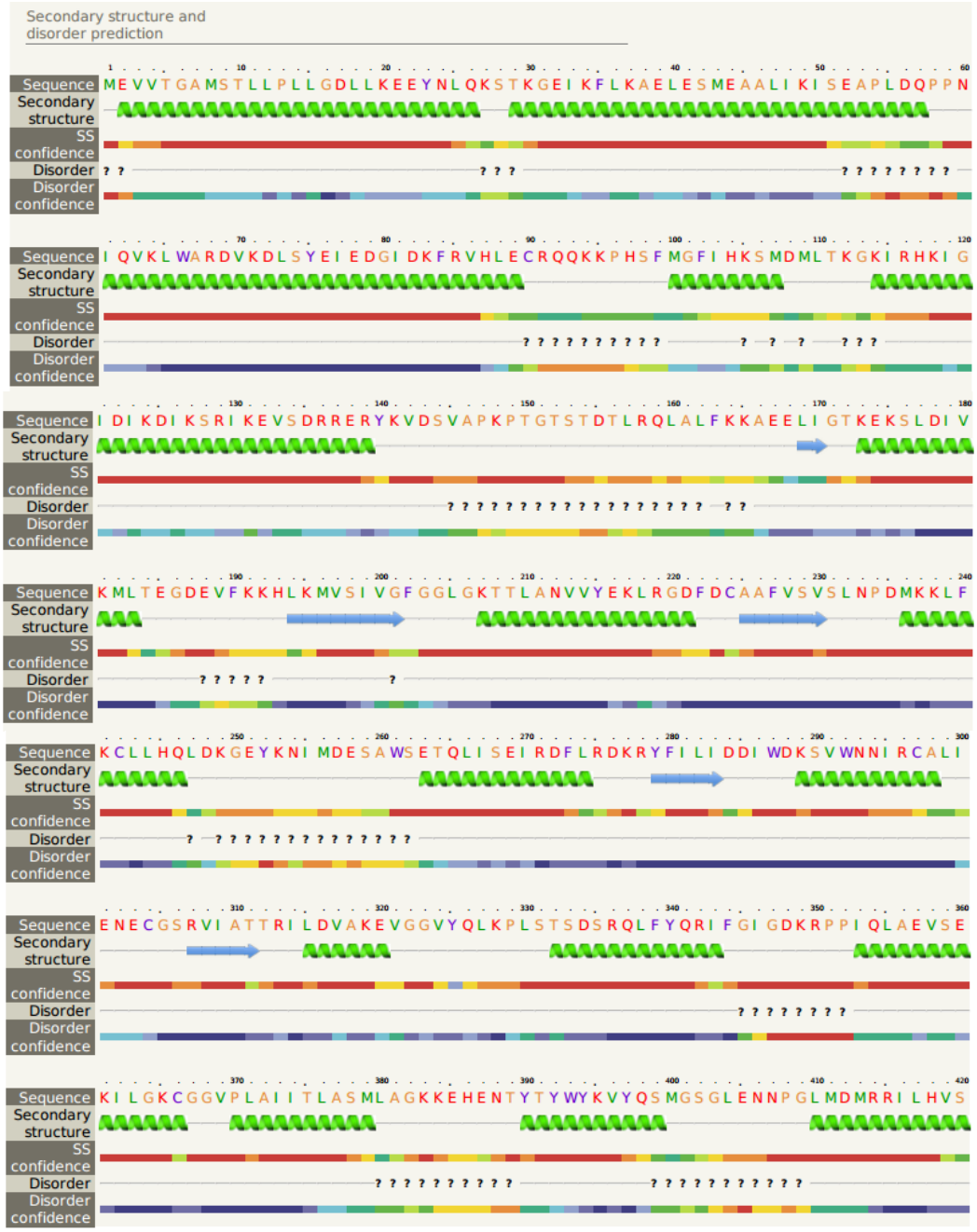


Figure C4. Pair Distribution Function vs. Distance (Å) curve

D. SECONDARY STRUCTURE AND DISORDER PREDICTION OF TaYr10



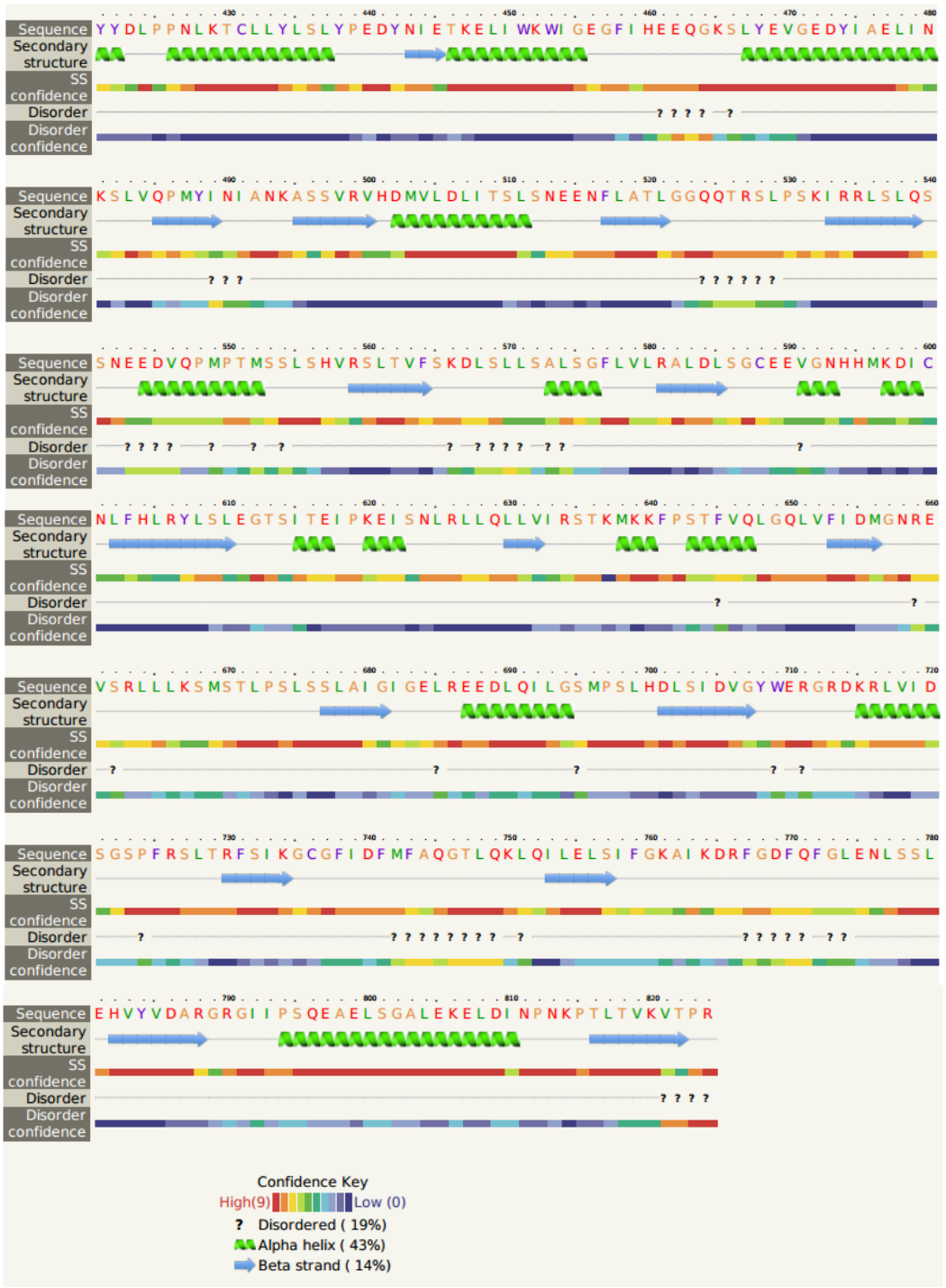


Figure D. Secondary structure prediction and disorder of TaYr10 predicted by Phyre2. Sequence shows the amino acid sequence of TaYr10. Secondary structure (SS) confidence is presented with the colours; the highest confidence with red and lowest confidence with violet colour. Disordered residues are expressed with a question mark (?), alpha helix is presented with green helix image and beta strand is presented with a blue arrow (Kelley et al., 2015) (Wass, Kelley, & Sternberg, 2010).

