Marta Grzechowiak

Arabidopsis thaliana proteins involved in signalling pathways: structural and functional studies

Thesis presented to the Scientific Council of the
Institute of Bioorganic Chemistry
Polish Academy of Sciences in Poznan
as a Ph.D. dissertation

Poznań 2015

The research described in this thesis has been carried out
at the Institute of Bioorganic Chemistry, Polish Academy of Sciences in Poznan
in Protein Engineering Laboratory
Department of Crystallography – Center for Biocrystallographic Research
under the supervision of Associate Professor Michal M. Sikorski, Ph.D., Dr.Sc.

Financial support for this work was provided
by the European Union within the European Regional Developmental Fund.
International PhD Programme was supervised by the Foundation for Polish Science.

In year 2015, the dissertation was partially supported by grant for Young Scientists provided
by the Polish Ministry of Science and Higher Education.

I would like to thank:
my advisor, Prof. Michal Sikorski
for support, care and understanding

Prof. Mariusz Jaskolski
for valuable comments

dr Miłosz Ruszkowski
for introducing me to the secrets of crystallography,
many scientific discussions and friendship

my Family
my Lab Mates
my Friends
for endless support, help and understanding

Table of content

Part II-Structural studies of enzymes involved in phosphate metabolism

Abbreviations

ABA abscisic acid

ACC 1-aminocyclopropane-1-carboxylic acid

APS ammonium persulfate

ADP adenosine diphosphate

ATP adenosine triphosphate

BSA bovine serum albumine

CASP Critical Assessment of Structure Prediction

CBVS calcium bond-valence sum

CD Circular Dichroism

DBD DNA-binding domain

DDM dodecyl β-D-maltoside

DLS Dynamic Light Scattering

DTT dithiotreitol

EMSA Electromobility Shift Assay

ETI effector-triggered immunity

FPLC Fast Protein Liquid Chromatography

FTIR Fourier Transform infrared spectroscopy

HR hypersensitive response

GST glutathion-S-transferase

IDP intrinsically disordered protein

IMAC Immobilized Metal Affinity Chromatography

IPTG isopropyl-D-thiogalactopyranoside

ITC Isothermal Titration Calorimetry

JA jasmonic acid

LB Luria Bertani

LIC ligation-independent cloning

LRRs leucine-rich repeats

MAMPs microbe associated molecular patterns

MBP maltose binding protein

MoRF molecular recognition feature

MPD 2-methyl-2,4-pentanediol

NMR nuclear magnetic resonance

NPS nitrate, phosphate, sulphate

NusA N-utilization substance A

PEG polyethylene glycol

PEG MME polyethylene glycol monomethyl ether

PCR polymerase chain reaction

PDB Protein Data Bank

P phosphorus

Pi phosphate

PNP methyl-phosphonic acid mono-(4-nitro-phenyl) ester

PPase inorganic pyrophosphatase

PPi pyrophosphate

PIPE polymerase incomplete primer extension

PR pathogenesis-related proteins

PRRs pattern recognition receptors

PTI pattern-triggered immunity

PVDF polyvinylidene fluoride

RMSD root mean square deviation

SA salicylic acid

SAR Systemic Acquired Resistance

SAXS Small Angle X-Ray Scattering

SDS sodium dodecyl sulfate

SEC Size Exclusion Chromatography

STS Static Light Scattering

TB terrific broth

TEMED tetramethylethylenediamine

TEV Tobacco Etch Virus

TF transcription factor

TCEP tris(2-carboxyethyl)phosphine

TLS translation/libration/screw

TRX thioredoxine

Preface

This dissertation describes several novel findings concerning structural biology of plants. It is focused on proteins that regulate transcription reprogramming during biotic and abiotic stress conditions and proteins that are indirectly involved in signal transduction. Studies on this proteins were performed mainly with the use of biomolecular crystallography, biophysical methods as CD, DLS, SLS and they were also characterized by various *in vitro* assays and bioinformatics predictions.

Signal transduction occurs when an extracellular signaling molecule activates a specific receptor located on the cell surface or inside the cell. In turn, this receptor triggers a biochemical chain of events inside the cell, initiating a response of stress reactions. Depending on the cell, the response causes changes in expression of certain genes, metabolic processes and triggers cell division or apoptosis. The signal can be amplified at any step of stress response and thus, one signaling molecule can cause many responses. The signal transduction in plants involves different receptor proteins, transcription factors and enzymes such as kinases, phosphatases, apyrases and phytohormones. Transmission of signals rarely is direct, usually it is multistage process and requires the participation of many different proteins simultaneously. The chain reaction allowing the plant response to environmental signals or internal signals generated within the organism, leads to physiological, morphological and developemental changes in the individual cells, tissues and in the whole plant and supports the maintenance of homeostasis.

WRKY transcription factors belong to a large family consisting of 74 proteins and regulate plant responses to pathogens and abiotic stress like salinity, heat, drought or wounding. They manage multiple enzymatic processes and affect hormone levels necessary for proper functioning of plant. While the processes regulated by individual WRKY proteins have been identified, for many of them, still little is known about their structure. So far only the crystal and NMR structures of the DNA-binding domains were solved. My attention was focused on structural studies of AtWRKY50 protein, a positive regulator of abscisic acid signaling pathway and a repressor of signaling *via* jasmonic acid. All attempts to obtain crystals and the crystal structure of AtWRKY50 and AtWRKY18^{DBD} failed. Also trials to obtain its complex with DNA were unsuccessful. Therefore, I applied biophysical methods: circular dichroism (CD) and complementary bioinformatics sequence analyses to characterize secondary

structure of protein of interests. I also used biophysical methods such as electromobility shift assay (EMSA) and isothermal titration calorimetry (ITC) to test DNA binding ability of AtWRKY50 and DNA-binding domain from AtWRKY18. Results of the mentioned analyses allowed me to determine the partially disordered nature of the AtWRKY50 and confirm the DNA-bindig activity of both full length AtWRKY50 and AtWRKY18^{DBD}.

Among my research interests are also enzymes involved in the metabolism of phosphate. Apyrases that remove the rest of the diphosphate from NTP and pyrophosphatases that hydrolyze diphosphate to phosphate making it available for further transformations, closing the phosphorus cycle in the cell. In plants there are several homologues of pyrophosphatases and apyrases. In *A. thaliana*, there are five homologous pyrophosphatases (based on the genome sequencing) and two apyrase homologs.

In my research work, I solved the crystal structure of recombinant pyrophosphatase from *A. thaliana* (AtPPA1). This is the first 3D model of plant pyrophosphatase. The biologically active form of AtPPA1 forms a trimer in contrast to homologous yeast *S. cerevisiae* pyrophosphatase forming dimer and *E. coli* forming hexamer. Structural studies were performed using X-ray crystallography. The diffraction data were collected using synchrotron radiation facility. This protein has been solved at high resolution at 1.93Å.

Due to my broad scientific interests I divided the thesis into two main parts. First part is dedicated to structural studies of AtWRKY transcription factors and the second one describes structural studies of enzyme that hydrolyzes inorganic pyrophosphate. Second part dedicated to structural studies of WRKY transcription factors is divided into four chapters. The first, *Introduction*, provides biological background of the subject. The second part, *Materials and Methods*, presents briefly all techniques used within this thesis, including protein expression, purification and all the methods used for structural and functional characteristic of the protein of interests. *Results* summarizes all experimental outcome. *Discussion*, focuses on a very comprehensive structural analysis of the ultimate results.

Part II is divided into three chapters: *Introduction, Results and Discussion* and *Materials and Methods*. The last describes only main techniques and the background of basic information about protein crystallography. I also decided to include results and discussion in one chapter because in this case it allows to describe the structural information including many comparisons to other structures more clearly avoiding repetition.

List of publications related to this thesis

Grzechowiak M, Sikorski M, Jaskolski M (2013) Inorganic pyrophosphatase (Ppase) from higher plant. BioTechnologia 94, 35-37.

Crystal structure of Inorganic Pyrophosphatase PPA1 from Arabidopsis thaliana- crystal structure deposited in Protein Data Bank under the accession code **4LUG**

Part I

Structural studies of WRKY transcription factors

1. Introduction

1.1. WRKY Transcription Factors

The WRKY transcription factors were broadly investigated in plants for more than 20 years. The first report about WRKY transcription factor SPF1 from sweet potato (*Ipomoea batatas*) revealed its role in induction of gene expression by sucrose [82]. The initial reports on WRKYs also defined their potential involvement in regulation of ABF1 and ABF2 genes expression during germination [147]. In one of the first reports on regulation of parsley response to pathogen, the name WRKY (pronounced 'worky') family was created, together with identification of the other WRKY proteins: WRKY1, WRKY2 and WRKY3 [149]. Since then, enormous progress in this field was achieved. Recently, an access to genome sequencing programs allowed to identify a putative WRKY proteins in different plant species as well as many members of this family have been cloned and characterized. Moreover, using system biology approaches such as transcriptomic and promoter analyses allows to define the WRKYs function in signaling network. Last years brought subsequent progress towards the understanding of WRKYs function in many distant physiological and developmental processes revealed a complex network of their relationships.

1.1.1. Distribution among species

Since their first discovery in sweet potato (*Ipomea batata*) multiple genes for WRKY transcription factors have been experimentally identified from more than 80 other plant species [91], including *Arabidopsis thaliana*, tobacco (*Nicotiana tabacum)*, wild oats (*Avena fatua*), rice (*Oryza sativa*), parsley (*Petroselinum crispum*), barley (*Hordeum vulgare*), wheat (*Triticum aestivum*), soybean (*Glycine max)*, potato (*Solanum tuberosum*), orchardgrass (*Dactylis glomerata*), chamomile (*Matricaria chamomilla*), sugarcane (*Saccharum*), cotton (*Gossypium arboreum*), grape (*Vitis vinifera*), poplar *(Populus trichocarpa*), sorghum (*Sorghum bicolor*) and coconut (*Cocos nucifera*). Most reports refer to angiosperm plants but WRKY were reported also from gymnosperm *Pinus monticola* [116]. Recently, some members of the WRKY family were also identified by searching all available sequence data from lower plants such as ferns (*Ceratopteris richardii*) and mosses (*Physcomitrella patens*). Homologues of WRKY genes were found only in two non-photosynthetic species: slime mold *Dictyostelium discoideum* closely related to the lineage of animals and fungi and also in

unicellular protist *Giardia lamblia,* a primitive eukaryote and the green algae *Chlamydomonas reinhardtii,* an early branching of plants. The WRKY proteins are a large superfamily of transcription factors. WRKY genes have been identified from a various plants as mentioned above and the number of them range from a single WRKY gene copy in the unicellular green alga *Chlamydomonas reinhardtii,* through 37 in the moss *Physcomitrella patens,* 74 in *Arabidopsis thaliana,* at least 109 in rice *Oryza sativa [184]* and over 230 in soybean *Glycine max* [194]. WRKY genes identified in the *Arabidopsis* genome by sequence similarity comparisons are a single copy randomly distributed over the five chromosomes. WRKY proteins vary in molecular weight from 14,3 kDa (AtWRKY43) to 210,3 kDa (AtWRKY19) [78]. The number of WRKY genes varies in different species and increases during the evolution of plants. WRKY family shows evolution from simpler to more complex multicellular organisms, demonstrating the ancient origin of the gene family. Comparing to fern, moss and pine, in flowering plants evolutionary expansion of WRKY gene family occurs. The ancestral WRKY gene seems to be duplicate many times, resulting in a large family in evolutionarily advanced flowering plants. It has been proposed that this expansion has associated with the increasing complexity of the body plants and development of highly sophisticated defense mechanisms adapted against pathogens.

To date the WRKY genes have been cloned only from plant species although a genome sequence data for species representing several major eukaryotic lineages are already available. There is still no evidence for presence of WRKY TF in animal kingdom. The absence of WRKY homologues in the animal genomes i.e. *Caenorhabditis elegans* and *Drosophila melanogaster* and *Saccharomyces cerevisiae* may suggest that WRKY transcription regulators are restricted to the plant kingdom.

1.1.2. Structural features and classification of WRKY proteins

The characteristic feature of WRKY transcription factors is their DNA binding domain known as the WRKY domain. There are about 60 amino acids region with characteristic almost invariant amino acid sequence Trp-Arg-Lys-Tyr-Gly-Glu-Lys (WRKYGQK) at its amino-terminal end and with a putative zinc-finger motif at its carboxy-terminal end. In a few representatives of WRKY proteins from rice (*Oryza sativa*), the consensus WRKY amino acid sequences have been replaced by WRRY, WSKY, WKRY, WVKY or WKKY suggesting that

W(R/K)(K/R)Y might be considered as a new consensus WRKY motif [187].

All known WRKY proteins contain either one or two WRKY domains and unique among all already described zinc-finger-like motifs. Despite of the strong conservation of their DNA-binding domain, the overall sequence homology of the WRKY proteins outside this conserved region, is low. Some WRKY transcription factors can be large and have a number of additional domains, others are slightly larger than the highly conserved DNA-binding domain, which is common in all WRKY transcription factors. Therefore the 74 *Arabidopsis thaliana* WRKY proteins were initially classified into three main groups and five subgroups on the basis of the number and type of their WRKY domains, differences within their zinc-finger motif and presence of additional characteristic features (Fig. 1). Members of group I contain two WRKY domains, while most proteins which possesses single WRKY domain belong to group II or III. Generally, the WRKY domains of group I and group II members have the same type C2–H2 of the zinc-finger motif with sequence pattern C–$X_{4–5}$–C–$X_{22–23}$–H–X1–H. In members assigned to group III, the WRKY domains contain a C2–HC zinc finger motif with sequence pattern C–X_{5-8}–C–X_{25-28}–H–X_{1-2}–C. Additionally group II splits up into five distinct subgroups (IIa–e). This classification is based on the presence of ten additional structural motifs that are conserved among the different subsets of the *At*WRKY family members. Each of these motifs is unique for certain subgroup. In some cases, these motifs can reveal clues about their potential functions. They seems to be nuclear localization signals, phosphorylation or calmodulin binding sites or allow protein dimerization which is characteristic for leucine zippers (LZs). A few AtWRKY proteins (AtWRKY10 and AtWRKY38, and AtWRKY52) do not fit precisely to any of previously established group. For example AtWRKY10 posses only one WRKY domain more related to group I. It might be a result of the secondary loss of the N-terminal WRKY domain. Moreover, two others AtWRKY 38 and AtWRKY52 could either belong to group III or represent members of a novel group when we take into consideration a pattern of Cys and His residues within their WRKY domains. AtWRKY52 posses also leucine-rich repeat (LRR) characteristic for R proteins.

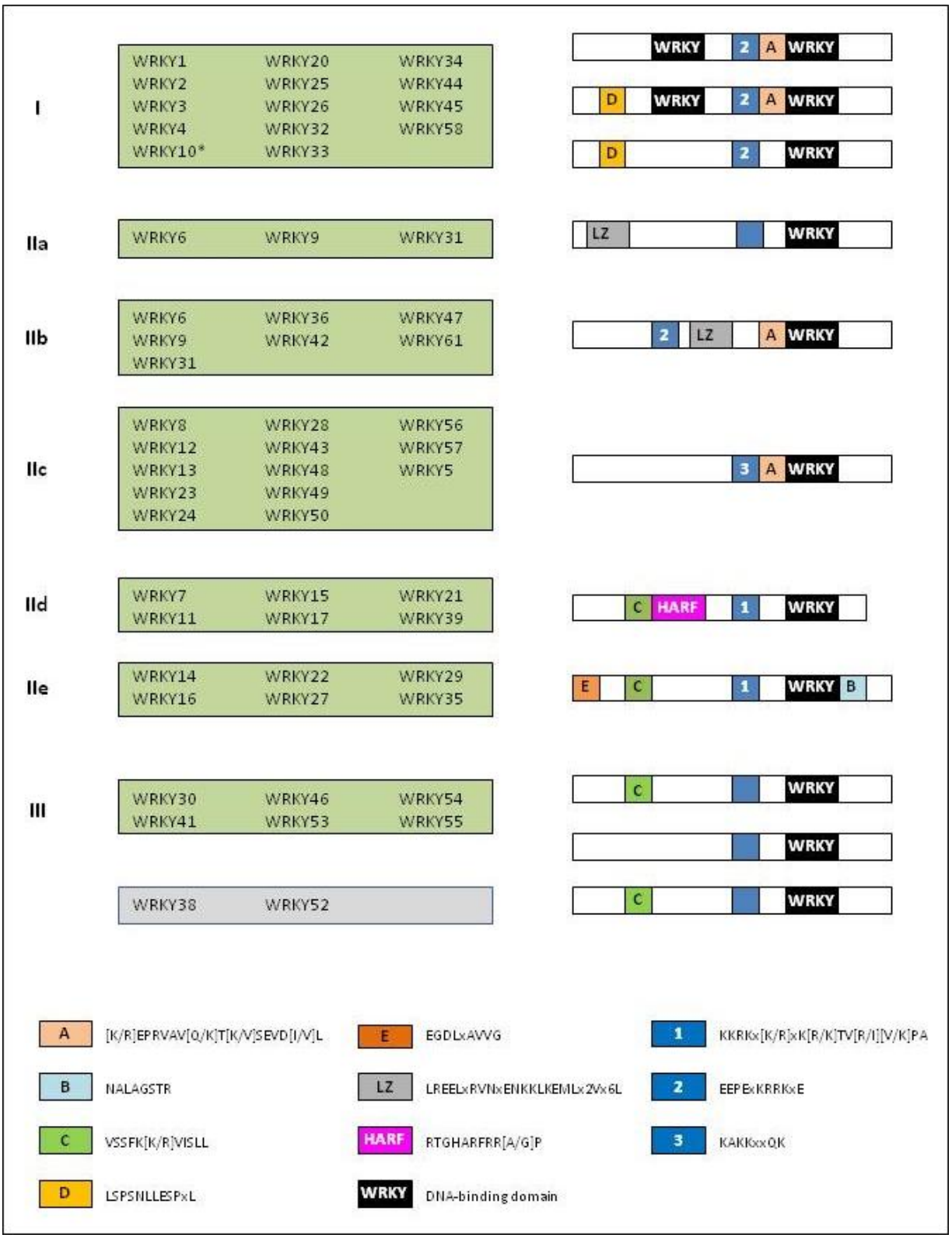


Fig.1. Classification of AtWRKY Transcription Factors Family members, according to Eulgen et al. 2000. [56]

Nevertheless, *in vivo* and *in vitro* experiments proved that members of all three groups of WRKY proteins possess clear binding preference for the same DNA sequence termed "W-box element" (TTGACY, where Y is C or T) found in the promoter region of a large number of plant target genes [31, 39, 57, 147, 149, 180, 192]. The W-box elements contain invariant TGAC core, which is essential for function and WRKY binding [56, 122]. Functional W boxes frequently cluster in short promoter regions and act synergistically [57]. Both the WRKY domains as well as the zinc finger motif are required for proper DNA-protein binding [122]. The two WRKY domains of group I appear to be functionally distinct and interestingly the C-terminal WRKY domain but not the N-terminal domain in I group representatives, is responsible for the DNA-binding activity [39, 57, 82]. The function of the N-terminal WRKY domain remain unclear. Presumably it might participate in the binding process increasing the affinity or specificity of these proteins for their target sites or it might provide an interface for protein-protein interactions. Unexpectedly, the single WRKY domains of group II and III family members are more similar in sequence to the C-terminal than to the N-terminal WRKY domain of group I proteins (Fig. 2), suggesting that C-terminal and single WRKY domains are functionally equivalent and constitute the major DNA-binding activity. Moreover the C-terminal WRKY domain sequence is supposed to be the ancestral type of WRKY gene because of its presence in primitive organisms such as protists or mosses.

Fig. 2. Comparison of WRKY domain sequences from representatives of different groups of *At*WRKY Transcription Factors. Gaps shown as dots have been inserted for optimal alignment. Residues that are highly conserved are in red and residues that binds zinc are highlighted in red boxes.

Despite differences in zinc fingers motives between groups I, II and III experimental evidences have shown that members of all three groups bind specifically to various W-box elements. Experiments with use of metal-chelators such as o-phenantroline and EDTA abolished DNA binding and the inhibitory effect was relieved when Zn^{2+} was further added to the protein. Other metal cations such as Mg^{2+}, Cu^{2+}, Fe^{2+} or Cd^{2+} were ineffective and confirmed that Zn^{2+} is required for DNA binding activity [122]. Few researches have described substitutions of the conserved cysteine and histidine residues to alanine in the C_2H_2-type zinc finger-like motif in the WRKY domain. This replacement abolished the DNA-binding activity because the domain is stabilized by Zn^{2+} cation chelated by two cysteines occurring respectively at the end of strand 2 and at the beginning of strand 3 and the two conserved histidines occurring at the end of strand 4 what indicates that this structural motif is crucial for DNA binding [122]. Similarly, mutations within the consensus/invariable WRKYGQK sequence at the N-terminal side of the zinc finger-like motif also significantly reduced the DNA-binding activity. The mutation experiments have shown that the replacement of each of the conserved residues: Trp, Arg, two Lys, Tyr, and Gly to Ala significantly decrease or almost completely abolished the DNA-binding activity. These amino acid residues play important role in stabilization of correct structure and are critical for maintaining a DNA-protein interactions [122] [51]. Those experiments were finally confirmed by solved structure of AtWRKY4 domain in complex with DNA (PDB:2LEX) suggesting that each of these residues together with Zn^{2+} cations are required for proper folding of the DNA-binding zinc finger and its binding activity.

1.2. Biological function

It is common for a single WRKY transcription factor to regulate transcriptional reprogramming associated with various biological processes. Studies carried out on different plants indicate that WRKY proteins are involved in regulation of biotic or abiotic stress responses [148] as well as plant development. The first experimentally confirmed function was that WRKY proteins play essential role in regulation of plant responses to pathogens as transcription factors. Many WRKY proteins are involved in the defence mechanism against attack of pathogenic bacteria [20, 24, 40, 41, 48, 50, 61], fungi [24, 30, 81, 152], viruses [22, 24, 180, 192] and oomycetes [24, 94, 129]. Furthermore, WRKY proteins are upregulated

upon the abiotic stress of wounding [27, 72, 129], salinity [7, 9, 17, 79], drought [7, 136, 143, 179], heat [35, 143], cold [81, 136], H_2O_2 effect [173] and UV radiation [85]. Some members of the family are implicated in other processes that are unique to plants such as morphogenesis of trichomes and seeds [92], senescence [24, 76, 144, 145], dormancy [136], growth [20], starch [162], lignin [68] antocyan biosynthesis [92] and also metabolic pathways [92, 147, 157, 162, 183]. Moreover single WRKY transcription factor might be involved in regulating several apparently disparate plant processes. A single WRKY gene often responds to several factors evenly as negative or positive regulator however, they might also regulate expression themselves. They have been isolated from different plants, but still, the role of individual representatives in regulating transcriptional reprogramming is not well characterized. It is due to cross-talk and very complex relationship between particular representatives.

1.2.1. The plant immune system

Plants are exposed to two types of stress: biotic and abiotic. Biotic stress is caused by parasitic microorganisms (viruses, bacteria, fungi), by other plants through overcrowding, alelopathy, or by trampling and gnawing animals. Plants become infected by pathogens of different lifestyles. Biotrophic pathogens are specialized to feed on living plant tissues and they have narrow host range. Additionally various strains of these pathogens have often adapted to a specific line of given plant species. Many biotrophs live in the intercellular space between leaf mesophyl cells and some produce haustoria. Necrotrophic pathogens are less specialized and they grow on plant tissues that are wounded, weakened or senescent. They frequently produce toxins to kill host tissue prior to colonization. Abiotic stress factors are naturally occurring, often intangible factors that may cause harm to the plants. The most basic stressors include: drought, wounding, salinity, extreme temperatures, H_2O_2 effect and UV radiation, as well as more extreme such as natural disasters: flood, tornadoes and wildfires. Abiotic stress is essentially unavoidable. Stress factors induce changes in plant hormone homeostasis, which can cause programmed cell death. Genetic basis of this mechanism is still poorly understood. Therefore studies of molecular basis of plant resistance to stress can contribute to a more resistant plants.

Plants have developed a highly complex immune system that enables them to respond to pathogen infection or environmental stress. Plants, unlike mammals, lack mobile defender

cells. Without the adaptive immune system, they rely on the innate immunity of each cell and on systemic signals originating from infection sites to defend against most potential pathogens.

Based mainly on studies with the model plant *Arabidopsis thaliana*, two branches of plant's innate immune system are currently distinguished: pattern-triggered immunity (PTI) and effector-triggered immunity (ETI), depending on the manner by which pathogens are recognized [47].

PTI is a type of plant innate immunity that is triggered upon the identification/recognition of microbe associated molecular patterns (MAMPs) through the corresponding pattern recognition receptors (PRRs) localized mainly in plasma membrane. MAMPs are common molecular structures characteristic of microbes that are not found in host cells. Both non-pathogenic and pathogenic microbes produce effective MAMPs to activate immune responses. Specific receptors with extracellular leucine-rich repeats (LRRs) subsequently transduce signal through MAP-kinase cascades, ultimately leading to the primary defense response. *A. thaliana* recognizes a variety of MAMPs including most characterized flagellin (flg22), lipopolisaccharide (lps) and elongation factor Tu (elf18) originated from bacteria or fungal chitin and β-glucan [165, 175, 199, 200]. Plants also respond to other factors such as small molecules (ATP) and cell wall or cuticular fragments. The first identified and best studied PRR is the flagellin receptor FLS2. It consists of the N-terminal signal peptide, 28 LRRs, a transmembrane domain, and a cytoplasmic kinase domain. In Arabidopsis, it perceives a minimal motif of 22 amino acid residues of the flagellin protein of bacterial flagella (flg22). Binding of flg22 to corresponding receptor FLS2 results in endocytosis of the complex. The internalization of endosome is kinase dependent and relies on the PEST motif that is related to ubiquitinylation. Upon MAMPs recognition the first line of defense is achieved and leads to range of defense responses and reprogramming of whole metabolism including activation, suppression, and modulation of various signalling pathways in plant cells which prevent further pathogen expansion. Then, cell wall modification, callose deposition and accumulation of defense-related proteins are initiated. Such processes negatively affect colonization of pathogens. PTI is an ancient conserved first layer of plant innate immune response. To successfully grow and proliferate on their host, virulent pathogens have to override the first line of defense. Plants do not have an adaptive immune system to eliminate pathogens that

have entered their intercellular spaces and vascular systems. PTI is therefore effective against a broad spectrum of invading microorganisms but is relatively weak immune response. Moreover plant pathogens are able to break or suppress basal defense activated in the primary innate immune system. They successfully proliferate on host plants and cause disease by producing effectors.

The second type of immunity involves recognition of pathogen virulence molecules called effectors by intracellular receptors. This induces effector-triggered immunity (ETI). ETI is result of co-evolution between pathogens and plants. Viral, bacterial, fungal and oomycete pathogens evolved secreted effectors targeting key PTI elements to interfere with plant defense. Some plants have evolved resistance (R) proteins to directly or indirectly detect these effectors named avirulence or Avr proteins. ETI is a faster and stronger version of PTI that often culminates in hypersensitive response (HR) being a form of programmed cell death. The hypersensitive response is a mechanism, that prevent the spreading of infection to other parts of the plant. The HR caused the rapid death of cells in the local region surrounding an infection. HR cell death typically may retard or stop pathogen growth in some interactions, particularly those involving haustorial parasites. The resulting necrotic lesions are one of the first visible manifestations of defense responses and are thought to aid the confinement of the pathogen to the dead cells. HR is not always observed, nor required for ETI. Particularly the mechanism of HR is initiated by activation of R genes that triggers ion flux and accumulation of reactive oxygen species (ROS), superoxide anions, hydrogen peroxide, hydroxyl radicals and nitrous oxide that induce lipid peroxidation and membrane damage. HR actually causes disease resistance by depriving the incoming pathogen of nutrients or by releasing compounds from dying cells which are destructive to microbes. For a subset of effectors, the mechanism of suppression has been elucidated. The *Pseudomonas syringae* effector AvrPto promotes infection in susceptible plants and abolish responses elicited by MAMPs. AvrPto binds receptor kinases, including *Arabidopsis* FLS2 and EFR, to block plant immune responses in the plant cell. The ability to target receptor kinases is required for the virulence function of AvrPto in plants. This model illustrates the dynamic coevolution between plants and pathogens [29]. Apart from suppressing hypersensitive response (HR), some plant pathogens produce small molecule effectors that mimic plant hormones. Pathogenic bacteria *P. syringae* AvrPtoB also induces production of coronatine, a jasmonic acid (JA) analogue that suppresses salicylic

acid-induced defense responses to biotrophic pathogens. It induces stomatal opening, helping pathogenic bacteria to gain access to the apoplast. Fungal pathogen of rice *Gibberella fujikuroi* produce plant hormone that cause hypertrophy, etiolation and chlorosis. Plants are infertile with empty panicles, producing no edible grains ("foolish seedling desease"). Cytokinin produced by many pathogens can promote pathogen success through retardation of senescence in infected leaf tissue. The interplay between PTI and normal ETI is qualitatively stronger, faster and often involves a localized cell death called the hypersensitive response (HR) [36]. PTI is generally effective against non-adapted pathogens in a phenomenon called non-host resistance, whereas ETI is active against adapted pathogens. However these relationships are not exclusive and depend on the elicitor molecules present in each infectious pathogen. Extreme diversification of ETI receptors and pathogen effectors within and between species is common.

Besides local immune responses, PTI and ETI activate long-distance defense reactions such as systemic acquired resistance (SAR) which predispose plants to become more resistant to subsequent pathogen attacks [128]. In *Arabidopsis thaliana* and other higher plants, local and systemic defense responses are controlled by the balanced action of distinct, but partially interconnected pathways involving several phytohormones, including salicylic acid (SA), jasmonic acid (JA) and ethylene (ET). In general SA signaling sectors are essential for resistance toward biotrophic and hemibiotrophic pathogens wheras the JA and ET sectors are important for immunity toward necrotrophs.

Systemic acquired resistance (SAR) is a mechanism of induced defense that confers long-lasting protection against a broad spectrum of microorganisms. SAR requires the signal molecule (salicylic acid) and is associated with accumulation of pathogenesis-related proteins (PR proteins), which are thought to contribute to resistance. Up to date, PR proteins have been classified into 17 families [170] based on their biological role and/or physicochemical properties (sequence similarity, molecular mass, isoelectric point). The biological functions of most classes of the defense proteins have been recognized, including chitinases, β-glucanases, peroxidases and protein inhibitors [171]. Some of them produce antimicrobial metabolites with a crucial role in induced plant disease resistance. The role of some PR proteins, including PR-10, in defense response still remains to be elucidated. In response to SA, the positive regulator protein NPR1 moves to the nucleus where it interacts with TGA transcription factors

and induces defense gene expression, thus activating SAR.

1.2.2. The role of WRKY transcription factors in plant defense

Extensive studies have demonstrated that plant WRKY transcription factors play important roles in the two branches of the plant innate immune system, PTI and ETI.

Studies using knockout or knockdown mutants or overexpession lines of WRKY genes have shown that WRKY TF can positively or negatively regulate various aspects of plant PTI and ETI. It was also well established that those regulators rarely act alone. Functional redundancy causes difficulties how to link specific WRKY with definite process. For example, AtWRKY70 protein acts as an integrator of cross-talk between SA and JA in plant defense responses [156]. It functions as activator of SA-dependent defense genes and a repressor of JA-regulated genes. Moreover, AtWRKY70 is required for both, basal defense and full R-gene mediated disease resistance against the oomycete *Hyaloperonospora parasitica* [99], bacteria *Erwinia carotovora* and *Pseudomonas syringae* [48] as well as the fungi *Erysiphe cichoracearum* [111]. Recent publications have provided conclusive genetic proof that *Arabidopsis* WRKY proteins are crucial regulators of the defense responses against both biotrophic and necrotrophic pathogens. For example, disruption of AtWRKY33 enhances susceptibility to the necrotrophic fungal pathogens *Botrytis cinerea* and *Alternaria brassicicola* [197]. Further investigation showed that AtWRKY33 physically interacts with genes involved in redox homeostasis, SA signaling, ethylene-JA mediated cross-communication, camalexin biosynthesis and thus is a key transcriptional regulator of hormonal and metabolic responses towards *Botrytis cinerea* infection [13]. Functional analysis based on T-DNA insertion mutants and transgenic overexpression lines indicates that AtWRKY3 and AtWRKY4 also function as positive regulators in plant resistance against *B. cinerea [107],* similarly to At*WRKY8* [23]. Several WRKY factors act as negative regulators of resistance. For instance, basal plant resistance triggered by avirulent *P. syringae* strain was enhanced in *Atwrky*7 and *Atwrky*11/*Atwrky*17 insertional mutants [93]. Likewise, disruption of AtWRKY38 or AtWRKY62 enhances plant basal defense against *P. syringae*. Overexpression of AtWRKY38 or AtWRKY62 reduces disease resistance and also PR1 expression, thus they function additively as negative regulators of plant basal defense [97].

A recent study suggests that AtWRKY51 may have function as a positive regulator of basal

defense against *P. syringae* [60]. In addition, AtWRKY25 and AtWRKY72 were also shown as regulators in response to biotrophs *Pseudomonas syringae* and *Hyaloperonospora arabidopsidis* [12, 196], whereas three representants of small subgroup IIa of WRKY genes, comprising AtWRKY18, AtWRKY40, and AtWRKY60, play important functions in regulating plant disease resistance toward *P. syringae, B. cinerea* and *Golovinomyces orontii* infection. Functional analysis of single, double, and triple combinations of *wrky*18, *wrky*40 and *wrky*60 mutants for response to microbial pathogens indicated that AtWRKY18, AtWRKY40, and AtWRKY60 proteins have partially redundant roles in activating defense to the fungal necrotroph *Botrytis cinerea* and repressing basal resistance to a virulent strain of the bacterial biotroph *Pseudomonas syringae* [188]. These three WRKY transcription factors play complex and antagonistic roles in plant disease resistance. *Arabidopsis* WRKY22 and WRKY29 are induced by a MAPK pathway that confers resistance to both bacterial and fungal pathogens and expression of WRKY29 in transiently transformed leaves led to reduced disease symptoms [6]. Two additional WRKY factors, AtWRKY53 and AtWRK58 were identified as modulators of SAR and act as positive and negative regulators respectively [176]. Furthermore, the AtWRKY52 representative of group III that posses an atypical structural feature - zinc finger motif, was shown to confer resistance toward the bacterial pathogen *Ralstonia solanacearum.*[41]. It combine a typical for R-proteins nucleotide binding leucine-rich repeat (NB-LRR) and Toll/interleukin-1 receptor (TIR) domain with WRKY domain.

These results indicate that the WRKY proteins interact functionally in a complex pattern of overlapping, antagonistic, and distinct roles in plant responses to different types of microbial pathogens. Above, there were mentioned only few examples of AtWRKY function in plant immunity to indicate complexity of this subject. More detailed information are included in Table 1.

1.2.3. The role of WRKY transcription factors in abiotic stress

Plants are unable to move and therefore they are simultaneously subjected to different stress factors. Adoptation of plants to unfavourable environmental changes involve a series of complex physiological and biochemical mechanisms. Moreover plants responses to abiotic stress condition are very diverse among species. Also single representatives of the same species, even from a plant living in the same area respond uniquely. There is no universal

defense response although some common mechanisms can be elucidated. Recent studies have demonstrated that many WRKY genes play roles in responses not only to biotic stresses but also to certain abiotic stresses such as wounding, drought, cold, heat or salinity/osmotic stress. Compared to the research on biotic stress, little is known about the involvement of these TFs in abiotic stress responses. A single WRKY protein is often involved in several stress responses, and some of them are even involved in abiotic and biotic stresses. Cross-talk between signaling networks involved in the responses to biotic and abiotic stress is very complex. Furthermore, it is ambiguous to distinguish which stress response is associated with a particular WRKY.

Microarray profiling/analyses of the *A. thaliana* root transcriptome revealed induction of 18 WRKY genes and repression of 8 WRKY genes under salinity stress. In other microarray experiments AtWRKY6 and AtWRKY75 were among the 27 transcripts elevated at least five-fold in data sets related to oxidative stress response [59]. Similarly, *Arabidopsis* WRKY18, WRKY40 and WRKY60 proteins were reported to respond in a complex pattern not only to pathogens but also to salt and osmotic stress [21].

In recent years, numerous groups have demonstrated that manipulation of WRKY TF levels in knockout or overexpressor plants affects specific stress responses. Two closely related AtWRKY25 and AtWRKY33 respond to heat, drought and osmotic stress [90]. The *wrky*25 mutants exhibited deficient thermotolerance at different stages of growth, while AtWRKY25 overexpressing plants displayed enhanced thermotolerance compared to the wild-type plants [113]. Furthermore, an earlier study showed the induction of WRKY25 during oxidative stress [142]. Thus AtWRKY25 is involved in various stress responses. In other work, the *AtbHLH17* and *AtWRKY28* TF genes which are known to be upregulated under drought and oxidative stress in *Arabidopsis* were expressed. The transgenic lines showed enhanced tolerance to NaCl, mannitol, and oxidative stress. Under mannitol stress condition also a higher root growth was observed [7]. These examples demonstrate that the WRKYs might be powerful tools for the generation of drought resistance plants.

WRKY might enhance cold or heat tolerance. The WRKY34 transcription factor negatively mediated cold sensitivity of mature *Arabidopsis* pollen. Otherwise, functional analysis indicated that the WRKY34 transcription factor was also involved in pollen development.

Mature pollen is very sensitive to cold stress in chilling-sensitive plants. AtWRKY34 gene might be involved in pollen viability, although the mechanism is unclear. Cold treatment increased AtWRKY34 expression in wild-type plants and promoter-GUS analysis revealed that At*WRKY34* expression is pollen-specific [201].

Arabidopsis WRKY39 provides an evident example for a TF that is involved in heat acclimation of plants. Heat-treated seeds and *wrky*39 knockdown mutants had increased susceptibility to heat stress, showing reduced germination, decreased survival and elevated electrolyte leakage compared to wild-type plants. Additionally, *At*WRKY39 overexpressing plants exhibited enhanced thermotolerance compared to wild-type plants [114]. WRKY also participate in tolerance to micro and macro nutrients deficiency. *At*WRKY6 and *At*WRKY42 are involved in *Arabidopsis* responses to low phosphate stress through regulation of *PHOSPHATE1* (*At*PHO1) gene expression [25]. Moreover transcriptome analysis around the root tip identified AtWRKY6 to be essential for normal root growth under low boron conditions [95].

WRKYs also participate in responses to wounding. Two wounding-responsive WRKY3 and WRKY6 genes were identified in tobacco *Nicotiana attenuata*. Moreover, *Na*WRKY3 is required for *Na*WRKY6 elicitation by fatty acid-amino acid conjugates from the larval oral secretions that are released into the wounds during feeding. Silencing either WRKY3 or WRKY6, or both, by stable transformation makes plants highly vulnerable to herbivores and is associated with impaired accumulation of jasmonates. This observations indicate an important role of WRKY3 and WRKY6 in sustaining active JA levels during continuous insect attack [159].

1.2.4. The role of WRKY transcription factors in other processes

In the past few years, there is increasing evidence that WRKY proteins actively participate in certain plant developmental and physiological processes such as trichome developement [92], seed germination, senescence [76, 145], fruit maturation and carbohydrate metabolism [162]. Finally, biosynthesis of anthocyanin [92], starch [162], and sesquiterpene [189] are also dependent on WRKY proteins.

Expression of root genes in *A. thaliana* was mapped and obtained gene expression pattern indicated a possible specialized role for 12 members of WRKY TF family in the root cell

maturation [14]. AtWRKY44 is the first member of the WRKY family involved in morphogenesis of trichomes. AtWRKY44 is presumed to have a role in non-hair epidermis development, due to its preferential expression in differentiating non-hair cells [92].

Several WRKY genes from different plant species are expressed during different stages of seed development. The WRKY gene *DGE1* of orchardgrass (*Dactylis glomerata*) is expressed during somatic embryogenesis [1]. Similarly, *Sc*WRKY1 gene, is strongly and transiently expressed in fertilized ovules at the late torpedo stage in wild potato and posses a specific role during embryogenesis [105]. In barley, *SUSIBA2* is expressed in the endosperm and regulates starch production [162]. Likewise, *Arabidopsis* WRKY10, also known as MINISEED3, is expressed in pollen, globular embryo as well as in developing endosperm from the 2-nuclei stage through the cellularization stage. Furthermore, WRKY genes may control seed germination and postgermination in rice. OsWRKY71 encodes a transcriptional repressor of GA signal transduction in aleurone cells [195].

AtWRKY44 plays additional role in mucilage and tannin synthesis in seed coat and is expressed in seed integument or endosperm. Experiments with w*rky*44 mutants showed that, they were defective in proanthocyanidin synthesis and seed mucilate deposition thus seeds were yellow colored and their size was reduced when the mutant allele was transmitted through the female parent [92]. AtWRKY18 and AtWRKY60 have a positive effect on plant ABA sensitivity for inhibition of seed germination and root growth. On the other hand, AtWRKY40, antagonizes AtWRKY18 and AtWRKY60 effect [21].

WRKY participate in carbohydrate metabolism. AtWRKY45 and AtWRKY65 are involved in regulating genes which respond to carbon starvation [33]. Three rice WRKY genes are also upregulated in sucrose-starved rice suspension cultures [178]. Furthermore, sugar regulates the expression of the *Arabidopsis NUCLEOSIDE DIPHSOPHATE KINASE 3a* (*NDPK3a*) gene. NDPK3a is located in mitochondria because sugar metabolism is intricately connected with this organelle through the conversion of sugars to ATP, and through the production of carbon skeletons that can be used in anabolic processes. Regarding the *NDPK3a* gene, glucose induction is decreased in the *wrky*34 mutant, while sucrose induction is increased in the *wrky*4 mutant. AtWRKY4 and AtWRKY34, are involved in the sugar regulation of the *NDPK3a* gene exerting opposite effects [69].

In cotton plant *Gossypium arboreum*, sesquiterpene phytoalexins are secondary metabolites

induced and by fungal and bacterial infection or other environmental stimuli. They accumulate in epidermal and subepidermal cells of roots. *Ga*WRKY1 is a transcriptional activator of the *CAD1* gene participating in cotton sesquiterpene biosynthesis. [189]

Table 1. List of WRKY transcription factors and its fuction.

Gene	Induction factor	Function	Ref.
AtWRKY1	SA	defense response, SAR	[51]
AtWRKY2	NaCl, mannitol	negative regulator in ABA signaling, regulation of seed germination and post germination growth	[87, 88]
AtWRKY3	*Botrytis cinerea*, SA, JA, ACC	positive role in plant resistance to necrotrophic pathogens	[107]
AtWRKY4	*Pseudomonas syringae*, SA, JA, sucrose, senescence, cold, salinity	negative effect on plant resistance to biotrophic pathogens, carbohydrate metabolism	[69, 107]
AtWRKY6	H_2O_2, methyl viologen, Pi and B starvation	negative regulator in low Pi stress and positive regulator in low B stress	[25, 95]
AtWRKY7	SA, *Pseudomonas syringae*	negative regulator of plant defense against *P. syringae*	[96]
AtWRKY8	NaCl, wounding, *Pseudomonas syringae*	salinity stress tolerance, repressor of plant PTI signaling, defense response against TMV-cg	[22, 61, 79]
AtWRKY10		seed developement	[120]
AtWRKY11	*Pseudomonas syringae*	negative regulator of basal resistance toward *Pseudomonas syringae*, regulation of JA-dependent responses	[93]
AtWRKY17	*Pseudomonas syringae*, NaCl	negative regulator of basal resistance toward *Pseudomonas syringae*, regulation of JA-dependent responses, NaCl tolerance	[93] [89]
AtWRKY18	ABA, SA, *Pseudomonas syringae*, *Botrytis cinerea*	ABA signaling, NaCl and mannitol tolerance, regulation of defense response to bacteria and fungi, resistance to *Pseudomonas syringae*	[20, 21, 152, 154]
AtWRKY22	H_2O_2, dark, chitin, flagellin	regulation of dark-induced senescence, resistance to pathogens	[6, 175, 198]
AtWRKY23	*Heterodera schachtii*, auxin	resistance to nematode, stem cell specification	[66, 67]
AtWRKY25	*Pseudomonas syringae*, ABA, ethylene, NO, NaCl, mannitol, cold, heat	tolerance to heat and NaCl, increased sensitivity to oxidative stress and ABA, negative regulator of defense response to *Pseudomonas syringae*	[90, 112, 113]
AtWRKY26	heat	heat tolerance, dehydration stress	[112]
AtWRKY28	NaCl, mannitol, H_2O_2	dehydration, salt and oxidative stress	[7]
AtWRKY29	chitin, flagellin, *Pseudomonas syringae*	defense response	[6, 175]
AtWRKY30	H_2O_2, ozone, SA	abiotic stress tolerance, regulation of senescence	[11, 151]
AtWRKY33	NaCl, mannitol, cold, heat, H_2O_2, ozone, UV, chitin, *Botrytis cinerea*, *Pseudomonas syringae*, *Alternaria brassiciola*	heat and NaCl tolerance, redox homeostasis, resistance to *Botrytis cinerea and Pseudomonas syringae* , SA signaling, ethylene-JA-mediated cross-communication, camalexin biosynthesis	[13, 90, 112, 175, 197]

Gene	Induction factor	Function	Ref.
AtWRKY34	cold, sucrose	cold tolerance, carbohydrate metabolism, pollen developement	[69, 201]
AtWRKY38	chitin, SA, *Pseudomonas syringae*,	negative regulator of plant basal defense, regulation of HR	[69, 97]
AtWRKY39	heat, drought	tolerance to heat, dehydration stress	[45, 201]
AtWRKY40	ABA, SA, chitin, wounding, *Pseudomonas syringae*, *Botrytis cinerea*	ABA signaling, defense response, thermotolerance	[21, 113, 118, 154, 155]
AtWRKY41	*Pseudomonas syringae*, *Erwinia carotovora*	resistance to *Pseudomonas syringae*, susceptibility to *Erwinia carotovora*, regulator in the cross talk of salicylic acid and jasmonic acid pathways	[75]
AtWRKY42	Pi starvation	Pi deficiency stress	[25]
AtWRKY44		proanthocyanidin synthesis , seed mucilate deposition, seed coat development,	[92]
AtWRKY45	Pi starvation	Pi deficiency stress	[177]
AtWRKY46	heat, NaCl, K starvation, *Pseudomonas syringae*	thermotolerance, osmotic stress, K deficiency stress, basal pathogen resistance	[45, 80, 113, 127]
AtWRKY48	*Pseudomonas syringae*	repressors of plant PTI signaling	[61]
AtWRKY50	*Botrytis cinerea*	SA- and low 18:1-dependent repression of JA signaling.	[60]
AtWRKY51	*Botrytis cinerea*	SA- and low 18:1-dependent repression of JA signaling.	[60]
AtWRKY52	SA, *Ralstonia solanacearum*	resistance to Ralstonia solanacearum	[41]
AtWRKY53	Chitin, flagellin, *Pseudomonas syringae*, SA, H2O2, wounding	tolerance to oxidative stress, regulator of SAR and basal pathogen response, leaf development, senescence	[45, 80, 175, 186]
AtWRKY54	H_2O_2	oxidative stress, negative regulator of leaf senescence	[11]
AtWRKY58		regulator of SAR	[176]
AtWRKY60	NaCl, SA, *Pseudomonas syringae, Botrytis cinerea*	salt and osmotic stress, ABA signaling, defense response	[21, 118, 154, 155]
AtWRKY62	*Pseudomonas syringae*	negative regulator of plant basal defense	[97]
AtWRKY63	water deficiency, ABA	positive regulator in drought tolerance, negative regulator in ABA signaling	[140]
AtWRKY65	Fe starvation	Iron deficiency stress	[78]
AtWRKY70	H_2O_2, *Hyaloperonospora parasitica, Erysiphe cichoracearum, Pseudomonas syringae, Erwinia carotovora*	response to reactive oxygen species, activator of SA-dependent defense genes and a repressor of JA-regulated genes, basal and full R-gene mediated pathogen resistance, negative regulators of leaf senescence	[11, 48, 99, 111, 157]
AtWRKY72	oomycete *Hyaloperonospora arabidopsidis*	basal defense response	[12]
AtWRKY75	Pi starvation	positive regulator in Pi starvation	[42]

1.3. The WRKY interactions

Transcriptional gene expression regulation is very complex. The gene expression programs that maintain specific cell states are controlled by thousands of transcription factors, cofactors, and chromatin regulators. Transcriptional regulation tends to involve combinatorial interactions between several transcription factors, which allow sophisticated response to multiple conditions in the environment. This is associated with the harmonious modulation of a large number of different proteins that directly interact with DNA but also require participation of other regulatory elements indirectly influencing gene expression. WRKY similarly to other regulatory proteins rarely work alone and interact transiently or permanently with proteins that play role in transcription and chromatin remodelling, signalling and other cellular processes. WRKY were classified into 3 large groups and 5 subgroups. Slight variations within DNA-binding domains and other sequence motifs conserved within each group participate in protein-protein interactions and mediate complex functional interactions between WRKY and other factors that posses regulatory and modulatory effect. Among partners interacting with WRKY TF the following proteins were identified: MAP kinases, MAP kinase kinases, 14-3-3 proteins, calmodulin, histone deacetylases, resistance proteins and other WRKY transcription factors [28].