E. *Triticum aestivum* STRIPE RUST RESISTANCE PROTEIN YR10 mRNA

Gen Bank: AF149114.1

FASTA Graphics

Go to:

```
LOCUS       AF149114                2475 bp    mRNA    linear   PLN
23-DEC-2000
DEFINITION  Triticum aestivum stripe rust resistance protein Yr10
(Yr10) mRNA,
            complete cds.
ACCESSION AF149114
VERSION  AF149114.1
                     /product="stripe rust resistance protein Yr10"
                     /protein_id="AAG42168.1"
                     /translation="MEVVTGAMSTLLPLLGDLLKEEYNLQKSTKGEIKFLKAELESME
                  AALIKISEAPLDQPPNIQVKLWARDVKDLSYEIEDGIDKFRVHLECRQQKKPHSFMGF
                  IHKSMDMLTKGKIRHKIGIDIKDIKSRIKEVSDRRERYKVDSVAPKPTGTSTDTLRQL
                  ALFKKAEELIGTKEKSLDIVKMLTEGDEVFKKHLKMVSIVGFGGLGKTTLANVVYEKL
                  RGDFDCAAFVSVSLNPDMKKLFKCLLHQLDKGEYKNIMDESAWSETQLISEIRDFLRD
                  KRYFILIDDIWDKSVWNNIRCALIENECGSRVIATTRILDVAKEVGGVYQLKPLSTSD
                  SGQLFYQRIFGIGDKRPPIQLAEVSEKILGKCGGVPLAIITLASMLAGKKEHENTYTY
                  WYKVYQSMGSGLENNPGLMDMRRILHVSYYDLPPNLKTCLLYLSLYPEDYNIETKELI
                  WKWIGEGFIHEEQGKSLYEVGEDYIAELINKSLVQPMYINIANKASSVRVHDMVLDLI
                  TSLSNEENFLATLGGQQTRSLPRKIRRLSLQSSNEEDVQPMPTMSSLSHVRSLTVFSK
                  DLSLLSALSGFLVLRALDLSGCEEVGNHHMKDICNLFHLRYLSLEGTSITEIPKEISN
                  LRLLQLLVIRSTKMKKFPSTFVQLGQLVFIDMGNREVSRLLLKSMSTLPSLSSLAIGI
                  GELREEDLQILGSMPSLHDLSIDVGYWERGRDKRLVIDSGSPFRSLTRFSIKGCGFID
                  FMFAQGTLQKLQILELSIFGKAIKDRFGDFQFGLENLSSLEHVYVDARGRGIIPSQEA
                  ELSGALEKELDINPNKPTLTVKVTPR"
ORIGIN
        1 atggaggtcg tgaccggggc gatgagcacg ctcctgccct tgcttggcga cctgcttaag
       61 gaggagtaca acctgcagaa gagcaccaag ggtgagatca agttcctcaa agcagagctg
      121 gagagcatgg aggctgccct catcaagatc tcggaggcac ccttggatca gccacctaac
      181 attcaggtca agctctgggc gagggacgtc aaggacctgt cctatgagat cgaagatggc
      241 atcgacaaat tccgggtgca ccttgagtgc cggcaacaaa agaagccaca cagctttatg
      301 ggtttcatcc acaaaagcat ggacatgctg acaaagggca agatccgaca caaaataggc
      361 atcgatatca aagacatcaa gagccgcatc aaggaggtca gtgacaggcg tgaaaggtac
```

```
 421 aaggttgata gcgttgcgcc caagcccacc ggcacaagta ctgatacact tcgccagtta
 481 gccttgttca aaaaggcgga agagcttatt ggcaccaaag aaaagagcct tgacatagtc
 541 aagatgctga cggaaggaga tgaggtcttc aagaaacatc ttaagatggt ctctattgtt
 601 ggctttggag gcttagggaa gacaactctt gctaacgtgg tatatgagaa gcttcgcggg