1.3.1. WRKY-WRKY interactions

The WRKY promoters are statistically enriched with W-box elements and this observation suggest functional linkage of many WRKY genes by auto- and cross-regulatory mechanisms. Thus WRKY proteins provide dynamic regulation of target genes by cooperation or antagonism. The extensive protein-protein interactions were found within members of the same subclass, but also between members of different subclasses. In *A. thaliana* three members of group IIa (WRKY18, WRKY 40, WRKY60) interact through the leucine zipper motifs present at the N-terminal end. Interestingly, *in vitro* assays shown that hetero-complexes AtWRKY18/AtWRKY40 may have enhanced regulatory activities comparing to homodimers composed of one of the mentioned WRKYs. Furthermore AtWRKY60 alone has little DNA-binding activity for W-box sequences but could enhance the binding of AtWRKY18 to DNA in contrast to reduction of AtWRKY40 DNA-binding activity. This phenomena may have role in controlling intensity of basal defense response [188]. Moreover

AtWRKY40 and AtWRKY60 interact with AtWRKY36 (group IId) and AtWRKY38 (group III) as revealed by yeast two hybrid assay [32]. Within group IIb, AtWRKY6 and AtWRKY42 interact with each other [25]. Similar examples are interactions of AtWRKY30 with 3 others members of group III (AtWRKY53, AtWRKY54 and AtWRKY70) [11]. Analysis of the WRKY sequence draw attention to multiple leucine/isoleucine/valine residues at circa seven residue intervals. This is not the canonical leucine zipper but it seems to be responsible for dimer formation through hydrophobic interactions. There are two more possible mechanisms of WRKY-WRKY interactions considering DNA organisation. W-boxes that are recognized by WRKY proteins very often are clustered and separated by short spaces. Interacting WRKY may bind the closely-spaced W-boxes and regulate the target gene cooperatively and antagonistically. If the W-boxes are separated by substantial number of nucleotides, then the same WRKY complex may interact through DNA loop formation. Furthermore this mechanism could affect the binding of other TF.

1.3.2. WRKY-VQ interactions

WRKY transcription factors interact with proteins containing a conserved FxxxVQxLTG motif with two residues: valine (V) and glutamine (Q). There are 34 genes encoding proteins which posseses VQ motif in *A. thaliana*. They are relatively small, 100-200 amino-acid in length. The sequence beyond the short conserved motif with VQ residues is very diverse but as showed by yeast two-hybrid assay, all of these 34 VQ proteins are capable to interact with WRKY proteins [26]. *A. thaliana* WRKY protein, members of group I and group IIc are able to interact with VQ motif. Analysing the amino acid sequences of the C-terminal WRKY domain of group I and the single WRKY domain of group IIc, the conclusion is that these two groups share similar structural features that are part of interface for interaction with short VQ motif. The two aspartate residues preceding WRKYGQK motif and four residues interfering with two cysteines engaged in zinc finger are essential for interaction with VQ motif. What is interesting, the interaction is not restricted and single WRKY protein may interact with several VQ proteins. For example AtWRKY25 and AtWRKY33 may interact with majority of VQ proteins with varying degrees while AtWRKY51 interact with about 50% of all tested VQ proteins [26]. Within VQ proteins are: MKS1 (MAP kinase substrate1) interacting with AtWRKY33 [2, 138], HAIKU1, responsible ifor endosperm growth and seed size, that interact

with AtWRKY10 [120] and SIB1 (SIGMA factor interacting protein1) that enhance plant defense against necrotroph [106].

1.3.3. WRKY-MAP-kinase interactions

MAPKs (Mitogen-Activated Protein Kinases) plays crucial roles in plant response to pathogens and environmental stress conditions. Majority of WRKY transcription factors are also engaged in response to various stresses. Functional analyses indicate that among substrates identified as stress responsive MAPKs are WRKY TF from group I. These WRKY posses two WRKY domains and contain clustered proline-directed serines (SP clusters) that are postulated to be potential phosphorylation sites for MAPKs [83]. MAPK may phosphorylate also WRKYs from other groups, suggesting recognition of other phosphorylation sites. Some members of group I proteins contain MAPK-docking site named the D-domain with the cluster of basic residues upstream of LxL motif [83]. Diversity of MAPK interacting sites may force selectivity of their interactions with WRKY [83]. For example AtWRKY 33 interact with MKS1, a substrate for MPK4. More detailed analyses showed that in the absence of pathogens, MPK4 is presented in nucleus in complex with AtWRKY33 and AtWRKY is released when infection occurs [138]. AtWRKY33 is also up-regulated by the MPK3/MPK6 cascade and therefore played role in regulation of pathogen-induced camalexin biosynthesis [124].

1.3.4. WRKY interactions with other proteins

There are evidences for existence of other binding partners for WRKY transcription factors. They belong to various protein families.

Yeast two-hybrid screens identified *Arabidopsis* HDA19 (Histone Deacetylase 19) as an interacting partner of both AtWRKY38 and AtWRKY62 [97]. The interaction occurs in nucleus and is highly specific. Histone deacetylase removes acetyl groups from histones. Deacetylated histones have ability to wrap the DNA more tightly. Deacetylation of histones leads to repression of genes transcription. Overexpression of HDA19 represses the AtWRKY38 and AtWRKY62 activity as transcriptional activators.

Another binding partner for WRKY is calmodulin (CaM-Calcium Modulated Protein). CaM is a multifunctional intermediate messenger protein that transduces calcium signals by binding

calcium ions and then modifying its interactions with various target proteins. 10 Arabidopsis WRKY proteins from group IId were recognized as CaM binding. The binding was verified by gel mobility shift assay, split-ubiquitin yeast two-hybrid assay and a competition assay with Ca^{2+}/CaM-dependant phosphodiesterase [133]. WRKY from IId group contains a short region called C-motif responsible for calmodulin binding. This domain has conserved amino acids sequence DxxVxKFKxVISLLxxxR. Functionally characterized WRKY from IId roup: AtWRKY7, AtWRKY11 and AtWRKY17 act as regulatory repressors of plant basal defense [93, 96]. Moreover, if the C-motif is localized close to the WRKY domain, then binding of CaM will prevent RKY-WRKY interactions. This might be a possible mechanism for regulation of WRKY-WRKY interaction by cellular Ca^{2+} levels.

Seven WRKYs from *A. thaliana* (AtWRKY6, AtWRKY16, AtWRKY18, AtWRKY19, AtWRKY27, AtWRKY32 and AtWRKY40) were identified as complexes with 14-3-3 proteins [19]. This interactions are promoted by phosphorylation by pathogen-responsive kinase cascades. 14-3-3 proteins dimerize and might bind two target proteins. Among the targets are also other than WRKYs phosphorylated proteins. 14-3-3 proteins often function as adaptor proteins that bind multitude of regulatory and signaling proteins, thus they could have important function in complex WRKY interaction network [153].

1.4. Structural studies of WRKY proteins

Structural studies of WRKY proteins are crutial to understand the mechanism of their interaction with both DNA and other potential binding partners. Each WRKY possess outside invariable DNA-binding WRKY domain other motifs responsible for interaction with different protein partners. Thus, the global structure determination is essential to help us understand the complex mechanisms of signalling and transcriptional reprogramming of cell functioning controlled by WRKY proteins. Unfortunately a solution structure is available only for highly conserved DNA-binding domain but not for full-length WRKY protein and there is no topological data regarding subgroup-specific motifs available. Structural data of full length WRKY protein will help us to understand how do they act as transcription regulators and locate potential DNA and interacting partners binding sites. So far there are only 3 structures of AtWRKY DNA-binding domains deposited in PDB (Fig. 3). Up to date one structure is solved using X-ray crystallography and the two others using NMR spectroscopy. Known

crystal structure represent C-terminal domain of AtWRKY1 (PDB code: 2ayd) [51] and the two NMR structures referred to the corresponding domain from AtWRKY4. One of them was solved in complex with DNA (PDB code: 1wj2) [190] and the other without ligand (PDB code: 2lex) [191]. Both, crystal and NMR structures posses very similar globular architecture composed of β-sheet. The crystal structure of the C-terminal part of AtWRKY1(PDB: 2ayd) [51] determined at 1.6 Å resolution revealed that this domain is composed of a globular structure with 5 β-strands, forming an antiparallel β-sheet. A zinc binding site was found at one end of the β-sheet, between strands β4 and β5. DNA-binding residues of WRKY1 are located at β2 and β3 strands [31]. 2-5 β-sheets correspond to 1-4 β-sheets from NMR structure of AtWRKY4 domain (PDB: 1wj2) [190].

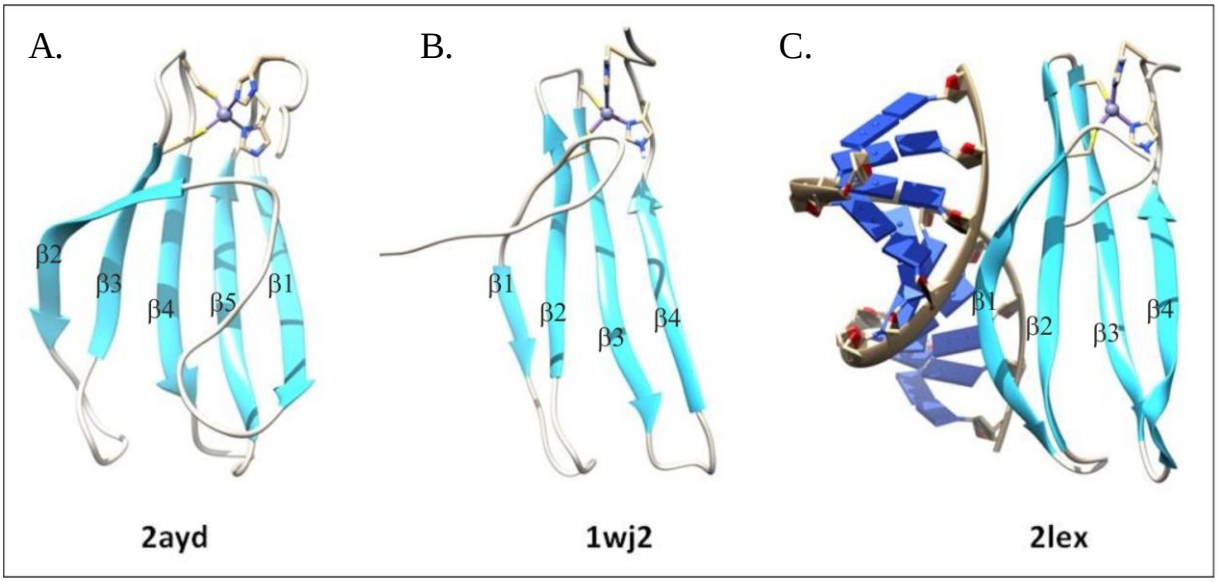


Fig.3. Known structures of *Arabidopsis thaliana* WRKY domains solved by biomolecular crystallography (2ayd) and NMR spectroscopy (1wj2 and 2lex). Detailed description in the text.

The major differences between the known structures were observed in the region considered as C- and N-termini of the domain. The secondary structure elements of above mentioned WRKY domains are β-strands forming an antiparallel β-sheet. Conserved Cys/His residues located at C-terminus of the β-sheet formed zinc binding pocket. WRKYGQK residues are present at the N-terminus of β-sheet. Structure of the C-terminal AtWRKY4 domain in complex with the DNA fragment (W-box) solved by NMR allowed to deduce the DNA-binding mechanism [191]. A four stranded β-sheet enters the major groove of DNA in an

atypical mode termed β-wedge, where the sheet is nearly perpendicular to the DNA helical axis. Residues in the conserved WRKYGQK motif (except tryptophane, W) contact DNA bases mainly through extensive apolar contacts and hydrogen-bonding interactions with thymine methyl groups [191]. The structure of the protein in complex with DNA consists of four-stranded β-sheet which is similar to that without DNA with a backbone root mean square deviation of 1.9 Å. The 16 bp DNA is in the B-form with slight bent toward the protein. The major molecular interface is created by the β1-strand that contains invariant WRKYGQK sequence (Fig. 3C). The formation of the complex significantly altered the position of this strand to the others. The kink at Gly enabled the close contact of β1-strand with DNA bases [191].

2. Goal of the thesis

WRKY proteins regulate many fundamental life processes of plants, allowing them to survive under various stress conditions such as invasion of pathogens, drought, cold, high salinity or wounding. Due to that fact, understanding the molecular mechanisms underlying their actions may be useful in the future in plant breeding or plant biotechnology for obtaining new varieties of plants resistant to biotic and abiotic stress.

The main goals of my dissertation were structural studies of plant WRKY transcription factors from *A. thaliana*. Despite extensive studies in last two decades, there are still limited informations about structure of WRKY proteins. So far the only structural studies of the DNA-binding domain were performed. It is due to the difficulties in preparation of soluble full-length recombinant WRKY proteins. Production of recombinant WRKY TF in prokaryotic system is challenging because of troubles appeared at different stages of applied procedures. The bottle neck is not only a toxicity for bacteria and inhibition of bacterial growth but also incorrect folding, inclusion bodies formation or vestigial expression. Apart from efficient expression, there are very often problems with solubility after cleavage of fusion tag such as protein aggregation and precipitation.

For my experimental work carried out in this dissertation, the coding DNA were provided by Dr. Imre Sommsich from the Max Plank Institute for Plant Breeding Research in Cologne, Germany. They were interested in determination of the three-dimensional structure because it would help to understand the mechanism of WRKY proteins action in regulation of gene transcription in plants. This was a very ambitious as well as difficult task, because structural studies required high amount of soluble protein.
The Laboratory of Protein Engineering at the Institute of Bioorganic Chemistry in Poznan, where my thesis was completed specializes in recombinant protein production. New technologies available in the laboratory: wide range of vectors, bacterial strains and growing conditions convinced me to accept this challenge.

3. Results

3.1. WRKY selection

20 target AtWRKY genes of 74 available were choosen for solubility screening. The selection was done according to the various functional properties and presence of characteristic sequence motives in individual coding sequences. Among them, 18 full length proteins and 2 DNA-binding domains were present (Table 2). The main criterion of selection was to choose representatives of all 3 main grups and 5 subgroups. Further selection was done based on molecular mass, isoelectric point and the information available in literature about the function of individual WRKY protein in plants. General features of WRKY proteins selected for experimental procedures are summarized in Table 2.

Table 2.
AtWRKY proteins selected for experimental procedures.

AtWRKY	GROUP	AA	MOLECULAR MASS (kDa)	pI
6	**IIb**	**553**	**60.6**	**6.2**
11	**IId**	**325**	**35.8**	**10.2**
17	**IId**	**321**	**35.0**	**10.4**
18	**IIa**	**310**	**34.8**	**7.4**
18^{DBD}	**IIa**	**76**	**8.8**	**9.3**
22	**IIe**	**298**	**32.3**	**7.6**
25	**I**	**395**	**44.1**	**6.4**
29	**IIe**	**304**	**33.8**	**7.0**
30	**III**	**303**	**33.9**	**6.6**
30^{DBD}	**III**	**76**	**9.2**	**9.5**
33	**I**	**519**	**57.2**	**7,9**
38	**III**	**289**	**33.3**	**5.4**
40	**IIa**	**302**	**33.7**	**8.5**
43	**IIc**	**109**	**12.9**	**9.9**
50	**IIc**	**173**	**19.3**	**6.4**
51	**IIc**	**194**	**22.0**	**5.3**
53	**III**	**324**	**36.3**	**6.8**
56	**IIc**	**195**	**21.8**	**8.2**
62	**III**	**263**	**30.4**	**6.4**
70	**III**	**294**	**32.9**	**6.2**

3.2. Screening for soluble recombinant AtWRKY proteins

3.2.1. Cloning of WRKY genes

The pDONR plasmids harbouring the coding sequences of the following WRKY proteins: AtWRKY6, AtWRKY11, AtWRKY17, AtWRKY18, AtWRKY22, AtWRKY29, AtWRKY30, AtWRKY33, AtWRKY40, AtWRKY50, AtWRKY51, AtWRKY56 and AtWRKY70. were kindly provided by Max Plank Institute for Plant Breeding in Cologne, Germany. All of them were then subcloned into suitable expression vectors.

cDNA of *Arabidopsis thaliana* Columbia-0 wildtype plants were used as templates to isolate and cloning other target WRKY genes of following WRKY proteins AtWRKY25, AtWRKY38, AtWRKY43, AtWRKY53 and AtWRKY62.

The coding sequences were obtained with the following procedure. First, RNA was isolated and cDNA with poliA primers were synthesized. cDNA was further used as template for PCR with primers specific to the coding sequence. Primers were designed according to the sequences available in the TAIR database:

(http://www.arabidopsis.org/browse/genefamily/WRKY.jsp).

For amplification of the AtWRKY25, AtWRKY38, AtWRKY53, AtWRKY62 coding DNA leaves were used as the best source of RNA but transcripts for AtWRKY43 were present only in flowers.

All target genes were successfully amplified using PCR method and the cDNA or pDONR plasmids as templates and afterwards it was verified by electrophoresis on agarose gel. Almost all genes were obtained as single product, but for AtWRKY43, two fragments were observed after amplification. One of those bands referred to exact size of the amplified *wrky*43 gene and the second one was nonspecific. A proper band was extracted from the agarose gel, purified and used PCR reamplification.

All amplified genes were cloned into expression vectors. The cloning of *At*WRKY sequences was carried out using three strategies: 1) TOPO cloning (into pET151/D-TOPO vector), 2) LIC (into pMCSG7, pMCSG9 and pMCSG48 vectors) and 3) restriction enzyme cloning (into pET32a vector) as described in Materials and Methods, 5.2.1.4 section. 49 constructs of several AtWRKY proteins with different tags and fusion proteins were obtained (see table 3).

TOP10 cells were transformed with the constructs of expression vectors for selection. The cells were streaked on selective solid growth media containing selective antibiotics. After overnight growth at 37°C, few colonies were picked for plasmid isolation. Isolated plasmids were then analyzed *via* PCR. To select correctly oriented insertion of the target gene fragment. Selected plasmids were used for sequencing analysis. Chosen constructs with correctly oriented protein coding DNA fragments were used for protein expression.

3.2.2. Expression and purification

Each of obtained 49 constructs was used for overexpression of individual AtWRKY proteins in *E. coli* cells. For optimization trials only constructs with significant protein overexpression level were chosen. During this procedure following difficulties were encountered: lack of expression, weak expression, insolubility or protein aggregation. Some of proteins obtained as fusion proteins were insoluble, others precipitated after cleavage of fusion tags. If even the protein was soluble after cleavage of fusion tag, usually it formed aggregates with other bacterial proteins. It was impossible to overcome this problem and obtain pure protein. In Table 3 all expression and purification results are summarized.

Table 3. List of AtWRKY proteins constructs including solubility of the recombinant proteins.

AtWRKY	pET151/D-TOPO	pMCSG7	pMCSG9	pMCSG48	pET-32a
6	IN	X	X	INAC	X
11	IN	X	X	IN	X
17	IN	X	X	IN	X
18	IN	IN	IN	SA	IN
18^{DBD*}	X	S	X	S	X
22	IN	X	X	SA	X
25	X	X	X	IN	X
29	IN	X	X	SA	X
30	IN	IN	IN	IN	X
30^{DBD*}	X	S	X	S	X
33	X	IN	X	INAC	X
38	X	X	X	IN	X
40	IN	X	INAC	INAC	IN
43	X	SA	X	SA	X
50	IN	IN	SA	S	X
51	IN	IN	IN	INAC	X
53	X	X	X	SA	X
56	IN	X	IN	SA	IN
62	X	X	X	IN	X
70	IN	X	X	IN	X

Legend:
IN-insoluble fusion protein
INAC-insoluble after cleavage of fusion tag
SA-soluble aggregates
S-soluble
X-not tested
*DBD-DNA binding domain

3.2.2.1. TOPO-cloning

Competent *E.coli* BL21(DE3)STAR cells were transformed with the plasmid constructs each containing one of the 12 WRKY genes. In order to test overexpression of the target genes, the transformants were grown in LB or TB medium supplemented with appropriate antibiotic at 37°C. Expression of the WRKY proteins was induced by isopropyl-β-D-thiogalactopyranoside (IPTG) when the growth of bacteria was between the mid-log phase and stationary phase. After five hours of IPTG induction, overexpression of all WRKY proteins indicated as AtWRKY6, AtWRKY11, AtWRKY17, AtWRKY18, AtWRKY22, AtWRKY29, AtWRKY30, AtWRKY33, AtWRKY40, AtWRKY50, AtWRKY51, AtWRKY56 was observed with diverse efficiency.

Overexpressing bacteria were grown at different conditions in small scale cultures for testing protein solubility. These were five hours growth in LB media at 37 °C induced with 0.3-1 mM IPTG or overnight growth in LB or TB media at 18°C induced with 0.5 mM IPTG. After overexpression, pelletted cells were lysed in a buffer containing 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% Glycerol, 0.5% Triton X-100, 10 mM Imidazole, 100 µM Lysozyme. Soluble and insoluble fractions were analyzed by SDS-PAGE.

All 12 proteins expressed in bacterial system were insoluble (Fig. 4). Attempts were made to purify the protein from insoluble fraction (inclusion bodies). This method required use of urea as a denaturation agent. In this case the pelletted cells were lysed in a buffer containing 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% Glycerol, 0.5% Triton X-100, 10 mM Imidazole, 100 µM Lysozyme. After sonication and centrifugation, insoluble inclusion bodies were resolubilized with a denaturing agent containing 50 mM Tris-HCl, 5% Glycerol, 7.2 M urea, 5 mM DTT. pH of the buffers was adjusted according to pI values of the proteins.

Directly after resolubilization, the denatured proteins were subjected to renaturation procedure. Refolding trials were performed by different methods: quick dilution, on column refolding or overnight dialysis against 50 mM Tris-HCl, 500 mM NaCl, 5% Glycerol, 5 mM DTT at 4 °C. Renaturation trials for all tested WRKY resulted in precipitation and aggregation. The highest rate of refolding was observed for AtWRKY 18 and AtWRKY30 if using overnight dialysis as renaturation method. However, obtained renatured proteins were very unstable with tendency to precipitation. AtWRKY 18 and AtWRKY30 were purified on HisTrap™ column and HiLoad Superdex 200 16/60. Unfortunately, they precipitated during

purification procedure. Also concentration of those proteins using Amicon Ultra 10 filters (Millipore) or ultrafiltration membranes under nitrogen pressure triggered precipitation. Any applied refolding method did not enable to obtain suitable amount of purified AtWRKY proteins for structural studies.

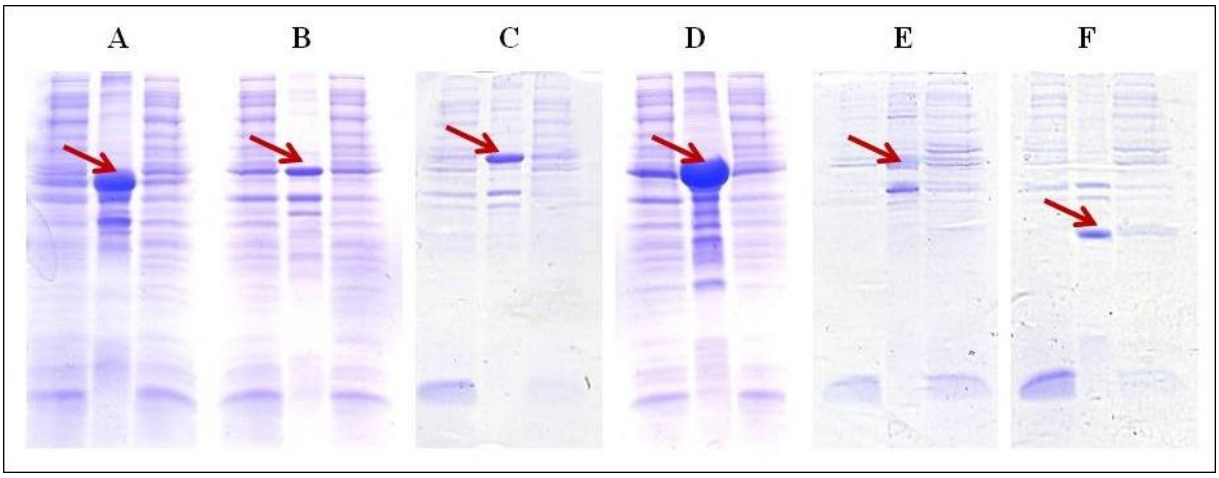


Fig. 4. Few examples of recombinant WRKY proteins expression in *E. coli* BL21STAR cells. Lane 1: before induction Lane 2: pellet after induction with IPTG. Lane 3: supernatant after induction with IPTG. Red arrow indicate position of individual AtWRKY protein.
A) AtWRKY18, B) AtWRKY22, C) AtWRKY29 D) AtWRKY30, E) AtWRKY40, F) AtWRKY56.

3.2.2.2. Ligase Independent Cloning - LIC

Competent *E.coli* BL21Magic cells were transformed with the 34 sequenced plasmid constructs of the 20 WRKY genes. In order to test overexpression of the target genes, the transformants were grown in LB medium containing appropriate antibiotic at 37°C and the temperature was lowered to 18°C when the expression of the WRKY protein was induced by 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The bacteria were cultivated overnight and the overexpression of recombinant WRKY proteins as fusion with NusA, MBP or HisTag was observed with diverse efficiency. Overexpressed proteins were analyzed for their solubility from small scale cultures (Fig. 5). After overexpression, pelletted cells were lysed in a buffer containing 50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 1 mM TCEP, 20 mM imidazole. Soluble and insoluble fractions were analyzed by SDS-PAGE. All soluble fusion proteins were overexpressed in large scale (1l culture) and purified as described in Materials and methods section 5.2.1.4.

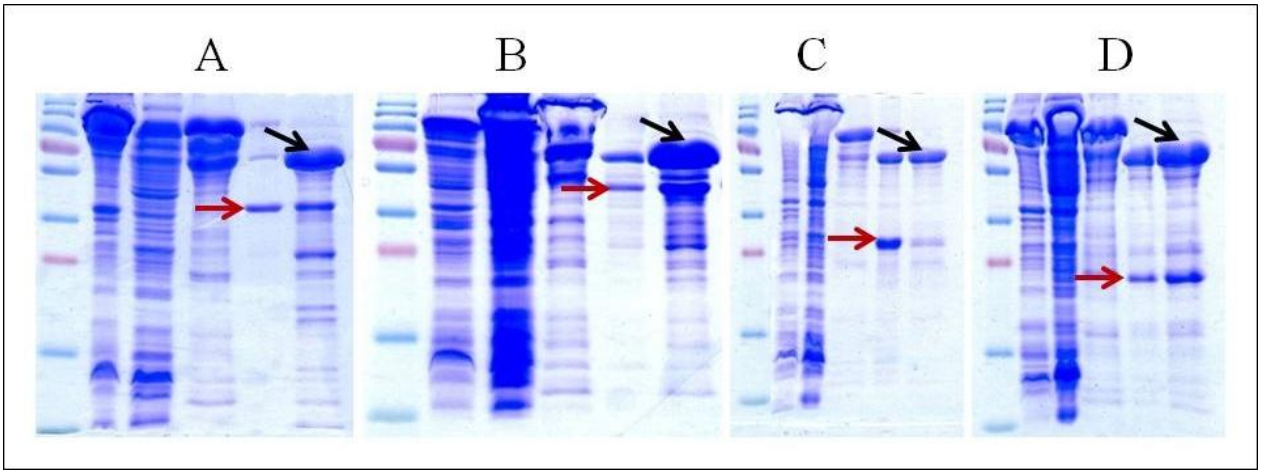


Fig. 5. Examples of solubility screening within constructs in pMCSG48 vector. (A) AtWRKY18, (B) AtWRKY29, (C) AtWRKY51, (D) AtWRKY56. Red arrows correspond to AtWRKY proteins, black arrows indicate NusA fusion protein.
Lane 1- protein ladder (250, 130, 100, 70, 55, 35, 27, 15, 10 kDa), Lane 2-insoluble fraction after lysis, Lane 3- supernatant after lysis, Lane 4- fraction after Ni-Sepharose purification, Lane 5- pellet after TEV cleavage, Lane 6- supernatant after TEV cleavage.

From 34 plasmid constructs of different AtWRKY proteins with His-Tag, MBP or NusA fusion only one full-length and DNA binding domains from two AtWRKY proteins were soluble after cleavage of fusion protein. They did not form aggregates. Majority of overexpressed protein fusions were insoluble even those produced with MBP or NusA proteins known as solubility promoting. The most effective for WRKY's expression was pMCSG48 plasmid with NusA fusion. Among constructs with NusA, 7 proteins were totally insoluble as fusion, 13 gave soluble proteins, 4 precipitated after cleavage of NusA and 8 were soluble after cleavage but formed aggregates with other proteins including NusA (Fig. 6 and Fig. 7). Those aggregates were very difficult to purify even in presence of detergent (0.1-0.3% DDM) (Fig. 6) or other additives like 50-200 mM arginine and 50-200 mM glycine (Fig. 7). Finally only one full length protein - AtWRKY50 (Fig. 8) and two domains AtWRKY18^{DBD} (Fig. 17) and AtWRKY30^{DBD} appeared soluble, stable and did not form aggregates. For crystallization and all other experiments described in next sections, both AtWRKY50 and AtWRKY18^{DBD} were used.

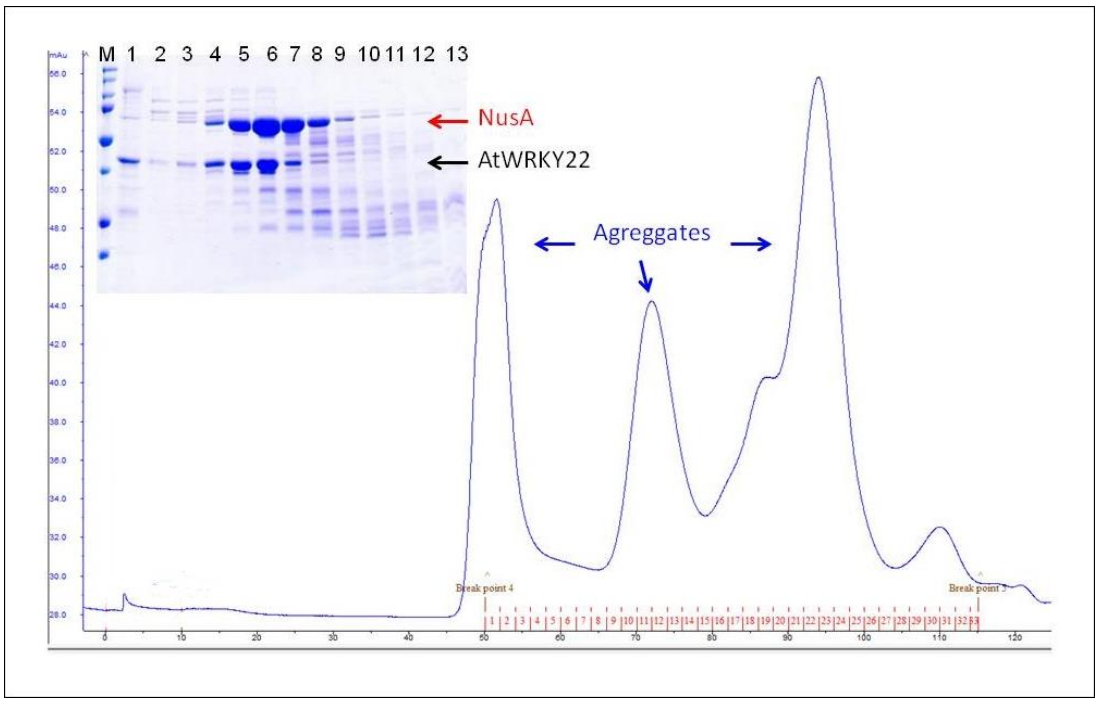


Fig. 6. Size exclusion chromatography on Superdex 200HL column (FPLC) of AtWRKY22 in buffer containing 50 mMTris pH 8.0, 500 mM NaCl, o.o1% DDM, 1 mM TCEP. (B) protein SDS-PAGE electrophoresis of AtWRKY22 purification steps. Lane M- protein ladder (250, 130, 100, 70, 55, 35, 27, 15, 10 kDa), 1- protein eluate after second Ni column, 2-13 - peak fractions after gel filtration.

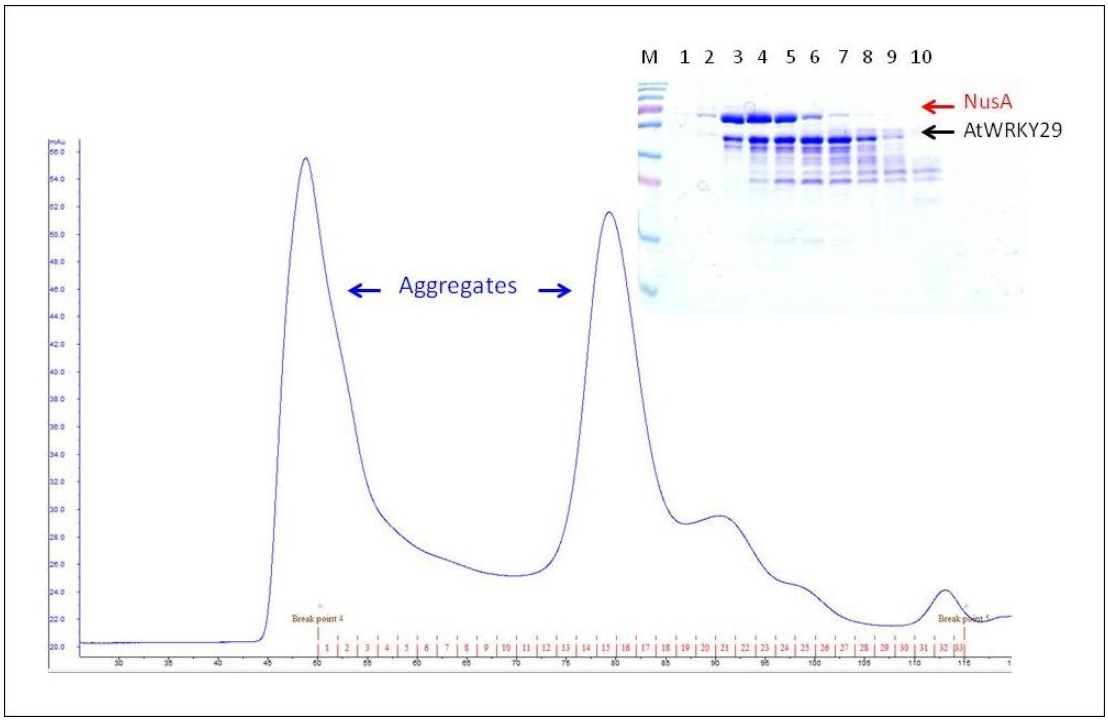


Fig. 7. Size exclusion chromatography on Superdex 200HL column (FPLC) of AtWRKY29 in buffer containing 50 mM Tris pH 8.0, 500 mM NaCl, 100 mM Gly, 100 mM Arg, 1 mM TCEP. (B) SDS-PAGE electrophoresis of AtWRKY29 protein from different steps of purification. Lane M- protein ladder (250, 130, 100, 70, 55, 35, 27, 15, 10 kDa), 1-10 - peak fractions after gel filtration.

3.2.2.3. Cloning into pET-32a vector

E.coli BL21Magic competent cells were transformed with the plasmid constructs harbouring AtWRKY18, AtWRKY40 and AtWRKY56 genes. In order to overexpress target genes, the transformants were grown in LB medium containing appropriate antibiotic at 37°C and the temperature was lowered to 18°C when expression of the WRKY protein was induced by 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The bacteria were cultivated overnight and overexpression of recombinant WRKY proteins as fusion with Trx (thioredoxin) was tested. In this conditions, AtWRKY18 and AtWRKY56 recombinant proteins were found in insoluble fraction. In case of AtWRKY40 only thioredoxin was expressed at high level despite the insert was found in proper orientation and in expected position of the construct as verified by sequencing.

3.3. *Arabidopsis thaliana* WRKY50 protein

3.3.1. Cloning and overexpression

AtWRKY50 was cloned into pMCSG48 vector using LIC method and the protein was produced as fusion with His-tag followed by NusA and TEV protease cleavage site between tags and protein sequence. Bacteria were grown at 37°C prior induction, then the temperature was decreased and the protein expression was induced with 0.5 mM IPTG. Soluble recombinant AtWRKY50 was overexpressed during overnight cultivation of *E. coli* cells at 18°C.

3.3.2. Purification

The protein was purified as described in Materials and methods (section 5.2.1.7). The supernatant separated from cell debris was purified using 2-step nickel column followed by gel filtration (FPLC). The supernatant was applied on a column packed with Ni-Sepharose HP resin (GE Healthcare). The eluted protein was cleaved with TEV protease to get rid of the His-NusA-tag and the excess of imidazole was removed simultaneously by dialysis (overnight, 4°C). The solution was mixed with Ni-Sepharose HP resin to get rid of the His-NusA-tag and the His-tagged TEV protease. The flow-through was collected, concentrated to 5 ml and applied on a HiLoad Superdex 200 16/60HL column (GE Healthcare). The final purification step (e.g. size exclusion chromatography) yielded a homogenous protein fraction of

monomeric AtWRKY50 (10 mg of pure protein from 1 l culture). After all chromatographic steps pure protein visible as one band on SDS-PAGE was obtained. The purification steps are shown on SDS-PAGE gel (Fig. 8.). The pure protein was concentrated to 10 mg/ml and used immediately or was flash frozen in liquid nitrogen.

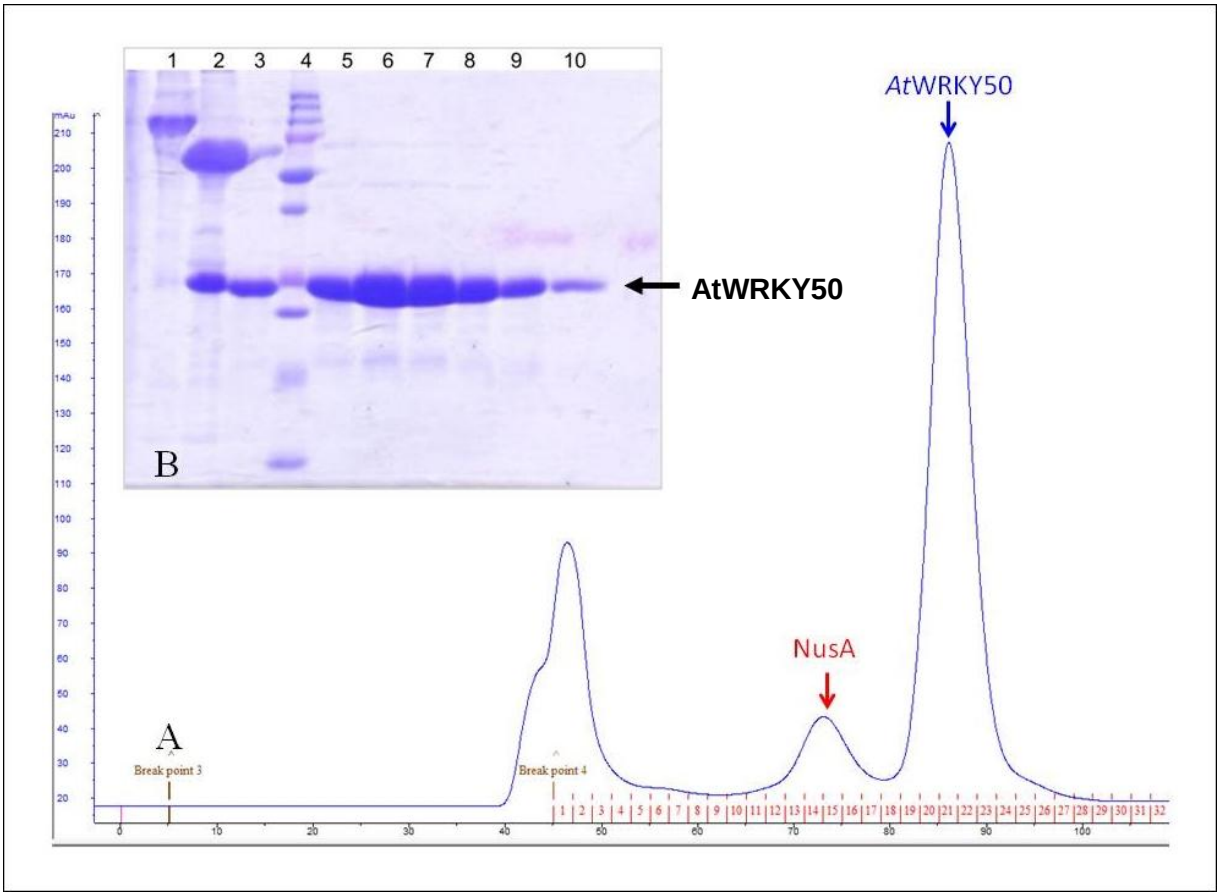


Fig. 8. (A) Size exclusion chromatography on a Superdex 200 FPLC column (GE Healthcare) of the AtWRK50 protein (B) SDS-PAGE electrophoresis of AtWRKY50 protein from different purification steps. Lane 1- protein after elution from first Ni-column, 2 - protein after TEV cleavage, 3 - protein eluate after second Ni-column, 4- protein ladder (250, 130, 100, 70, 55, 35, 27, 15, 10 kDa), 5-10 - peak fractions after gel filtration. The black arrow indicates the electrophoretic migration of the AtWRK50 protein.

3.3.3. Crystallization

Crystallization of AtWRKY50

Homogenous AtWRKY50 protein preparation at concentration about 10 mg/ml was used for high-throughput crystallization screening. Six screens from Molecular Dimensions were used for the initial experiments. Unfortunately none protein crystals were obtained. Precipitate was observed in a vast majority of crystallization drops. Crystallization was repeated using lower

protein concentration but still without success. Other crystallization trials were performed manually using hanging drop method. Initial hit for ligand-free AtWRKY50 crystallization was visible in Structure Screen (Molecular Dimensions) set manually (Fig. 9A). Microcrystals were observed within 4 hours after setting the drops at presence of 800 mM sodium potassium tartrate and 100 mM HEPES pH 7.5 in crystallization wells. Crystals were soaked with $(Ta_6Br_{12})^{2+}$ (Fig. 9A). These conditions were optimized to obtain large crystals. During optimization, different concentrations of protein, precipitant and salt, including pH and temperature were screened. The main problem was a reproducibility of crystals growing and their very small size. Unfortunately multiple trials to grow bigger crystals were unsuccessful. Despite testing various protein concentration (from 5-20 mg/ml), using additives (Additive Screen of Hampton Research, Lysine, Arginine and Serine) and optimizing concentration of buffer and other crystallization components, the obtained crystals were always very small. After optimization trials, the best crystals were obtained at 19°C in 750 mM sodium potassium tartrate, 100 mM HEPES, pH 7.5 (Fig. 9 B). The crystals appeared very quickly and after few days achieved the final dimensions. Crystals from initial screen soaked with $(Ta_6Br_{12})^{2+}$ and also crystals after optimization were frozen in liquid nitrogen with 20% glycerol as cryoprotectant and diffraction data were collected at synchrotron. Few frames of X-ray diffraction were collected at the EMBL beamline Petra of the DESY synchrotron in Hamburg. The crystals diffracted to very low quality when exposed to X-rays (Fig. 10). Annealing of crystals did not improve diffraction limit. It was not possible to index the images and to determine the space group and cell dimensions. Diffraction limit of about 8 Å was achieved. Nevertheless, the diffraction pattern indicated that the crystallized material was a protein. Cocrystallization trials of AtWRKY50 with DNA resulted precipitation in most droplets. Several crystallization screens were tested but without any success.

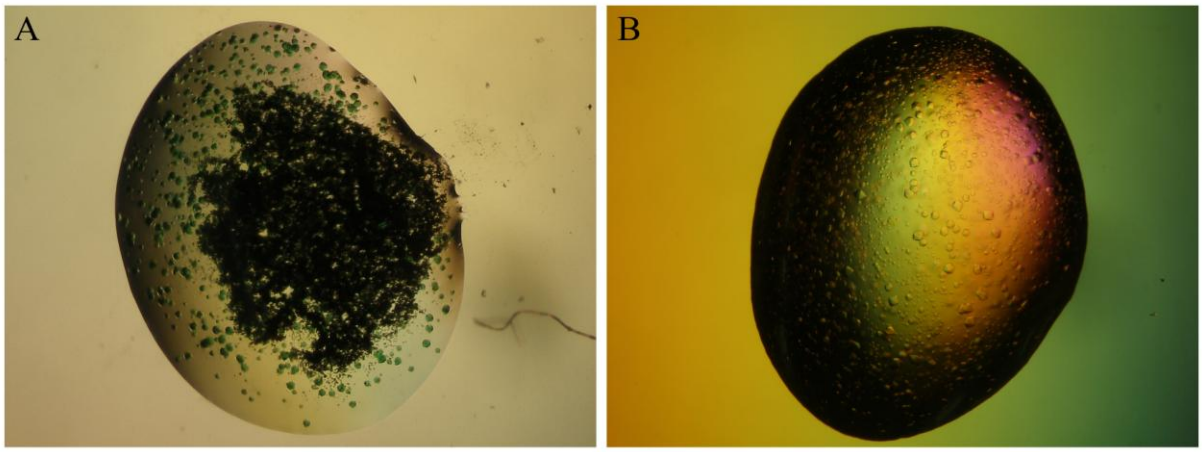


Fig. 9. Crystallization of ligand free AtWRKY50.
(A) Initial hit after screening, soaked with $(Ta_6Br_{12})^{2+}$ (B) Crystals after optimization.

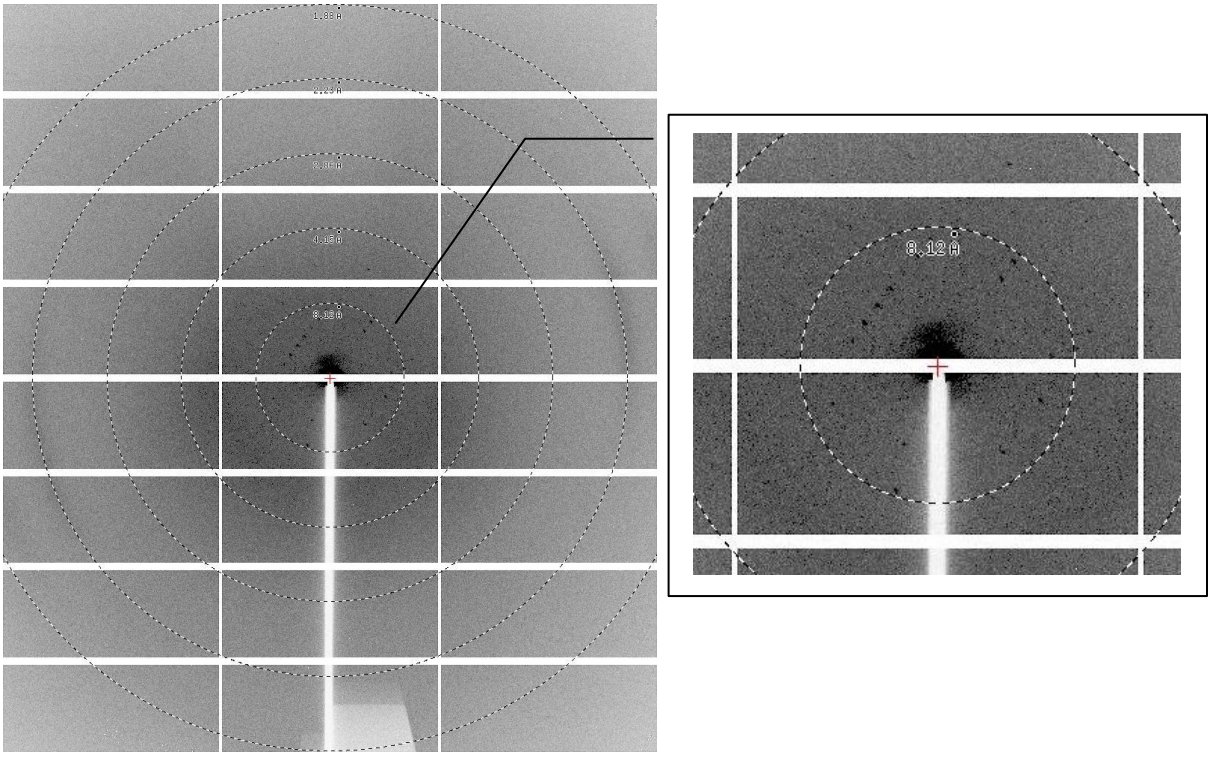


Fig. 10. X-ray diffraction pattern recorded for a single crystal of ligand free AtWRKY50 using synchrotron radiation. The edge of the detector corresponds to the last visible spot of 8 Å resolution.