 661 gactttgatt gtgcagcttt tgtctctgtg tctcttaatc ctgacatgaa gaagcttttc
 721 aagtgtttgc tccatcagct tgacaagggc gagtacaaga acatcatgga cgagtcagcg
 781 tggagcgaaa cacaactcat aagtgagata agagatttcc ttcgagacaa gaggtacttc
 841 attctcattg atgacatatg ggataaatct gtgtggaata atattagatg tgctctgatt
 901 gagaatgaat gtggtagtag agtaatcgca acaactcgca ttctagatgt tgccaaagaa
 961 gttggtggtg tttatcagct taaacctctt tctactagtg actcaggaca attattctat
1021 caaagaatat ttggaattgg cgacaagcgt cctcctattc agttggctga agtaagtgag
1081 aaaattttag ggaaatgtgg cggagtacca ttagctatca ttacactggc tagtatgttg
1141 gctggtaaaa aggaacatga aaatacatat acttattggt acaaggtgta ccaatctatg
1201 ggttctgggc tagaaaataa tcctggcctg atggacatga ggaggatact acatgtcagt
1261 tactatgacc tacctccaaa tctgaagact tgtttactgt atctcagttt gtatccagag
1321 gattataata ttgaaaccaa agagttgata tggaaatgga taggcgaagg attcattcat
1381 gaagagcaag ggaagagctt gtatgaagta ggagaggatt acattgctga gctcattaac
1441 aaaagcttgg tccaaccaat gtatatcaat attgctaata aggcgagctc tgtccgtgta
1501 cacgacatgg tgcttgacct tatcacttcc ttgtcaaatg aggagaactt tctcgcaaca
1561 ttgggtggtc agcagaccag gtcactgcca cgtaagatcc gtcgactgtc tctccaaagc
1621 agtaatgaag aggatgtcca gccaatgcca accatgagca gcttgtccca tgtgaggtca
1681 cttactgtgt tcagtaaaga cctcagtttg ctgtcggcac tttcaggttt tcttgtcctg
1741 cgtgcattgg atttaagtgg ttgtgaggaa gtgggtaatc atcacatgaa ggacatttgc
1801 aatttatttc acttgaggta tctgagtcta gaagggacat ctatcactga gatcccgaaa
1861 gagataagca atctacggct tctgcaattg ctagtcataa ggtccacaaa aatgaaaaaa
1921 tttccatcaa cctttgttca gctaggacaa ctggtgttca ttgacatggg taatagggaa
1981 gtctctagat tgcttctaaa atcaatgtcc accttgccct ccctctcttc acttgcaatc
2041 ggaataggag aactgagaga ggaagacctc caaatacttg ggagcatgcc gtctctgcat
2101 gacctctcta ttgatgtagg ttattgggaa agaggcagag ataaaaggct agtcattgac
2161 agtggttctc ccttccggtc tctgacaagg ttcagtataa agggttgcgg cttcattgat
2221 ttcatgtttg cacaaggaac cttgcaaaag ctccagatcc tggagttaag tatttttggt
2281 aaggcaataa aagacagatt tggtgatttt caatttgggc tggagaatct ctcatcactt
2341 gagcatgtct atgtggacgc tcgtggtcgt ggtataatcc caagtcaaga agctgagctg
2401 agcggtgcac ttgagaaaga gcttgatatt aatcccaaca agcccacact gacggtgaag
2461 gtaactccac gctga
//
```

F. ZINC BINDING ASSAY - ZnSO4 CONCENTRATIONS

Table 15. *DSF Analysis of PstSCR1 protein with different concentartions of Zinc ion to identify the zinc binding property of the protein.*