Crystallization of modified AtWRKY50

AtWRKY50 was modified by reductive lysine methylation and treated with proteases as described in Materials and methods. Entire reductive lysine methylation of AtWRKY50 was carried out at 4°C. During methylation reaction a small amount of precipitated protein

occurred and it was removed by centrifugation. Soluble fraction of the methylated AtWRKY50 was concentrated up to 5 ml and subjected to SEC with HiLoad Superdex 200 16/60 column (GE Healthcare) equilibrated with a buffer composed of 50 mM Tris-HCl, pH 8.0, 200 mM NaCl and 1 mM TCEP. The fractions were analyzed using SDS-PAGE and those of pure protein were concentrated up to 8 mg/ml and subjected to crystallization.

The initial digestion of AtWRKY50 with proteases was prepared with all 12 enzymes from Proti-Ace and Proti-Ace 2 screens (Hampton Research) in small scale. Based on the proteolyses pattern of each enzyme visualised by SDS-PAGE (Fig. 11), two proteases: elastase and pepsin were used as additives for crystallisation trials in large scale.

Multiple crystallization screens (Morpheus, JCSG plus, PACT Premier, PGA from Molecular Dimensions) were used to crystallize modified protein. AtWRKY50 sample with methylated lysine residues in nearly all crystallization conditions formed heavy, brown precipitate without protein crystals. Crystallization trials of protein treated with proteases did not bring any hit. Drops with precipitation as well as clear drops were observed but even after few months, none of crystals occurred.

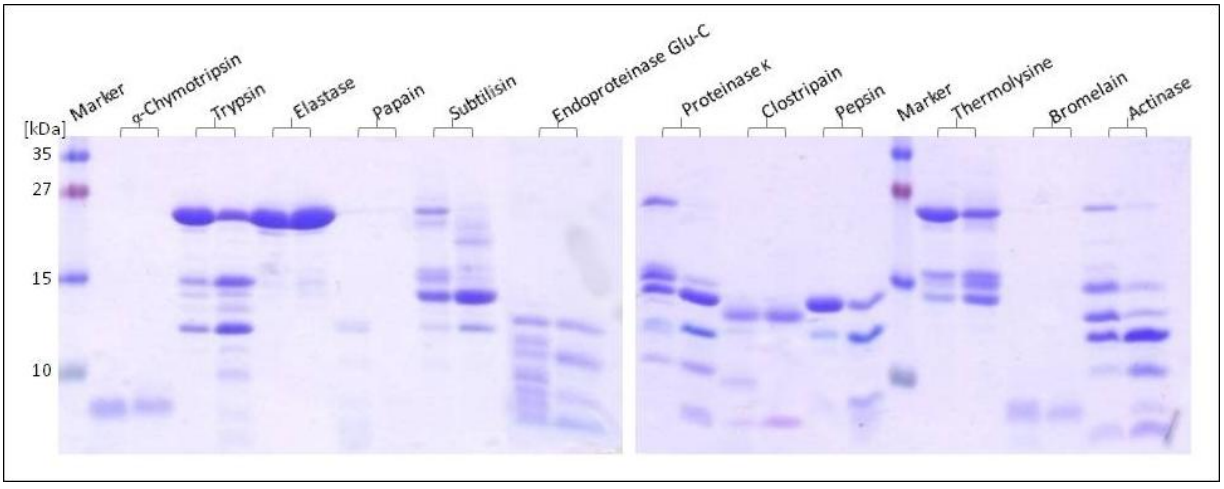


Fig. 11. SDS-PAGE of limited proteolysis of AtWRKY50. A and B correspond to Proti-Ace 1 and 2, respectively. Two measuring points were examined for each protease. Left lane represents the sample after 2 h and right lane after overnight incubation.

3.3.4 Functional and structural studies of *At*WRKY50

Structural studies and physicochemical characteristics were performed on recombinant AtWRKY50 protein using the procedures described in the Material end methods (section 5.2.3). From all tested recombinant WRKY proteins only full-length *At*WRKY50 was soluble, did not form aggregates, it was stable in solution and the expression was sufficient for setting crystallization. Unfortunately, despite the crystals appear, all trials of collecting diffraction data failed. To characterize the secondary structure of full-length AtWRKY50, the CD spectra and bioinformatics tools were employed. EMSA and ITC methods were applied to characterize DNA association behavior. These techniques allowed to estimate DNA-binding parameters.

3.3.4.1. DNA-binding analyses (EMSA, ITC)

For protein-DNA binding experiments, a pair of synthetic oligonucleotides (15 and 30 bp) containing the optimal binding site identical to sequence in the region of the parsley PR1-1 promoter containing one W-box (W2) [149] was used. The W2 DNA element was previously shown to be specifically bound by WRKY transcription factors [149]. AtWRKY50 belongs to a small IIc subgroup of WRKY proteins in which the DNA binding domain contains the WRKYGKK sequence opposed to WRKYGQK present in majority of other WRKY proteins.

Extensive attempts were made with different protein concentrations and experimental setups to assess AtWRKY50-DNA binding parameters. The DNA binding activity of AtWRKY50 was studied using isothermal titration calorimetry and electromobility shift assay. Both techniques confirmed binding ability of recombinant AtWRKY50 to tested DNA. In electrophoretic mobility shift assay, DNA duplex was incubated with the AtWRKY50 protein (0.27-10 μg), and specific DNA was detected by non-standard procedure as described in Methods (section 5.2.3.3). As shown in Fig. 12, recombinant AtWRKY50 binding activity was clearly detected by electromobility shift assay as visible shift of DNA-protein complexes. This method is not very precise and the quantitive analysis of the protein-DNA complex was not possible. This test was used only to confirm binding of studied molecules, but ITC (Isothermal Titration Calorimetry) was used for dissociation constant determination.

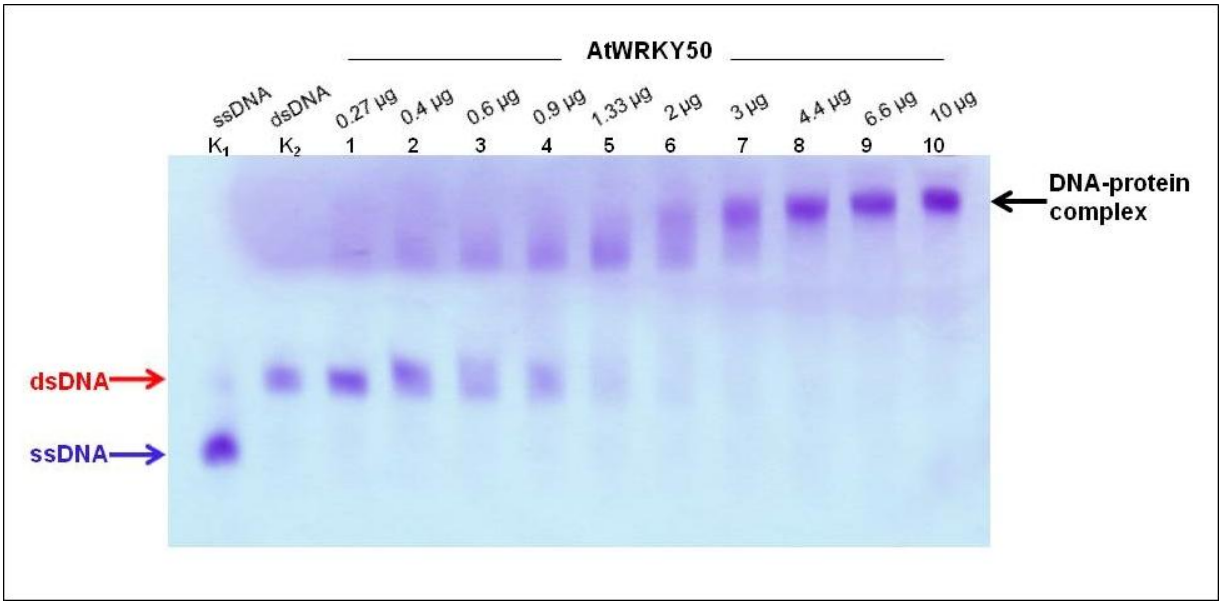


Fig. 12. Binding of AtWRKY50 transcription factor to W-box containing sequence. The specific retarded DNA-protein complexes are marked by black arrow, whereas red arrow designate positions of the free-running probe. The blue arrow indicate the single stranded DNA position.
Lane K_1- ssDNA, K_2-dsDNA, 1-10-dsDNA incubated with different protein amount 0.27-10 µg.

The titration revealed exothermic character of interaction between AtWRKY50 protein and DNA duplex. From the fitting sigmoidal titration curve stoichiometry N = 1.15 ± 0.01 and K_d = 237 ± 27 nM were obtained (Fig. 13A.). Titration of DNA duplex into the buffer alone revealed hyperbolically decreasing heat effect, which could be related to DNA dilution (Fig. 13B). This reference data were not subtracted from experimental points to avoid multiplification of the data errors of already relatively weak signal. Instead of that, the data were analyzed without first seven points, after which the plateau of above-mentioned effect could be observed.

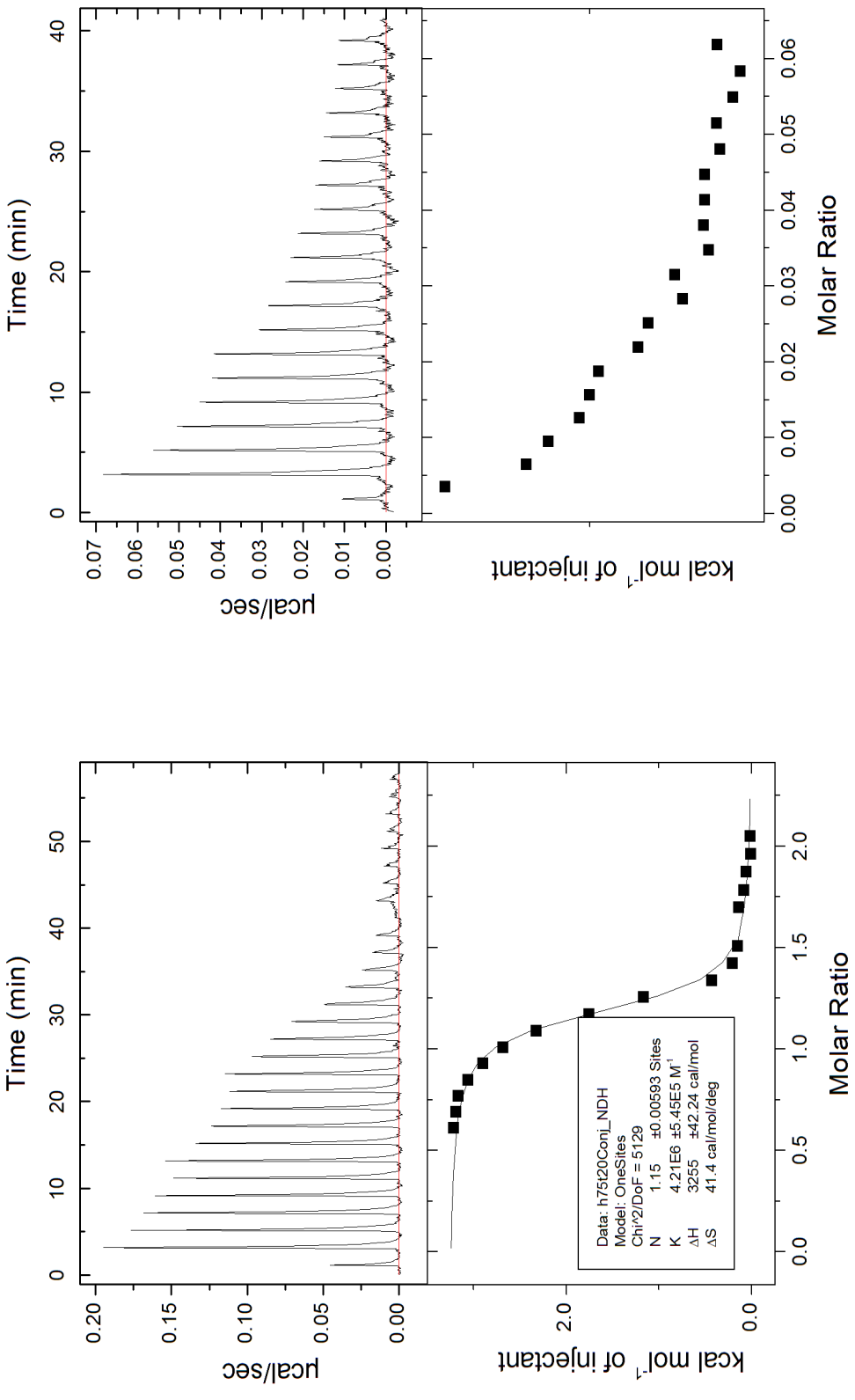


Fig. 13. A) calorimetric titration of AtWRKY50 protein with DNA duplex. The top panels show raw heat data obtained from 29 consecutive injections of 297 µM duplex into the sample cell (200 µl) containing 41 µM protein (at 290.15 K). The bottom panels show the binding isotherm created by plotting the heat peak areas against the molar ratio of DNA added to protein present in the cell. The line represents the best fit to the model of one independent site, B) titration data of dialysis buffer with DNA duplex

3.3.4.2. Secondary structure prediction

3.3.4.2.1. Circular dichroism

The content of secondary structure in AtWRKY protein was determined using circular dichroism spectroscopy. CD experiments were carried out at far-UV region (185-350 nm). The experimental CD spectrum obtained at far-UV in the range of 185-240 nm, collected for AtWRKY50 protein is presented in Fig. 14.

To estimate the secondary structure content, the CD spectrum was analyzed using DICHROWEB online server [182]. The results indicate that 40% of AtWRKY50 structure is disordered, while the rest retains a certain degree of organization and is composed of approximately 26% turns, 5% helix and 26% strands .

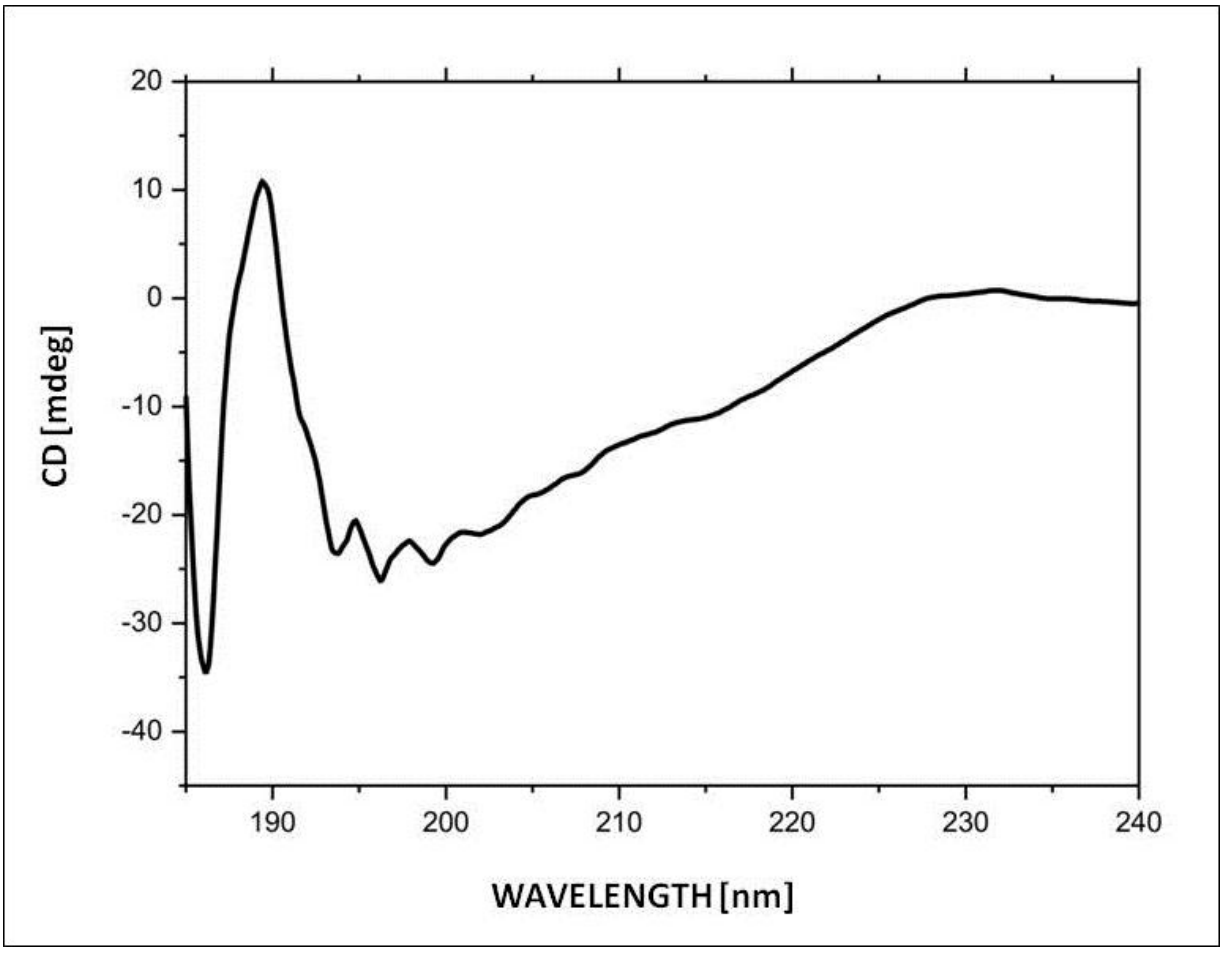


Fig. 14. CD spectra of AtWRKY50

3.3.4.2.2. Bioinformatics analyses

AtWRKY50 consists of 173 aminoacids. This protein is characterized by high content of charged amino acid residues (49), including 16 lysines and 17 aspartic acids. It is worth to notice that AtWRKY50 protein have also a high content of serines (24 of 173 residues).

Secondary structure content elucidated from CD spectrum indicate that AtWRKY50 posses 40% of disordered regions. It was the tip for prediction of AtWRKY50 disordered regions using an online Metadisorder server [102]. This tool allowed to calculate "consensus" from results returned by other primary disorder prediction methods. Metadisorder web service consists of four parts: MetaDisorder, MetaDisorder3D, MetaDisorderMD and MetaDisorderMD2. The detailed description of calculation method is given in Materials and Methods, section 5.2.3.5.2.

Analysis of AtWRKY 50 secondary structure using disorder predictors is shown in Fig. 15. This bioinformatics analysis showed disordered N-terminal region of about 100 residues and well ordered C-terminal 70-aminoacid region corresponding to the WRKY domain. Predicted structure is in a good agreement with Circular Dichroism experimental data.

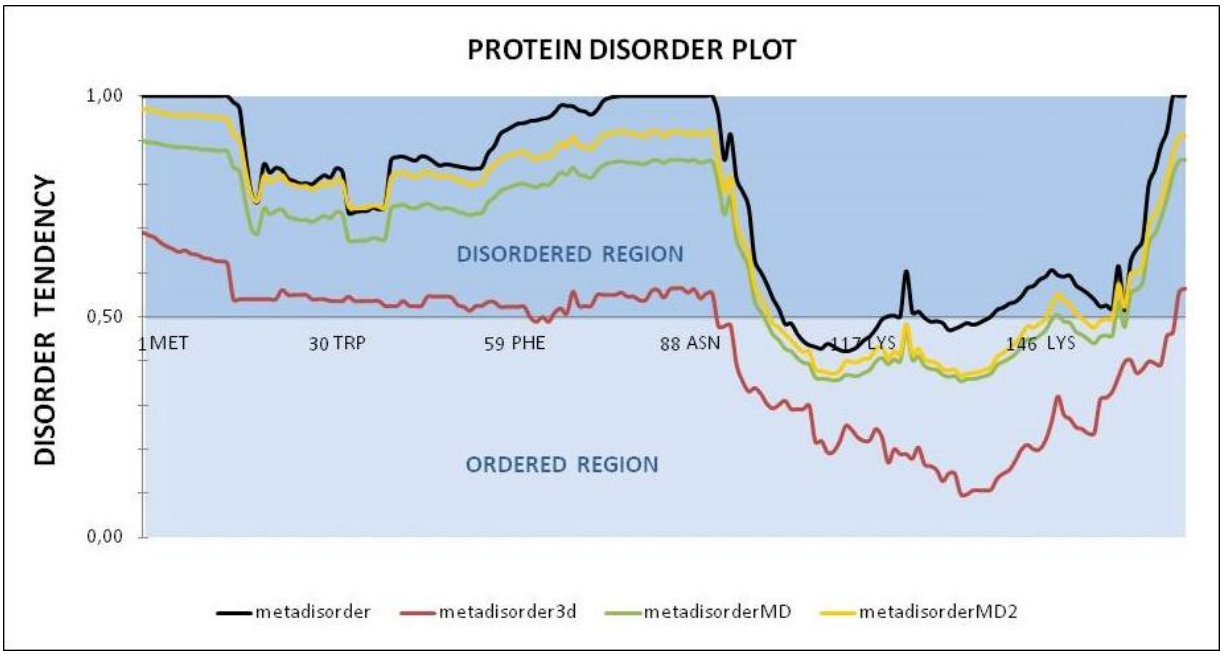


Fig. 15. Analysis of AtWRKY 50 secondary structure using disorder predictors. The local disorder is shown as a function of residue number. The residues with disorder tendency over 0,5 are considered to be disordered.

Additional sequence analysis was performed using MoRFpred online tool [46] which found one MoRF (Molecular Recognition Feature) region consist of 4 aminoacids (102Ala, 103Phe, 105Thr, 106Arg) as indicate in red in Fig. 16. This region is located exactly between disordered and ordered regions predicted using Metadisorder server.

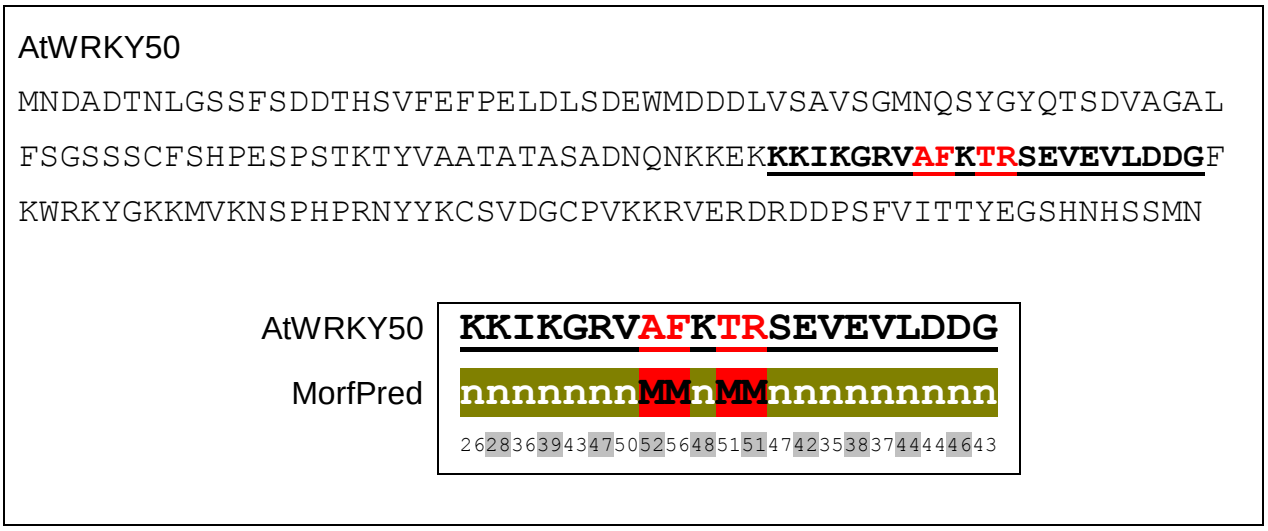


Fig. 16. The first line displays the query sequence followed by predictions which are shown in two rows: the first row annotates Molecular Recognition Feature (MoRF) (M) and non-MoRF (n) residues; the second row gives prediction scores (the higher the score the more likely it is that a given residue is MoRF).

AtWRKY50 sequence contain few characteristic regions: (1) WRKY domain- responsible for interaction with DNA, (2) zinc finger, a small structural motif within WRKY domain responsible for coordination zinc ion, which is necessary for DNA binding, and (3) MoRF – 5-amino acid sequence responsible for stabilization disordered domain. MoRF sequence is localized between disordered and ordered region of protein.

3.4. *Arabidopsis thaliana* WRKY18-DNA binding domain

3.4.1. Cloning and overexpression

AtWRKY18^{DBD} was cloned into pMCSG7 and pMCSG48 vectors using LIC method. The protein was obtained as His-tagged protein or as fusion with His-tag followed by NusA, both possesed TEV protease cleavage site between tags and protein sequence. Bacteria were grown at 37°C prior induction, then the temperature was decreased and the protein expression was induced with 0.5 mM IPTG. Soluble recombinant AtWRKY18^{DBD} was obtained by overnight cultivation of transformed *E. coli* cells at 18°C. Due to the large molecular mass of the NusA fusion (ca. 70 kDa) the efficiency of expression was lowered, therefore for further experiments the protein that possesed only His-tag without any fusion was choosen.