Well Number	$ZnSO_4$ Concentation
A-01	0
A-02	0
A-03	0
A-04	0
A-05	100 pM
A-06	100 pM
A-07	100 pM
A-08	100 pM
A-09	1 nM
A-10	1 nM
A-11	1 nM
A-12	1 nM
A-12	1 nM
B01	10 nM
B02	10 nM
B03	10 nM
B04	10 nM
B05	100 nM
B06	100 nM
B07	100 nM
B08	100 nM
B09	1 µM
B10	1 µM
B11	1 µM
B12	1 uM
C-01	10 µM
C-02	10 µM
C-03	10 µM
C-04	10 µM
C-04	10 µM

C-05	100 μM
C-06	100 μM
C-07	100 μM
C-08	100 μM
C-08	100 μM
C-09	1 mM
C-09	1 mM
C-10	1 mM
C-10	1 mM
C-11	1 mM
C-12	1 mM
C-12	1 mM
D01	10 mM
D01	10 mM
D02	10 mM
D02	10 mM
D03	10 mM
D04	10 mM
D04	10 mM
D05	100 mM
D06	100 mM
D07	100 mM
D07	100 mM
D08	100 mM

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Erdoğan, Sayıt Mahmut
Nationality: Turkish
Date and Place of Birth: 1 July 1978, Kırşehir
Marital Status: Single
Phone: +90 541 350 89 66
email: sayit.erdogan@metu.edu.tr

EDUCATION

Degree	Institution	Year of Graduation
MS	Philipps Universität Marburg and Max Planck Institute for Terrestrial Microbiology, Marburg, Germany	2010
BS	Hacettepe University, Department of Chemistry, Ankara	2002
BS	Anadolu University, International Relations	2015
High School	Kırşehir High School, Kırşehir	1997

WORK EXPERIENCE

Year	Place	Enrollment
2012-Present	Ministry of Agriculture and Forestry, Ankara	European Union Expert
2016 Dec.-2017 July	University of Houston, Department of Bio. and Biochemistry, Texas, USA.	Visiting Scholar
2018 June	Kasetsart University, Department of Food Science and Technology, Bangkok, Thailand.	Visiting Scholar
2018 Dec.-2019 Dec	University of Copenhagen, Department of Immunology and Microbiology, Copenhagen, Denmark	Post Doc. Researcher

FOREIGN LANGUAGES

Advanced English, Fluent German

AWARDS

2013, METU Thesis and Course Performance Awards "*Most successful PhD student of the Department of Biotechnology, Grad. School of Nat. and Appi Sci. - METU* "

PUBLICATIONS

1. S.M.Erdogan, M.Sen, S.Ide, M.S.Akkaya, Purification of *E. coli* expressed apoplastics effector of yellow rust wheat disease causing *Puccinia striiformis f. sp. tritici* and its structural analysis by SAXS and NMR, (Article in Preparation, 2019).

2. European Union Expertise Thesis on "Comparison of GMO Legislations, Commerce and Applications in the World and Turkey" written under the supervision of Prof. Dr. Mahinur S. Akkaya – METU, submitted to the Turkish Ministry of Food, Agriculture and Livestock, 2015, Ankara.

3. Master Thesis on "Isolation and characterization of type IV secretion system core complex of *Neisseria gonorrhoeae*" written under the supervision of Assoc. Prof. Dr. Chris van der Does in Max-Planck Institute for Terrestrial Microbiology, submitted to Philipps Universität Marburg, 2010, Marburg, Germany.

4. M. Sen, C. Uzun, O. Kantoglu, S. M. Erdogan, V. Deniz and O. Güven, Effect of gamma irradiation conditions on the radiation-induced degradation of isobutylene – isoprene rubber, Nuclear Inst. and Meth. Phys. Res. Section B. Beam Interact. with Materials and Atoms 208 (2003), pp. 480–484.

HOBBIES

Calligraphy, International Festival Movies.