3.4.2. Recombinant protein purification

Recombinant AtWRKY18^{DBD} was purified as described in Materials and methods (section 5.2.1.7). The supernatant separated from cell debris was purified using 2-step Ni-Sepharose column followed by gel filtration. The supernatant was applied directly on a column packed with Ni-Sepharose HP resin (GE Healthcare). The eluted protein was cleaved with TEV protease to get rid of the His-tag and the excess of imidazole was removed by dialysis (overnight at 4°C) simultaneously. The solution was mixed with Ni Sepharose HP resin to get rid of the His- tag and the His-tagged TEV protease. The flow-through was collected, concentrated to 5 ml and applied on a Superdex 200 HL 16/60 column (GE Healthcare). The final purification step, size exclusion chromatography, yielded a homogenous fraction of monomeric AtWRKY18^{DBD} (6 mg of pure protein from 1 l culture). After all chromatographic steps pure protein visible as single band on SDS-PAGE was obtained. The purification steps are presented on an SDS-PAGE gel in Fig. 17. The pure protein was concentrated to 10 mg/ml and used immediately or flash frozen in liquid nitrogen.

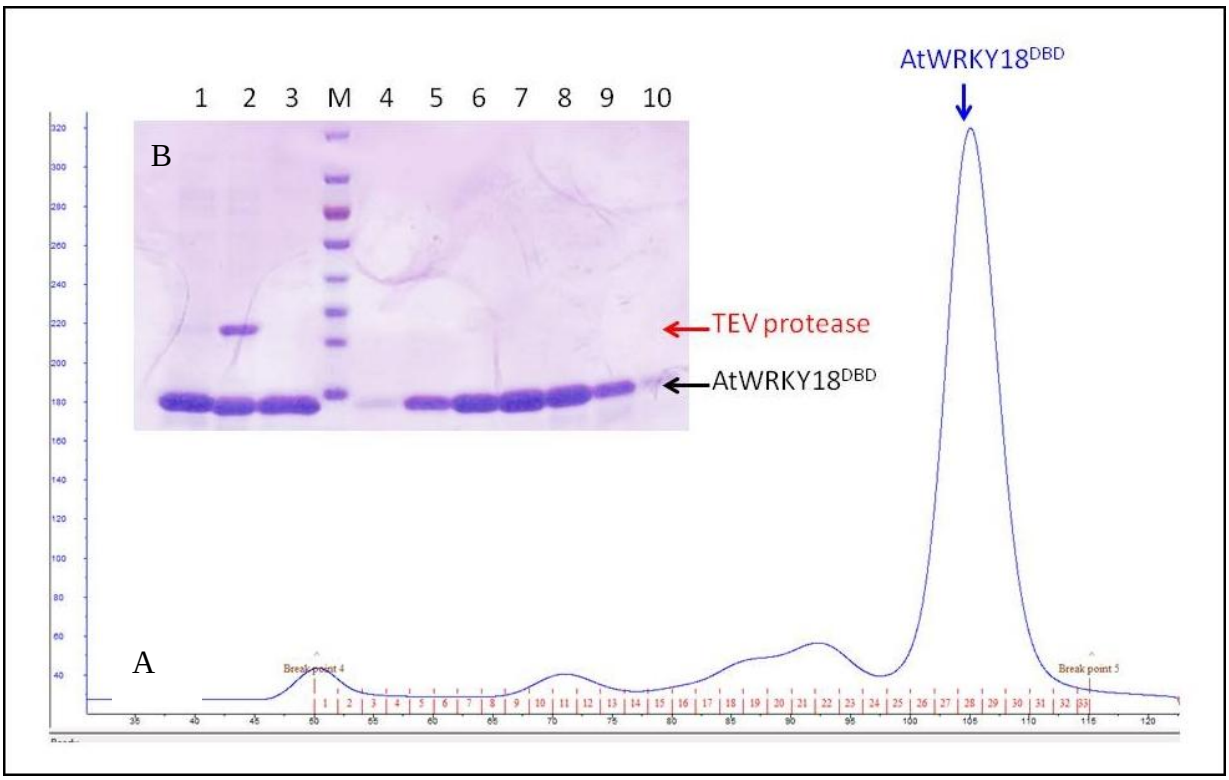


Fig. 17. A) Size exclusion chromatography on a Superdex 200 FPLC column (GE Healthcare) of the AtWRKY18^{DBD} B) SDS-PAGE electrophoresis of AtWRKY18^{DBD} purification steps. Lane 1-protein after elution from first Ni column, 2- protein after TEV cleavage, 3- protein eluate after second Ni-column, M-protein ladder (130, 100, 70, 55, 35, 27, 15, 10 kDa), 4-10 - peak fractions after gel filtration (FPLC).

3.4.2. Crystallization of AtWRKY18^{DBD}

Prior to crystallization trials, a homogenous solution of AtWRKY18^{DBD} protein was concentrated to approximately 10 mg/ml and kept in 50 mM Tris buffer, pH 7.5, 200 mM NaCl, 2mM TCEP at 4°C. The crystal structure of similar AtWRKY1 C-terminal domain (AtWRKY1^{DBD}) showing 58% amino acid sequence identity is known [51]. The initial crystallization condition were based on variants of the known condition described in literature for AtWRKY1^{DBD}. Despiete similar crystallization condition for AtWRKY18^{DBD} were checked: 1.2 M succinic acid, 100 mM Tris-HCl pH 7.0, 1% PEG MME 2000 and further optimized, the crystal did not appear and precipitate was observed in a vast majority of crystallization drops. Next, several crystallization screening experiments were carried out with the AtWRKY18^{DBD} protein sample. In parallel, some of the screens were set manually and others using high-throughput Robotic Sitting Drop Vapor Diffusion setup (Mosquito). The subsequent screens included Crystal Screens I and II and PEG/ion screen from Hampton Research as well as a six of Molecular Dimensions screens. Initial hit for ligand-free

AtWRKY18^{DBD} was found in Morpheus screen (Molecular Dimensions). The bunch of rod shaped crystals (Fig. 18) were observed within 7 days after setting in the drop where the reservoir was filled with 90 mM NPS, 100 mM Tris/bicyne pH 8.5, 37% MPD/PEG1K/PEG 3350. Few rod-shaped crystals were broaken off, frozen in liquid nitrogen and submitted for synchrotron radiation. When exposed to X-rays, the crystals produced a diffraction pattern of very low quality, it was not possible to index the images and to determine the space group and cell dimensions. Crystallization conditions were optimized; unfortunately the crystallization could not be repeated. Cocrystallization trials of AtWRKY18^{DBD} with DNA resulted precipitation in most droplets. Several crystallization screens were tested but without success.

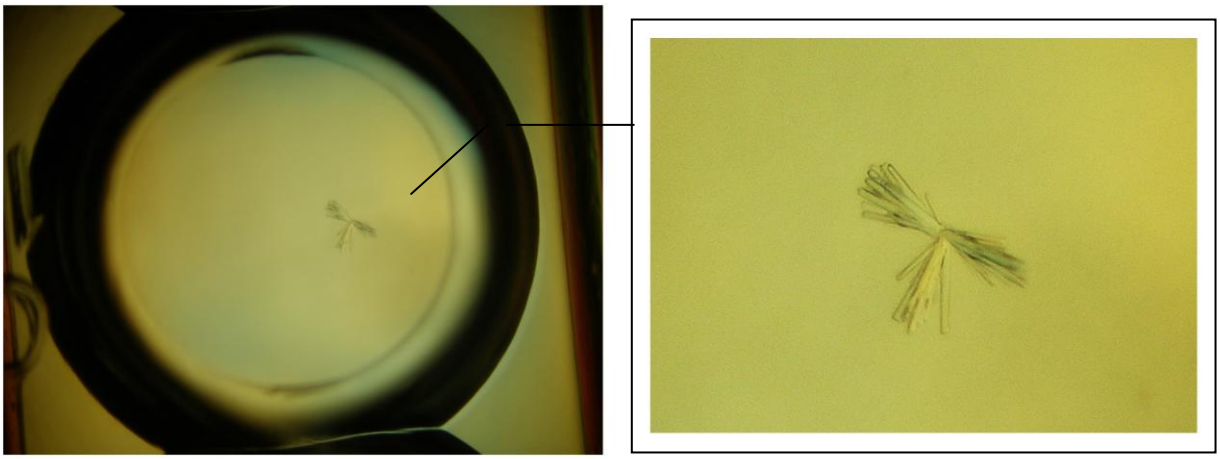


Fig. 18. Rod shaped crystal of AtWRKY18^{DBD}

3.4.3. DNA binding (EMSA, ITC)

The DNA binding activity of AtWRKY18^{DBD} was studied using electromobility shift assay (EMSA). DNA duplex was incubated with the AtWRKY18^{DBD} protein (0.15-10 μg), and specific DNA was detected by toluidine blue staining procedure as described in section Methods (EMSA). This test confirmed binding ability to 15 bp oligonucleotides long, used in experiments the same as for AtWRKY50. As shown in Fig. 19, recombinant AtWRKY18^{DBD} binding activity was clearly confirmed by mobility shift of DNA complexed with protein.

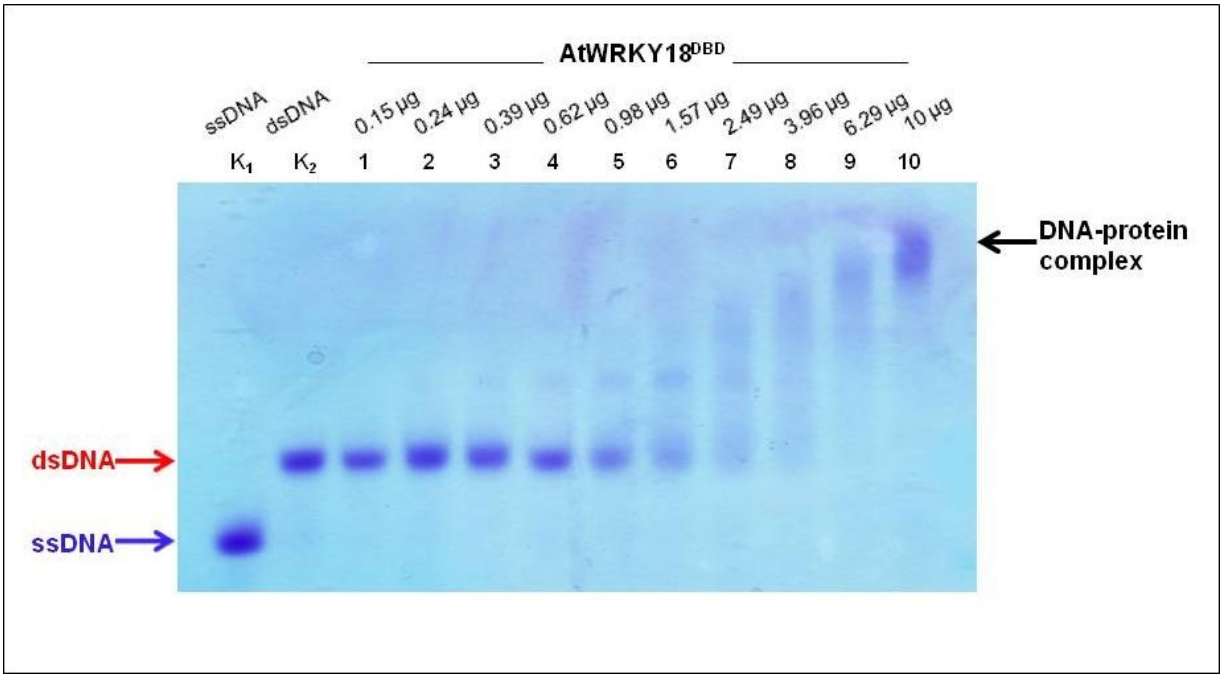


Fig. 19. Binding of AtWRKY18^{DBD} transcription factor to W-box containing DNA sequence. The specific retarded DNA-protein complexes are marked by black arrow, whereas red arrow designate positions of the free running probe. The blue arrow indicate the single stranded DNA position.
Lane K_1- ssDNA, K_2-dsDNA, 1-10-dsDNA with different protein amount 0.15-10 µg.

The Isothermal Titration Calorimetry experiment was performed to calculate DNA-binding constant. Interestingly in case of the AtWRKY18^{DBD}, the heat effect was endothermic, what was oposite to the exothermic effect obtained for AtWKY50. So far the effect of DNA dilution during measurements give significant background and it was difficult to determine K_d. The measurements require further optimization.

4. Discussion

WRKY proteins were studied for past 25 years. Since then the enormous progress of knowledge has been made in this field. The main experiments were performed *in vivo* and focused on their engagement in transcription regulation of genes under stress conditions. Research of plant systems based on microarrays, real-time PCR or others techniques require analysis of WRKY expression level in wild type plants, as well as in knock-out mutants or overexpressor but usually there was no need for obtaining recombinant WRKY proteins for functional studies.

The main goals of the thesis were structural studies of AtWRKY proteins using X-ray crystallography. The key and essential task in my research was to obtain full-length WRKY protein without fusion tag that is stable, pure and not forming aggregates. However, there is no data in literature on production of soluble form of any of recombinant full length copy of WRKY proteins, but only fragments corresponding to DNA binding domains or this domain with quite long flanking sequences were obtained. In recent years, the analysis of WRKY transcription factors in terms of their structure, function and mode of interaction with DNA and other components of the transcriptional machinery has been greatly facilitated by the identification of their genes. Collection of all WRKY gene sequences is now available in database: http://www.arabidopsis.org/browse/genefamily/WRKY.jsp

Recent progress in recombinant DNA technology permits engineering of fusion proteins labeled with specific affinity tags that greatly simplify the purification of the recombinant proteins which usually involves a single chromatography step with an appriopriate affinity resin. It facilitate synthesis of large quantities of recombinant proteins for use in DNA binding assays or protein crystallography as well as in protein-protein interaction and *in vitro* transcription assays.

4.1. WRKY cloning, overexpression and purification

Earlier experimental attempts to express full length WRKY proteins in *E. coli* systems were unsuccessful and the problem with overexpresson, purification and stability was highlighted in literarature [31]. Their expression was detrimental for bacterial growth or if bacterial growth was not significantly influenced, individual WRKY proteins were found in inclusion bodies. Almost all attempts to purify WRKY proteins from insoluble fraction employing denaturation

and subsequent renaturation steps resulted in misfolding of recombinant proteins and finaly in loss of W-box binding ability [31]. So far, a majority of research on WRKY's refers to *in vivo* studies of gene expression and transcriptional regulation. However, in few functional experiments the recombinant WRKY protein fragments or their biologically functional domains were used [21, 31].

WRKYs are unique for plant kingdom and they are absent in bacteria, fungi and animals. It might be the reason why the proper folding of recombinant proteins in bacteria is difficult. Due to this difficulties, usually analyses were done on fragments of selected proteins, mainly on DNA-binding domains [16]. It depends on purpose of experiment but the results obtained with isolated fragment of protein might not be reliable and may influence on final results of experiments. The same doubts cause analyses performed on fusion proteins (i.e. GST) that were not cleaved before experiment. It was observed especially when the binding to DNA or other partners was analyzed. The GST fused to native WRKY protein might influence the whole topology of the molecule.

To carry out structural studies of AtWRKY proteins, it was necessary to obtain full-length WRKY protein without fusion tag and protein preparation which would be stable, pure and not forming aggregates. Additionally to perform any crystallographic experiment, I needed at least several milligrams of prepared protein. Protein expression is an art of science and technology. Each protein is different and needs special treatment. Production of proteins, protein domains or fusion proteins that are soluble, functional and applicable for certain use is usually puzzle. Having the knowledge about the physicochemical properties of WRKY proteins taken from literature [31] and from consultations and discussions with researchers from Max Plank Institute in Cologne, Germany, that production of those proteins is laborious, I assumed that the solubility screening approach could be the best solution for this subject. Such approach could increase chance of success and allows selection of protein from whole family that meet all general criteria of crystallographic studies mentioned above.

From 74 WRKY proteins found in model plant *Arabidopsis thaliana*, I selected initially 12 proteins in a way that gave me the representative of each from 3 main groups and 5 subgrups according to their classification [15]. During the selection I also managed the availability of literature data about the functions of the individual WRKY in plant, as well as the size and their characteristic physicochemical parameters such as isoelectric point.

Optimization of protein production using a conventional approach when at a time only one

factor is modified, is laborious. It is due to the large number of potential factors and their interactions that can affect protein expression. A number of factors affect recombinant protein expression including, construct length, vector, cell strain, temperature, time, media, inducer concentration and additives in growing media. In order to express protein, many factors need to be examined experimentally, as it is difficult to predict *a priori* what will succeed. Due to limited time for research I decided to reduce the total number of experiments and eliminate at each step of whole procedure those proteins that have no potential for successful expression and purification for structural and functional studies. To obtained recombinant WRKY protein in amount sufficient for structural studies I decided to test 3 cloning strategies and vectors with different fusion proteins (MBP, NusA and Trx) to facilitate solubility using histidine affinity tag to make purification faster and more efficient.

For recombinant protein production, I used the most common bacterial expression systems. Although eukaryotic expression systems are also commonly used for producing functional recombinant proteins from higher organisms. Unfortunately eukaryotic expression systems yield usually small to moderate amounts of protein. In contrast, the *E. coli* expression system is the most widely used prokaryotic systems for genetic manipulations and easy to handle in any laboratory able to perform the basic molecular biology experiments. *E. coli* offers several advantages, including simple culture condition, inexpensive growth media, potentially very high expression levels and a simple scale-up process. Moreover there is plenty of comprehensive tools for *E. coli* systems, such as expression vectors, bacterial strains, protein folding and fermentation technologies to optimize expression of any protein. It is, however, not uncommon that overexpressed recombinant proteins fail to reach a correct conformation. Bacteria as prokaryotes are not equipped with the full enzymatic machinery to accomplish the required post-translational modifications or molecular folding. Hence, multi-domain eukaryotic proteins expressed in bacteria often appears non-functional. Also, many proteins become insoluble as inclusion bodies that are very difficult to recover without harsh denaturants and subsequent cumbersome protein-refolding procedures. Major disadvantages, apart from lack of protein processing or modification, include proteolytic degradation, poor expression of the protein due to toxicity or instability and potential problems with codon usage. The latter can be simply solved by application of the available commercial *E. coli* strains that posses extra copies of genes, which encode tRNAs that recognize the rare codons mainly AGG/AGA (arginine), but also AUA (isoleucine), CUA (leucine) and CCC (proline).

The first strategy of cloning I choose, was TOPO cloning system (Invitrogen) because is quite cheap, fast and does not require ligase. This method is also very efficient. Unfortunately, the vector I chose did not possess any fusion protein that might improve solubility, only His-Tag was present that facilitate purification using affinity chromatography. Without any complications I got all the 12 constructs of recombinant WRKY proteins although all of them were found insoluble. Heterologous expression of foreign genes in *E. coli* often leads to production of the misfolded proteins that form insoluble aggregates (inclusion bodies, IBs) [73]. Trials to optimize the growing condition (using different media, temperatures of induction, IPTG concentration) were unsuccessful.

Temperature and time are frequently critical factors, especially since these two variables often interact. In bacteria, there are some proteins that benefit greatly from a slower, longer induction, which generally requires low temperature. This approach minimize stress and improve recombinant protein production [71, 181, 185]. At high temperatures bacterial cells will reach a maximum density and eventually run out of nutrients, causing a point stress and finally cell death. If the protein of interests aggregates easily and cannot be overexpressed in a short frame, then lowering the temperature is essential. I tested expression of proteins at the temperature range of 15 to 37°C for 4-18 h. Expressing and inducing AtWRKY proteins at lower temperature had no influence on solubility. Although decreasing the growth temperature to 18-25°C resulted in a higher yield of protein, but it was accompanied by low solubility, as a large amount of protein was expressed as inclusion bodies. Thereafter, I performed expression at 18°C because of the greater protein yield. Subsequently, I attempted to increase the solubility of the protein by changing the concentration of IPTG (0.3-1 mM) but it have no significant influence on solubility, only on protein yield and thus routinely the 0.5 mM IPTG was used.

Using chemical additives or rich growth media sometimes have an effect on expression [115]. Additives are known to help forming correct conformation of proteins *via* different mechanisms. Improved solubility has been observed for the human phenylalanine hydroxylase when 0.4% glycerol was added to the growth medium [109]. It resulting in both higher solubility and activity. Accordingly, at the time of induction, I included chemical chaperone-glycerol at final concentrations of 0.4-1% to the LB medium but found that these additive was not beneficial for expressing AtWRKY proteins. For bacteria cultivation, I used also TB medium (Terrific broth) which is a phosphate buffered rich medium containing 20% more

peptone and about 4 times more yeast extract than LB broth. TB was supplemented with 0.4% glycerol as an extra carbon source. Buffered media have an advantage because prevent acidification during cultivation that may have negative effect on bacteria viability and protein production. TB medium increase the yield of protein production but not only of protein of interests but also all bacterial proteins. I observed this effect using TB medium instead of LB and it was undesirable, because AtWRKYs interact with other proteins. Large amount of contaminating proteins was present and therefore purification was tough.

During optimization of protein production, I tested also different strains of *E. coli* that have specific properties i.e. additional codons of rare amino acids, allowing disulfide bridges formation or toxic proteins production. There is a number of *E. coli* strains with genotypes engineered specifically to meet the needs of expressing recombinant proteins. I tested few commercially available strains derived from the BL21(DE3) cell line containing co-resident plasmids to address specific protein expression issues, including toxicity (C41 and C43), codon bias (STAR, Codon+RIPL) and folding (pLysS, RosettaGami, Origami). Unfortunately it does not improve solubility and any of those strains guarantee success in soluble AtWRKY proteins production.

Recombinant AtWRKY proteins were significantly expressed in the insoluble fraction, and were purified under denaturing conditions using 7.2 M urea to solubilize the insoluble proteins. However, sometimes it was possible to resolubilize the protein from the inclusion bodies (IBs) and further refolding [18, 158]; this approach was used in my research work as the last solution because protein refolding from IBs is not a straightforward process, often requiring an extensive trial-and-error approach [115] [110] [166]. Moreover IBs can be solubilized only by strong denaturants. Proteins from IBs must be solubilized and refolded into an active conformation. There are two important issues in recovering active proteins from IBs: (1) solubilization agent and (2) refolding method. Solubilization must result in monomolecular dispersion and minimum intra- or inter-chain interactions thus choice of solubilizing agents, e.g., urea, guanidine HCl, or detergents plays a key role in solubilization efficiency, structure recovery of the proteins from denatured state, and in subsequent refolding. Protein refolding means a change in protein conformation from unfolded to folded state. It is not a single reaction but competes with other processes, such as misfolding and aggregation, leading to inactive proteins. At high denaturant concentrations proteins are unfolded, well solvated and flexible, while in aqueous buffer are folded, rigid, and rather compact. Transfer of protein

molecules from high denaturant concentration to aqueous solvent will force them to shape into a compact structure and thus it should lead to refolding. However, such a drastic process usually does not work, since it will lead to misfolding and aggregation. Once misfolded or aggregated protein, in the absence of denaturants have no flexibility to disaggregate and refold into the native structure. There are few ways of removing denaturant: dilution, dialysis, size-exclusion chromatography, or solid phase refolding [18]. A success of applied method depends on the proteins and it need to be determined experimentally. Also the use of low molecular weight additives during the refolding process often helps in improving the yield of bioactive proteins from inclusion bodies [158]. However, the list of possible additives is quite extensive and it is difficult to guess which of these additives and at what concentration will work for particular protein. Optimization of buffer composition for the best solubility is achieved by screening. Reviewing the literature, additives such as urea, detergents, sugars, glycerol, amino acids, short-chain alcohols but also acetone, acetoamide, DMSO and PEG have been used to enhance the yield of bioactive protein during refolding [5, 49, 139, 164, 169, 193]. The most commonly used low molecular weight additives are: L-arginine (1–2 M) [49], urea or guanidine hydrochloride and detergents. All of tested AtWRKY 12 proteins obtained in pET151/D-TOPO vector were denatured with 7.2 M urea and renatured by dilution on column refolding or two variants of dialysis: one-step or stepwise using intermediate concentration of denaturant. Unfortunately, AtWRKY18 and AtWRKY30 were the only proteins among tested WRKYs, which were renatured from the inclusion bodies as a soluble proteins. To facilitate refolding, small molecules such as: urea, glycerol, arginine and glycine were added to refolding buffer. These additives reduces association of folding intermediates without interfering with refolding process. It is known from literature that several types of amino-acid derivative have been used as additives to decrease aggregation during refolding and heat-treatment of proteins [3, 167]. Moreover, simultaneous addition of charged amino acids: L-Arg and L-Glu at 50 mM to the buffer [65] or polyols i.e. glycerol [139, 164, 169] can dramatically increase the maximum achievable concentration of soluble protein and enhance protein stability [65]. It is necessary to reduce the extent of protein aggregation at each step of refolding starting from isolation to final purification. Therefore I also checked the WRKY protein behavior in buffer supplemented with the 2M urea during further purification steps. During refolding, the presence of anti-aggregation agents inhibits the aggregation of proteins. However, after refolding, when the urea, glycerol, arginine or glycine were completely

removed by dialysis, the proteins tended to precipitate precluding further purification and crystallization experiments.

In addition, it has been shown that the presence of reducing agents also improves protein solubility [5, 130]. AtWRKYs possess cysteine residues (AtWRKY18 - 5 Cys, AtWRKY30 - 3 Cys) in their sequences. Therefore the obtained misfolded proteins were subjected to refolding in the presence of a combination of different additives, glycerol and reducing agents, β-mercaptoethanol, Tris(2-carboxyethyl)phosphine (TCEP) and dithiothreitol (DTT). Presence of reducing agents in the refolding buffer might facilitate disulfide rearrangement [54]. Unfortunately, the procedure was unsuccessful and did not produce functional AtWRKY proteins. Despite many attempts of optimization mentioned above, renaturing the protein with urea and using additives failed to obtain the functional proteins. According to the procedures presented here, it was evident that the refolding method appeared inefficient. Refolded AtWRKYs were very unstable, easily precipitated and formed aggregates and therefore the purification and performing any crystallization or DNA-binding experiment were feasible. To conclude, the recovery of biologically active proteins from inclusion bodies is a complex process. In order to find the optimum refolding conditions, they have to be determined case by case. It is the major bottleneck in recovering high amounts of protein from inclusion bodies. The His-tag was mainly developed as an affinity tag [77, 161], but it is not commonly used to increase solubility. Because using pET151/D-TOPO as an expression vector failed to yield a suitable amount of soluble protein and the refolding trials were unsuccessful, I continued expression experiments by sub-cloning AtWRKY genes into the pMCSG and pET32a vectors. One commonly used strategy to increase solubility is to make a fusion with a protein that is known to have high solubility. Single His-tag usually works well with soluble, small or medium size proteins, but in many cases it lower the solubility. Interestingly, a comparative study of several N-terminal fusion tags and selected 32 human proteins indicated that protein tags, especially thioredoxin and MBP are preferable to the His-tag with respect to confer solubility, and appear to be particularly powerful [70]. There are several fusion proteins that have been shown to increase solubility in *E. coli* cells. For instance glutathione-S-transferase (GST) [160] or the maltose binding protein (MBP) [8, 43], which were developed as affinity fusion proteins, have been shown to increase solubility. The other more specialized fusion proteins are thioredoxin [108], NusA [38] and SUMO [134] that very often possess also His-tag and therefore do not require specific reagents or chromatographic columns for purification.

None of the tags is universal and often parallel expression experiments determine what the best strategy is. I decided to change cloning system and use vectors, that besides His-Tag possess also fusion protein known as solubility and folding promoting such as MBP (maltose binding protein), Trx (thioredoxine) and NusA (N-utilization substance). I learned from literature [31] that GST is ineffective as solubility enhancer in case of particular WRKY proteins. I utilized ligase independent (LIC) strategy for cloning [44, 98] into vectors with MBP or NusA and His-Tag for DNA-binding domain, that allowed simultaneous very simple cloning using the same primers compatible with all 3 vectors. For cloning into vector with TrxTag, the commonly used method with restriction enzymes was applied.

First of all I wanted to focus only on full length proteins. When it turned out that obtaining soluble WRKY was so difficult and rather impossible, I chose additional 6 WRKY proteins and decided to obtain DNA-binding domains (DBD) from two WRKYs. So far, only structures of DNA-binding domains of AtWRKY1 and AtWRKY4 transcription factor (both belongs to group I) are known. Therefore I chose domains from others WRKY TF group: AtWRKY18^{DBD} (group IIa) and AtWRKY30^{DBD} (group III). Although the DNA-binding domain is considered to be conserved and the hallmark of the whole protein family, the sequence alignment of the domains showed only 55-58% sequence identity with known structures. Thus I thought that it would be worth to compare structures of DBD from different groups of WRKY. Unfortunately overexpression trials revealed that even with NusA as fusion protein, 7 AtWRKY proteins were completely insoluble, 4 were insoluble when cleaved from NusA, 6 were soluble after cleavage but only in presence of additives. Again, similarly to proteins obtained with application of refolding method, they tended to aggregate, were unstable and easily precipitated.

In general, I tried to improve solubility and overcome aggregation by using additives such as glycerol, arginine, glycine and non-ionic detergent *n*-dodecyl β-D-maltoside (DDM). All these additives had an impact on proteins behavior in solution. Their preferentially interacted with the protein, changing the surface tension or amino acid solubility. First choice was glycerol, because it is the cheapest solution and is compatible with His-Tag resins. Unfortunately there was very little promising effects on solubility and aggregation. Among the additives, the best effect in improving solubility and reducing aggregation appeared at presence of L-arginine/HCl. It has been demonstrated that this additive works also for other proteins [3]. Moreover arginine is known to stabilize proteins during storage. Arginine at concentration 0.1

to 1 M may help to reduce protein aggregation and thus increase the purification yield of expressed protein [3, 167]. For plasminogen kringle 5 (pK5), supplementation buffer with glycine and arginine significantly improved the solubility of protein of interests [119]. Indeed in case of few AtWRKY proteins (see Results, Table 3, SA), addition both amino acids: glycine and arginine simultaneously allows to maintain protein in solution, prevents against precipitation and reduces aggregation. The suppression of protein aggregation by arginine cannot be readily explained by either surface tension effects or interactions with particular amino acids. Little is known about this mechanism. It has been suggested that favorably interactions of the guanidinium group from arginine with tryptophan side chains and peptide bonds might be one of the proposed mechanisms of suppression protein aggregation [117, 167].

In my opinion, application of amino acids as stabilizers, solubility enhancers or anti-aggregation factors are one of the best solutions. In living systems, proteins are surrounded by various solutes including macromolecules which stabilizes itself. Hence the isolated proteins are often marginally stable in aqueous solutions and must be stabilized by additional compounds. Use of additives like: sugars, polyols, amino acids, amino acid derivatives and polyamines is kind of lesson from nature because they naturally occur in cytoplasm. Certain amino acids, including glycine, alanine, proline and other are called osmolytes due to fact that these low molecular weight substances accumulate in the cells to raise osmotic pressure when they are exposed to high environmental salt concentrations. They do not interfere neither with protein activities i.e. enzymes nor functional structures and the use of L-Arg, L-Gly, L-Glu or L-His is approved for pharmaceuticals production [4]. Thus, amino acids can be used to stabilize proteins for structural studies where the maintaining native functions is pivotal. AtWRKYs purification was performed also in presence of non-ionic detergent. *n*-dodecyl β-D-maltoside (DDM) is most often used for the solubilization and purification of membrane proteins. It has a feature that low-dose did not interfere with protein crystallization. Detergents are commonly used to isolate membrane proteins because they interact with the hydrophobic sites of proteins, which are then solubilized by water layer, thus allow separation of membrane proteins [121, 137]. DDM has been applied in my experiments to prevent protein aggregation during purification but in case of WRKY protein it did not work.

Despite the additives, AtWRKY proteins still were unstable and prone to aggregation. They formed aggregates also with other bacterial proteins making their purification impossible.

Moreover, the yields of many protein preparations were too low to allow crystallization. Some proteins adhere to the filter units used during sample concentration. Few proteins were subjected to purification using an immobilized-metal affinity column, however, the purifications were unsuccessful, as a large amount of contaminating proteins was present. They also adhere to the plastic tubes used during purification. Finally, presence of aggregates can impact protein function or create point defects in crystal growth, while presence of additives might interfere crystal formation. It explains why I found this samples useless for crystallographic studies.

From all generated constructs, only one carrying full-length AtWRKY50 was expressed as a soluble and stable protein that do not form aggregates. It appeared a good candidate for crystallographic studies. Overexpression and purification was very efficient giving suitable amount homogenous protein for crystallization trails. The other 2 soluble, homogenous and stable proteins suitable for crystallization screening were DBD of the AtWRKY18 and AtWRKY30 but the amount of protein after cleavage was satisfactory only for AtWRKY18^{DBD}. Finally I focused on structural studies of AtWRKY50 and performed also few experiments on AtWRKY18^{DBD}.

4.2. WRKY crystallization

Protein crystallization is traditionally considered to be challenging due to the restrictions of the aqueous environment, difficulties in obtaining high-quality protein samples, as well as sensitivity of protein samples to temperature, pH, ionic strength, and other factors. Proteins vary greatly in their physicochemical characteristics, and therefore crystallization of a particular protein is rarely predictable. Determination of appropriate crystallization conditions for a given protein often requires empirical testing of many conditions before a proper ones are found. Moreover the nature of protein crystals is unique because individual molecules when pack in a repeating array are held together by a very weak noncovalent interactions.

The AtWRKY50 was extensively tested for crystallization conditions either in the apo form or in complex with DNA ligand. Also a lot of crystallization trails of AtWRKY18^{DBD} were made. The comprehensive crystallization trials engaged changing the crystal growth conditions (buffer, temperature, pH, protein and precipitant concentration) as well as protein modifications. In addition to manual screening for crystallization conditions, the AtWRKY50 protein and AtWRKY18^{DBD} were submitted to an extensive high-throughput crystallization

screens. In crystallization trials, except most widely used screens such as Structure screen, PACT or JCSG, I used also modern, alternative screens such as: Morpheus, PGA and Midas developed by Molecular Dimensions. PGA and Midas are based on alternative to PEGs polymeric precipitants. Additionally, Morpheus incorporate a range of low molecular weight ligands (e.g. alcohols, carboxylic acids, salts) and different type of precipitants that are also a cryo-protectants. Unfortunately, no well-diffracting crystals were obtained in any of tested conditions. In a comprehensive optimization process of AtWRKY50 cryatallization, the drops were individually inspected and some of them were selected for optimization. Only one set of conditions i.e. 0.8 M Na/K tartrate, 0.1 M HEPES pH 7.5 was found that produced weak-diffracting protein crystals (see Results, Fig.10). Unfortunately, obtained crystals of AtWRKY50 were always very small. Usually it was a shower of micro crystals. Occasionally a little bit bigger crystals appeared. Several available crystals were checked on synchrotron for diffraction quality, however the diffraction was weak. Only one crystal diffracted to circa 8 Å resolution, precluding any attempt to solve the crystal structure of the protein. This is known that crystals can deteriorate within days of growth thus fresh crystals were shoot. Different cryo-protectants were tested and the weak diffraction was obtained using glycerol. The attempts to improve diffraction quality of existing crystals were made also by special treatment of crystals, such as thermal annealing [103] and dehydration [55, 163] but without any improvement in diffraction quality. A lot of factors were considered to grow bigger crystals and to improve their quality. The observed "crystal shower" suggested lowering the nucleation point. It was done by reducing either the protein concentration or the precipitant concentration. Few different ratios of protein to crystallization solution were applied. The microcrystals appeared usually few hours after setting the crystallization. To slow down this process, decreasing the crystallization temperature was also tested. The use of temperature to control the level of super-saturation in crystallization experiments is well-established. Earlier studies have shown that temperature influences protein solubility, affecting both nucleation and crystal growth [63]. In case of AtWRKY50, changes in crystallization temperature resulting in non-reproducible experiment and no crystals appeared. In further work, I used few different techniques like makro, mikro and streak seeding and its combinations with dilution series [10] to optimize initial hit from screening. I used an existing microcrystal or crushed crystal for seeding but introducing them into new drop resulting again in shower of microcrystals. In some cases crystals did not appear.

I made also attempts to improve ability of AtWRKY50 to crystallize at the presence of different additives. The solution used for protein crystallization usually contained buffer, salts and precipitant. Recently, several studies have proposed that aggregation suppressors should be used in the crystallization of proteins and viruses [150] [126]. These studies hypothesizes that a new aggregation suppressor that is added as a fourth component could favor protein crystallization by suppressing protein aggregation under supersaturated conditions at the presence of precipitant [84]. In case of protein standard - hen egg-white lysozyme, addition of certain types of amino acids and amino acid derivatives, such as Arg, Lys and esterified or amidated amino acids, to solution containing various precipitants owing to decrease in aggregation. This additional component of the solution increases the probability of crystallization. These results suggest a simple method of improving the successful rate of protein crystallization. [84]. In addition to the crystals described in the Results section, some other crystallization experiments also produced crystals, but their diffraction quality did not improve. The crystals were grown by the sitting-drop vapor-diffusion method at room temperature and set using Mosquito Robotic setup. Each drop consisted of 0.6 µl protein solution and 0.3 µl reservoir solution equilibrated over 100 µl reservoir solution. Crystals of the AtWRKY50 grew in two very similar crystallization conditions from Morpheus screen (Molecular Dimensions): 30 mM $MgCl_2$, 30 mM $CaCl_2$, 20% PEG 550 MME, 10% PEG 20K that differ in buffers composition: 100 mM imidazol/MES pH 6.5 or 100 mM HEPES/MOPS pH 7.5. The morphology of obtained crystals was very similar, but only slight difference in size was observed. In this case, the AtWRKY50 protein crystallized when the protein used for crystallization was supplemented with 20 µl equimolar mixture (2 M) of arginine and glutamic acid [65] for each 100 µl of protein. Unfortunately obtained crystals did not diffract. As additives, I also used other amino acids such as lysine, serine and Additive screen (Hampton research) without any improvement of crystallization. The most commonly useful class of additives, are physiological ligands. They may be bound by the protein with consequent favorable changes in its physicochemical properties or conformation. These include natural ligands such as coenzymes and prosthetic groups, inhibitors, substrates, substrate analogs or products of enzymatic reactions, ions, DNA and other effector molecules. Often the protein-ligand complex is structurally defined and stable, while the ligand-free form is not. Often the former exhibit improved ability to crystallize, better crystal packing or order the crystal lattice when the latter have opposite properties. I tried co-crystallization of AtWRKY50 with DNA

but all the trials failed and I did not obtain crystals.

In order to improve the crystallization of AtWRKY50, the protein was subjected to chemical and enzymatic modifications like reductive methylation, in which the primary amino groups of surface residues are modified to tertiary amines. The AtWRKY50 protein has 16 lysine residues which is a quite high content for protein built of only 173 residues. The protein was also treated with proteases. Unfortunately, the modified protein did not crystallized in any of screen tested.

In spite of numerous crystallization trials, using several improving tricks, obtaining well diffracted crystals of full length AtWRKY50 failed. In this case the use of protein engineering, removal of flexible regions including terminal and interior loops, as well as replacement of residues that affect solubility or overexpressing conserved part of a given protein (functional domain) might be helpful. Recently proposed methodology to engineer residues that are exposed on protein surface is noteworthy. Surface sequence variants are designed to form intermolecular contacts that could support a crystal lattice [64]. This approach can be used to obtain crystals of proteins recalcitrant to crystallization or to obtain well-diffracting crystals if wild-type protein crystals yielded limited resolution. This method relies on the concept of surface entropy reduction (SER) by the replacement small clusters of two to three solvent-exposed residues characterized by high conformational entropy with residues with lower conformational entropy i.e. alanine residues. The surface entropy reduction prediction server (SERp server) was designed to identify mutations that may facilitate crystallization. Predicted mutations are based on an algorithm incorporating a conformational entropy profile, a secondary structure prediction, and sequence conservation. [64]. This strategy renders crystallization thermodynamically favorable and has been successfully used for crystallization of more than 15 novel proteins that were difficult to crystallize [34, 64].

Very often, protein of interests precipitate at concentrations required for crystallization. Amino acid substitutions were shown to increase protein solubility without altering structure or function because solubility is the function of surface hydrophobicity and can be altered by mutational modification of selected hydrophobic surface residues. Such strategy was applied and lead to successful crystallization and structure determination [37, 53, 86]. Usually hydrophobic residues exposed to solvent might be directly identified if structure of a homologous protein is available. If there is no reliable model, then hydrophobic residues are mutagenesis targets [125].

Many proteins contain highly flexible or even completely unfolded fragments dramatically interfering with crystallization. This is particularly true regarding large multi-domain proteins or signaling proteins, in which the unstructured regions often account for more than 50% of the molecule. To gain structural information about the stable fragments of such proteins, it is necessary to extract individual functional domains from full length protein. Since the crystallization of full length AtWRKY50 was unsuccessful, the crystallization trials of DNA-binding domain from closely related AtWRKY18 were taken. AtWRKY18 DBD crystals were obtained. Extremely thin needles grew from a single nucleation center, but they did not diffract. Trials of obtaining larger, well diffracting crystals failed.

4.3. Structural studies of WRKY proteins

Structural studies of WRKY proteins are very important to understand the mechanism of their interaction with both DNA and other potential binding partners. Each WRKY possess outside invariable DNA-binding domain other motifs responsible for interaction with different protein partners. Therefore, the global structure determination is essential to help understanding the complex mechanisms of signalling and transcriptional reprogramming of cell functioning under control of WRKY proteins. Unfortunately a solution structure is available only for highly conserved DNA-binding domain but not for full-length WRKY protein. There is no topological data regarding subgroup-specific motifs available. DNA-binding domains of AtWRKY transcription factor (both of the group I) were solved and deposited in PDB: one crystal structure of AtWRKY1 C-terminal domain (PDB code: 2AYD) [51] and two NMR structures of AtWRKY4 C-terminal domain (PDB codes: 1WJ2 and 2LEX) [190] [191]; the latter include its complex with DNA. There are still any structural studies of full-length WRKY proteins. The solution structures of AtWRKY4 DNA-binding domains consists of four β-strands. The N-terminal strand contains the WRKY amino acid sequence which binds DNA. The other three strands form novel zinc finger structure [190]. The crystal structure of AtWRKY1^{DBD} has revealed that this domain possess globular structure with five β-strands, forming an antiparallel β-sheet. A zinc-binding site is situated at one end of the β-sheet, between strands β4 and β5. DNA-binding residues of AtWRKY1^{DBD} are located at β2 and β3 strands [55]. There are differences between the structures related to the length. NMR structure of AtWRKY4^{DBD} lacking one beta strand, corresponds to the β1 from AtWRKY1^{DBD} crystal structure. The region assumed as WRKY domain in NMR structure seem to be short and may

not represent the structure of the whole domain.

Structural data of full length WRKY protein would help to elucidate how do they act as transcription regulators and let identify potential DNA binding sites of interacting partners.

Despite a lot of effort I put for my studies on WRKY proteins, reasonably there was a little chance to obtain diffracting crystals and crystal structure of full length AtWRKY50. Continuation of structural studies of AtWRKY18^{DBD} might be duplication of earlier reports and that is why I moved to biophysical and bioinformatics methods to characterize overall shape of AtWRKY50 protein and DNA-binding properties.

Bioinformatics analyses are based on algorithms implemented as computer programs to perform automated calculations, and data processing. Mathematical and statistical calculations allow to create the more accurate structural model. This type of results have limited certainty and needs to be interpreted with caution if they are not verified using experimental data. There have been a large number of diverse approaches to solve the structure prediction problem. In order to determine which methods were most effective a structure prediction competition called CASP (Critical Assessment of Structure Prediction) was founded [104, 131]. Genesilico Metadisorder service was choosen because it is one of the best predictors of protein disorder evaluated during independent tests (CASP8 [104] and CASP9). In case of AtWRKY50, the results derived from bioinformatics prediction using Metaserver were compared to experimental data. Independently similar and complementary results were obtained. The CD spectrum analyses allowed to deduce that AtWRKY50 lack of defined secondary structure and is partially disordered. This results where content of disordered regions was established as 40%, were a hint for the prediction of disordered regions using bioinformatics tools. Bioinformatics analyses of AtWRKY50 sequence revealed also that its sequence is rich in charged residues which generate high net charge and might lead to strong electrostatic repulsion. This feature causes low driving force for protein compactness. AtWRKY50 belongs to the IIc group of WRKY TF family. This protein group was distinguished because of presence in the amino acid sequence two characteristic motifs: basic stretch KAKxxQK, a potential nuclear localization sequences, followed by the amino acid sequence motif [K/R]EPRVAV[Q/K]T[K/V]SEVD[I/V]L and WRKY domain which are close to the C-terminus [56]. Moreover, AtWRKY50 sequence is also enriched in “disorder promoting” residues (especially Ala10, Gly10, Pro7, and Ser24), as described earlier for several intrinsically disordered proteins [52, 146, 174]. Obtained results from Metadisorder server that

predict intrinsically unstructured proteins from amino acids sequence only, were in good agreement with previously determined CD spectrum. The disorder prediction indeed confirmed that the AtWRKY50 is partially disordered with significant degree of disorder (~ 40%). Structure prediction of AtWRKY50 suggest that this protein consist of well ordered C-terminal region containing WRKY domain that include zinc finger motifs responsible for DNA binding and disordered N-terminal region from residues 1-107. Additional analysis of AtWRKY50 sequence using MoRFpred showed the presence of one MoRF motif [101]. Molecular recognition features (MoRF) are very short structure motives in intrinsically disordered proteins which are responsible for interacting with proteins or DNA and stabilization of protein structure. In protein of interest only one 5AA helical MoRF [46] was predicted which is located exactly between the disordered and ordered region.

In cooperation with dr. Kamil Szpotkowski, additional set of biophysical analyses including: dynamic light scattering (DLS), small angle X-ray scattering (SAXS) and infrared spectroscopy (FTIR) to characterize the overall structure of full-length *At*WRKY50 proteins in solution were performed. Results of those analyses were not published yet and were not included in present work, however they clearly allowed to confirm results of meta-analysis of disordered regions. Moreover, analysis of radius of gyration, hydrodynamic radius and volume-parameters obtained from SAXS and DLS, let us demonstrate the AtWRKY50 elongated shape, with N-terminal flexible tail and globular C-terminal domain.

Results of my studies may explain difficulties in the crystallization and obtaining diffracting crystals of the full-length protein which may be related to the flexibility of the long N-terminal protein fragment (residues 1-107). The presence of large regions of disorder can block protein crystallization or block formation of regular crystal lattice, because disordered proteins lack a single and stable conformation in solution, where the conformations fluctuate over time and over the population [132]. In case of AtWRKY50 the latter feature is rather possible. Three-dimensional structure of proteins that are entirely or partially disordered, typically cannot be determined by high resolution methods (X-ray crystallography and NMR). Summarizing, biophysics and bioinformatics methods cannot give the detailed structural information in comparison with crystallographic studies which gives subatomic resolution however our findings pointed out to the structural polarity of the AtWRKY50 macromolecule and its overall shape.

4.4. DNA-binding

Binding of natural ligands such as small molecules, substrates, cofactors, other proteins, nucleic acids or membranes may induce folding of unstructured proteins. Some transcriptional activators that recognize specific DNA motifs possess regions largely unstructured in the absence of DNA. The addition of DNA induce the transition of this region from unstructured to structured e.g. α-helical form [100]. The homogenous recombinant AtWRKY50 protein as well as AtWRKY18^{DBD} preserve their physiological ability to bind DNA and form stable complex what was confirmed by non-standard EMSA. In the experiment presented in dissertation, synthesized oligonucleotides (15 and 30 bps) containing the optimal binding site identical in sequence to the region of the parsley PR1-1 promoter containing one W-box [149] were used. W-box was found in the promoters of many stress related plant genes. W-boxes are often over-represented and clustered in the promoters of stress inducible genes as shown by transcriptome studies [123]. The W2 element was previously shown to be bound specifically by WRKY factors [149, 168]. WRKY transcription factors have been shown to bind to promoter regions containing the W-box consensus sequence TTGACC. To better understand the interactions between the AtWRKY proteins and their ligand-DNA, the binding studies using the ITC method were performed. ITC measures the heat effects during molecular association, but since the thermal changes are very small, a special care has to be taken to avoid side effects arising from dilution or simple mixing of the binding partners. To accomplish this, the AtWRKY50 protein and AtWRKY18^{DBD} were extensively dialyzed and the ligands were dissolved in the dialysis buffer. Using Isothermal Titration Calorimetry, AtWRKY50 was demonstrated to bind dsDNA containing one W-box sequence motif with K_d of 237 nM and 1:1 stoichiometry. As expected, AtWRKY50 protein binds one DNA molecule because this protein possess only one region responsible for DNA binding, the N-terminal WRKY domain. This result are in good agreement with those obtained for WRKY4-C domain where surface Plasmon resonance was applied to determine the K_d. The calculated value was 260 nM [190]. Unfortunately, it was impossible to compare the results presented in cited publication with results for AtWRKY18^{DBD} having standard WRKYGQK sequence since I did not designate K_d value.

AtWRKY50 belongs to a small IIc subgroup of WRKY proteins that possess within DNA-binding domain the WRKYGKK sequence instead of specific WRKYGQK present in majority of other WRKY proteins [172]. AtWRKY50 has the highest homology to tobacco NtWRKY12

(68% sequence similarity) which also belongs to this GKK subgroup [172]. In addition to AtWRKY50, only two other Arabidopsis WRKY proteins, AtWRKY51 and AtWRKY59, possess the WRKYGKK sequence but only AtWRKY50 has the highest similarity to NtWRKY12. NtWRKY12 is involved in transcriptional activation of PR-1a promoter [172] and AtWRKY50 might be an *Arabidopsis* analog. Interestingly, lacking the almost invariant WRKYGQK consensus in AtWRKY50 may confer the specific binding to the 5'-TGAC-3' W-box core. As proved in recently published experimental publication, AtWRKY50 exhibit increased affinity to the DNA with W-box probe and also a weak affinity to the mutated W-box probe, which was not observed for two others tested *At*WRKY11^{DBD} or *At*WRKY33^{cDBD}, that contain canonical WRKYGQK sequence. Moreover, the same effect was observed even when higher protein amounts were added [16]. These differences in binding specificity directly refers to altered residues in the DNA-binding domain. The exchange glutamine to positively charged lysine within the WRKYG**Q**K consensus might be responsible for the increased affinity to the mutated W-box-probe [16].

4.5. Conclusions

The expression of stable and functional proteins remains a bottleneck in many scientific attempts. Recently, enormous progress have been made in methods development which can improve the production of soluble and active proteins in heterologous expression systems. These include modifications to the expression constructs, the introduction new expression systems, development of new purification methods, sometimes requiring use of sophisticated equipement.

In presented studies, I developed an efficient method for overexpression and purification of recombinant AtWRKY50 and AtWRKY18^{DBD} protein retaining the biological activity of the DNA binding. The methods presented in this study allow to produce a significant amount of active AtWRKY50 and AtWRKY18^{DBD} in bacterial expression system for further functional and structural studies. Obtained recombinant proteins were pure enough to carry out crystallization experiments, however all attempts to crystallize all attempts to obtain well diffracting crystals of AtWRKY18^{DBD} or AtWRKY50 protein failed.

The CD spectrum and bioinformatics sequence analyses allowed to deduce that AtWRKY50 lack of well defined secondary structure and is partially disordered. This may explain difficulties in crystallization and failure to gain the main goal of the thesis - solving the crystallographic structure of the protein of interests.

ITC and EMSA analyses provided evidence for activity of recombinant AtWRKY50 protein and AtWRKY18^{DBD} toward DNA binding.

5. Materials and methods

5.1. Materials

5.1.1. Materials used in the experiments

Category	Material	Producer
PCR reagents	dNTP Mix (10 mM)	Sigma-Aldrich
	betaine	Sigma-Aldrich
	oligonucleotides	Genomed
	PureLink® PCR Purification Kit	Invitrogen
DNA electrophoresis	Agarose	Bio Shop
	Midori green	Nippon Genetics
	Bromophenol blue	Serva
	Xylene cyanol	Sigma-Aldrich
SDS-polyacrylamide gel electrophoresis	Acrylamide	Molekula
	N,N'-Methylene-bis-acrylamide	Molekula
	Glycine	Sigma-Aldrich
	Tris-HCl	Sigma-Aldrich
	SDS	BioRad
	TEMED	Sigma-Aldrich
	APS	BioRad
	β-mercaptoethanol	BioShop
	Glycerol	Sigma-Aldrich
	Coomassie Brilliant Blue R	Serva
	Methanol	POCH
	Glacial acetic acid	POCH
Bacteria cultivating media	Agar	Lab Empire
	Bio-Tryptone	Serva
	Yeast extract	Serva
	Sodium chloride	Sigma-Aldrich
	LB Broth	BioShop
	Carbenicillin	Polfa Tarchomin S.A.
	Kanamycin	Sigma-Aldrich
	IPTG	Biosynth
Protein chromatography	HisTrap HP (Ni Sepharose High Performance)	GE Healthcare
	HiLoad 16/60 Superdex 200	GE Healthcare

Crystallization	PACT premier HT-96	Molecular Dimensions
	Morpheus	Molecular Dimensions
	JCSG-plus HT-96	Molecular Dimensions
	ProPlex	Molecular Dimensions
	PGA	Molecular Dimensions
	Structure screen I and II	Molecular Dimensions
	Crystal screen	Molecular Dimensions
	Midas HT-96	Molecular Dimensions
	Grid Screen Ammonium Sulphate	Hampton Research
	Grid Screen Sodium Malonate	Hampton Research
	Grid Screen PEG 6000	Hampton Research
	Grid Screen MPD	Hampton Research
	Additive screen	Hampton Research
	Crystallization plates	Hampton Research
	Cover slides	Hampton Research
Buffer ingredients	Sodium phosphate Monobasic	Sigma-Aldrich
	Sodium phosphate Dibasic	Sigma-Aldrich
	Sodium hydroxide	Sigma-Aldrich
	Hydrochloric acid	Chempur
	EDTA	Sigma-Aldrich
	Imidazole	Molekula
	Tris-base	Bio Shop
	TCEP	Biosyntch
	DTT	Fluka
	Sodium Fluoride	Fluka
	Glicyne	Sigma-Aldrich
	Arginine	Sigma-Aldrich
	DDM	Biosynth
	Urea	Lab Empire
	Glycerol	Sigma-Aldrich
	Triton X-100	Sigma-Aldrich
	Sodium pyrophosphate	Sigma-Aldrich
	Ammonium molybdate	Sigma-Aldrich
	Iron sulphate	Fluka
Enzymes	Reverse transcriptase Superscript III	Invitrogen
	EcoRI	Promega
	XhoI	Promega
	Ligase T4	NEB
	Taq Polymerase	Fermentas
	Hot start KOD polymerase	Novagen
	T4 polymerase	NEB
	Pfu DNA polymerase	Fermentas

	DpnI restriction enzyme	NEB
	Bensonaze	Merck
	Lizozyme	Sigma-Aldrich
	TEV protease	Self production
Vectors	pMCSG7 pMCSG9 pMCSG48	Midwest Center for Structural Genomics, Argonne
	pET32a	Novagen
	pET151/D-TOPO	Invitrogen
Bacteria strains	TOP10	Invitrogen
	BL21Magic	Midwest Center for Structural Genomics
	BL21(DE3)STAR	Invitrogen
	BL21(DE3)C+RIPL	Agilent
	BL21(DE3)pLysS	Novagen
	OverExpressC41(DE3)pLysS	Lucingen
	OverExpressC43(DE3)pLysS	Lucingen
	Rosetta(DE3)pLysS	Novagen
	RosettaGami2(DE3)pLysS	Novagen
	Origami2(DE3)pLysS	Novagen
	Arctic Express(DE3)RP	Stratagen
Size markers	GeneRuler 1 kb DNA Ladder	Thermo Scientific
	GeneRuler 100 bp Plus DNA Ladder	Thermo Scientific
	PageRuler Plus Protein Ladder	Thermo Scientific
Protein handling	Amicon Ultra Centrifugal Filters Ultracel -30K	Millipore
	Amicon Ultra Centrifugal Filters Ultracel -10K	Millipore
	Millex-GV Syringe-driven Filter Unit	Millipore
	Ultrafree-MC Centrifugal Filter Device	Millipore
	Snake skin Dialysis Tubing 10K	Thermo Scientific
	Snake skin Dialysis Tubing 3.5K	Thermo Scientific
Other	GeneMATRIX Plasmid Miniprep DNA Purification Kit	EURx
	Plasmid Mini Kit	Qiagen
	RNeasy Plant Mini Kit	Qiagen
	Z-Competent E. coli Transformation Kit	Zymo Research

BSA	Sigma-Aldrich
azofoska	Inco Veritas
Whatman paper 3mm	Whatmann
PVDF membrane 0,22µm	Millipore
parafilm	Sigma-Aldrich
Petri plates	Sarsted

5.1.2. Oligonucleotides

Name	Primer sequence (5'-3')	Purpose
WRKY22_FW	CACCATGGCCGACGAT	
WRKY22_RE	TCATATTCCTCCGGTGGTAGT	
WRKY30_FW	CACCATGGAGAAGAACCATAGTAG	
WRKY30_RE	CTAAGAATAGAACCCACCAAATCC	
WRKY18_FW	CACCATGGACGGTTCTTC	
WRKY18_RE	TCATGTTCTAGATTGCTCCATTAAC	
WRKY29_FW	CACCATGGACGAAGGAG	
WRKY29_RE	CTAGTAATTCCATAAATACCC	
WRKY17_FW	FCACCATGACCGTTGATATT	
WRKY17_RE	TCAAGCCGAACCAAACAC	
WRKY56_FW	CACCATGGAAGGGGTTGAC	TOPO cloning
WRKY56_RE	TTACAGATCAGAAACTCTTGAG	
WRKY70_FW	CACCATGGATACTAATAAAGC	
WRKY70_RE	TCAAGATAGATTCGAACATGAAC	
WRKY40_FW	CACCATGGATCAGTACTCATC	
WRKY40_RE	CTATTTCTCGGTATGATTCTGTT	
WRKY6_FW	CACCATGGACAGAGGATG	
WRKY6_RE	CTATTGATTTTTGTTGTTTCC	
WRKY50_FW	CACCATGAATGATGCAGACACAAACT	
WRKY50_RE	TTAGTTCATGCTTGAGTGATTGTG	
WRKY51_FW	CACCATGAATATCTCTCAAAACCCTAGCC	
WRKY51_RE	TTAAGATCGAAGAAGAGAGTGTTGG	
W6_LIC_FW	TACTTCCAATCCAATGCCATGGACAGAGGATGGTCTGGTCT	
W6_LIC_RE	TTATCCACTTCCAATGTTACTATTGATTTTTGTTGTTTCCTTCGCCGT	LIC
W17_LIC_Fwd	TACTTCCAATCCAATGCCATGACCGTTGATATTATGCGTTTACCTAAGAT	

W17_LIC_Rev	TTATCCACTTCCAATGTTATCAAGCCGAACCAAACACCAAACCA	
W18_LIC_Fwd	TACTTCCAATCCAATGCCATGGACGGTTCTTCGTTTCTCGACAT	
W18_LIC_Rev	TTATCCACTTCCAATGTTATCATGTTCTAGATTGCTCCATTAACCTC	
W18_LIC_Rev	TTATCCACTTCCAATGTTATCATGTTCTAGATTGCTCCATTAACCTC	
W18dom_LIC_Fwd	TACTTCCAATCCAATGCCTCGACTGTCTACGTGCCTACTGAAA	
W18dom_LIC_Rev	TTATCCACTTCCAATGTTATCAAGCATTTGGACCCAAGTGGTTATG	
W22_LIC_Fwd	TACTTCCAATCCAATGCCATGGCCGACGATTGGGATCTC	
W22_LIC_Rev	TTATCCACTTCCAATGTTATCATATTCCTCCGGTGGTAGTGG	
W25_LIC_FW	TACTTCCAATCCAATGCCATGTCTTCCACTTCTTTCACCGACCTT	
W25_LIC_RE	TTATCCACTTCCAATGTTATCACGAGCGACGTAGCGCG	
W29_LIC_Fwd	TACTTCCAATCCAATGCCATGGACGAAGGAGACCTAGAAGCAATA	
W29_LIC_Rev	TTATCCACTTCCAATGTTACTAGTAATTCCATAAATACCCACTGAAGAACT	
W30_LIC_Fwd	TACTTCCAATCCAATGCCATGGAGAAGAACCATAGTAGTGGAGAGT	LIC
W30_LIC_Rev	TTATCCACTTCCAATGTTACTAAGAATAGAACCCACCAAATCCTCCA	
W30dom_LIC_Fwd	TACTTCCAATCCAATGCCAGTTCAAAAGTCAGAATTGCCCCTGGA	
W30dom_LIC_Rev	TTATCCACTTCCAATGTTACTAATTTGCAGCTTGAGAGCAAGAATGTATT	
W33_LIC_FW	TACTTCCAATCCAATGCCATGGCTGCTTCTTTTCTTACAATGGACAATA	
W33_LIC_RE	TTATCCACTTCCAATGTTATCAGGGCATAAACGAATCGAAAAATGAG	
W38_LIC_FW	TACTTCCAATCCAATGCCATGGAAATGAACTCCCCACACGAAAAG	
W38_LIC_RE	TTATCCACTTCCAATGTTATCAAAAGTAAAACTGATCATAACGATCCCAC	
W40_LIC_Fwd	TACTTCCAATCCAATGCCATGGATCAGTACTCATCCTCTTTGGTC	
W40_LIC_Rev	TTATCCACTTCCAATGTTACTATTTCTCGGTATGATTCTGTTGATACAATTTT	
W43_LIC_FW	TACTTCCAATCCAATGCCATGAATGGCCTCGTCGACTCTTCT	
W43_LIC_RE	TTATCCACTTCCAATGTTATTAGGTGAACTTAGAGAGGAACTGCAATT	
W50_LIC_Fwd	TACTTCCAATCCAATGCCATGAATGATGCAGACACAAACTTGGGGA	
W50_LIC_Rev	TTATCCACTTCCAATGTTATTAGTTCATGCTTGAGTGATTGTGGGAA	
W51_LIC_Fwd	TACTTCCAATCCAATGCCATGAATATCTCTCAAAACCCTAGCCCTAATTTTA	
W51_LIC_Rev	TTATCCACTTCCAATGTTATTAAGATCGAAGAAGAGAGTGTTGGTTC	
W53_LIC_FW	TACTTCCAATCCAATGCCATGGAAGGAAGAGATATGTTAAGTTGGGA	
W53_LIC_RE	TTATCCACTTCCAATGTTATTAATAATAAATCGACTCGTGTAAAAACGCGG	
W56_LIC_Fwd	TACTTCCAATCCAATGCCATGGAAGGGGTTGACAACACAAATCCTA	
W56_LIC_Rev	TTATCCACTTCCAATGTTATTACAGATCAGAAACTCTTGAGAGGAACT	
W62_LIC_FW	TACTTCCAATCCAATGCCATGAACTCTTGCCAACAAAAGGCTATGG	
W62_LIC_RE	TTATCCACTTCCAATGTTATCATGATGATAAGTCGTGAGATGTCCAG	
W70_LIC_Fwd	TACTTCCAATCCAATGCCATGGATACTAATAAAGCAAAAAAGCTTAAAGTTATGAAC	
W70_LIC_Rev	TTATCCACTTCCAATGTTATCAAGATAGATTCGAACATGAACTGAAGATAGA	
W18_trx_Fw	ATAGAATTCATGGACGGTTCTTCGTTTCT	pET32a cloning
W18-trx_RevR	CTCGAGTCATGTTCTAGATTGCTCCATTAAC	

W40_trx_Fw	AGTAGAATTCATGGATCAGTACTCATCCTCTTTG	
W40_trx_Rev	CTCGAGCTATTTCTCGGTATGATTCTGTTGA	
W56_trx_Fw	ATAGAATTCATGGAAGGGGTTGACAACA	
W56-trx_Rev	CTCGAGTTACAGATCAGAAACTCTTGAGAGG	
ITC_1_FW	CGCCTTGACCAGCGC	ITC
ITC_1_RW	GCGCTGGTCAAGGCG	
EMSA_FW:	TTATTCAGCCATCAAAAGTTGACCAATAAT	EMSA
EMSA_RE:	ATTATTGGTCAACTTTTGATGGCTGAATAA	
TOPO_T7F	TAATACGACTCACTATAGGG	sequencing
Oligo (dT)	TTTTTTTTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTCTTTTTTTTTTT TTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTT	Reverse transcription

5.1.3. Media and antibiotics

Sterilized media was used for growing bacteria. For sterilization, media with or without agar was autoclaved for 20 min at 121°C and cooled down prior adding heat instable antibiotics or other supplements. Heat instable compounds were filter-sterilized before use (0.22 µm Millipore filter).

LB *(1000 ml)*	
Bacto trypton	10 g
Yeast extract	5 g
NaCl	10 g
Agar	15g

TB *(1000 ml)*	
Bacto trypton	12 g
Yeast extract	24 g
Glicerol	4 ml
KH_2PO_4	2,31 g
K_2HPO_4	12,54 g

SOC *(100 ml)*	
Bacto trypton	2 g
Yeast extract	0,5 g
5 M NaCl	200 µl
2M $MgCl_2$ x $6H_2O$	500 µl
20% glukoza	2 ml

Antibiotic stock solutions (1000x conc.) were prepared as indicated and stored at -20°C.

Antibiotic	**concentration**
Ampicilin	50 mg/ml in H20
Kanamycin	50 mg/ml in H20
Gentamycin	25 mg/ml in H20
Carbencilin	50 mg/ml in H20
Chloramfenicol	34 mg/ml in EtOH

5.1.4. Buffers:

SDS-Page buffers

4xLB-sample buffer *(10 ml)*	**10xLaemmli** *(1000 ml)*		
1 M Tris-HCl pH6,8	2,5 ml	Tris base	30,25 g
glicerol	4 ml	glicyna	144 g
14,3 M β-merkaptoetanol	2 ml	SDS	10 g
bromophenol blue	1,5 ml		
SDS	0,8 g		

Solutions for preparing resolving SDS-PAGE gel *(10 ml)*		
	12 % gel	15 % gel
Water	3.3 ml	2.3 ml
Acrylamide	4 ml	5 ml
1.5 M Tris pH 8.8	2.5 ml	2.5 ml
10% SDS	0.1 ml	0.1 ml
10% APS	0.1 ml	0.1 ml
TEMED	0.004 ml	0.004 ml

Solutions for preparing stacking SDS-PAGE gel *(3 ml)*	
	5 % gel
Water	2.1 ml
Acrylamide	0.5 ml
1.0 M Tris pH 6.8	0.38 ml
10% SDS	0.03 ml
10% APS	0.03 ml
TEMED	0.004 ml

Staining and destaining buffers for SDS- polyacrylamide gels.		
	Staining	Destaining
Methanol	5%	5%
Glacial acetic acid	12.5%	12.5%
Coomassie brilliant Blue R-250	0.1%	

Agarose electrophoresis

1xTAE	
Tris-acetate	40 mM
EDTA	1 mM

Cloning

PCR for LIC insert preparation *(50µl)*	
10xbuffer	5 µl
$MgSO_4$ 25 mM	3 µl
Betaine 5 M	10 µl
dNTP 2 mM each	5 µl
F primer 10 µM	1.5 µl
R primer 10 µM	1.5 µl
KOD polymerase	1 µl
Template 50 ng/µl	1 µl
water	22 µl

T4 reaction for LIC *(15 µl)*	
NEB2 buffer	1.5 µl
dGTP/dCTP 100 mM	0.75 µl
DTT 40 mM	1.5 µl
T4 polymerase 3U/	0.15 µl
BSA 100x	0.15 µl
Template	1 µl
water	9.95 µl

Protein purification

Cell lysis buffer *(TOPO)*	
Tris-HCl pH 7.5/Na-phosphate pH 7.5	50 mM/20 mM
NaCl	500 mM
Glycerol	5-10 %
DTT	2-5 mM
Triton-X	0.5%
imidazole	10 mM
lizozyme	100 µg/ml

Cell lysis buffer *(LIC, pET32a)*	
Tris-HCl pH 7.5	50 mM
NaCl	500 mM
Imidazole	20 mM
TCEP	1-2 mM
lizozyme	100 µg/ml

Denaturation buffer (*inclusion bodies resolubilization)*	
Tris-HCl pH 7.5/Na-phosphate pH 7.5	50 mM/20 mM
Urea	7.2 M
DTT	5 mM
Glycerol	5 %

HisTrap binding buffer	
Tris-HCl pH 7.5	50 mM
NaCl	500 mM
Imidazole	20 mM
TCEP	1-2 mM

HisTrap elution buffer	
Tris-HCl pH 7.5	50 mM
NaCl	500 mM
Imidazole	300 mM
TCEP	1-2 mM

Dialysis buffer	
Tris-HCl pH 7.5	50 mM
NaCl	500 mM
TCEP	1-2 mM

Gel filtration buffer	
Tris-HCl pH 7.5	20 mM
NaCl	200 mM
TCEP	1-2 mM
Optional additives:	
DDM	0.01%
Arginine	50-250 mM
Glycine	50-250 mM

5.2. Methods

Our research group collaborate with Max Planck Institute for Plant Breeding Research In Cologne, Germany. The coding sequences of the *Arabidopsis thaliana* WRKY genes were obtained from a cDNA library (property of the Max Plank Institute for Plant Breeding Research in Cologne, Germany) as pDONOR201 vectors (Gateway system, Invitrogen). I received from dr Imre Sommsich group cDNA of 13 AtWRKY transcription factors: AtWRKY6, AtWRKY11, AtWRKY17, AtWRKY18, AtWRKY22, AtWRKY29, AtWRKY30, AtWRKY33, AtWRKY40, AtWRKY50, AtWRKY51, AtWRKY56 and AtWRKY70.

The coding sequences of AtWRKY25, AtWRKY38, AtWRKY43, AtWRKY53, AtWRKY62 were obtained from fresh *A. thaliana* plant material by isolation of total RNA followed by reverse transcription and PCR with specific primers suitable for WRKY sequences . The sequence agreement was confirmed by comparison with TAIR database.

5.2.1. Recombinant protein production

5.2.1.1. Plant growing

A. thaliana seeds were sown in pots with steril soil mixed with sand in 1:1 ratio. The culture was grown in greenhouse without temperature control. Plants were watered as needed but once a week with addition of 3% azofoska (Inco Veritas). They were grown until flowering.

5.2.1.2. Isolation of total RNA

Total RNA was isolated separately at different developemental stages: leaves, stems and flowers. Isolation of total RNA was performed with use of the RNeasy Plant Mini Kit from Qiagen. Plant material (100 mg) was frozen at -70 ° C, then crushed in liquid nitrogen to obtain a uniform powder. Afterwards 450 µl of homogenization buffer RLT was added to each sample. After homogenization using shaker, lysate was subjected to 3 minutes incubation at 56°C and then applied on QIAshredder columns to remove cell debris and homogenize the lysate. Columns were centrifuged (2 min /10000 rpm /room temperature). To the cleared lysate, 0.5 volumes of 96% ethanol was added and the samples were mixed by pipetting. The resulting mixtures were transferred to an RNeasy spin columns for isolation of total RNA and followed further procedure provided by the manufacturer. Undesirable genomic DNA contamination was removed using “on column” digestion by DNase I (Qiagen). Buffer RW1

used for washing columns was supplemented with the DNase I (27U). The incubation lasted 15 min at room temperature. Total RNA was eluted from the column with 40 µl RNase-free H_2O (Ambion). A quality of RNA was evaluated by electrophoresis. The yield of total RNA was determined spectrophotometrically using a NanoVue Plus Spectrophotometer (GE Healthcare).

5.2.1.3. Reverse transcription

The reverse transcription reaction was performed using the SuperScript III Reverse Transcriptase (Invitrogen). For first strand of cDNA synthesis, 500 ng of total RNA was used. The reaction (Invitrogen) proceeded in two stages. In the first stage, mixture of total RNA with oligo dT primers, dideoxynucleotide (dATP, dCTP, dGTP, dTTP) and water was incubated (65°C/5 min). At this stage ribonucleic acid denaturation and primer binding occured. In a next step the enzyme reverse transcriptase, M-MLV (ang. Moloney Murine Leukemia Virus), the reaction buffer, dithiothreitol, and stabilizer RNA (RNase OUT) were added to the reaction mixture,. Transcription of genetic information from mRNA to cDNA was carried out at 50 ° C for 60 min, and then the enzyme (reverse transcriptase) was inactivated by 15 min incubation at 70 ° C.

The reaction was performed in a volume of 20 ul according to the protocol recommended by manufacturer.

Obtained cDNA was directly used as template for PCR with specific forward and reverse primers to individual AtWRKY sequence.

5.2.1.4. Cloning of the WRKY protein coding sequences

For cloning of AtWRKY transcription factors, several techniques utilized different fusion-tags were used.

5.2.1.4.1. TOPO® Cloning

pET151 / D-TOPO utilizes a highly efficient, 5-minute cloning strategy ("TOPO® Cloning") to directionally clone a blunt-end PCR product into a vector for high-level, T7-regulated expression in *E. coli*. N-terminal fusion tags: V5 epitope, 6xHis and TEV protease cleavage site simplify purification of recombinant fusion proteins.

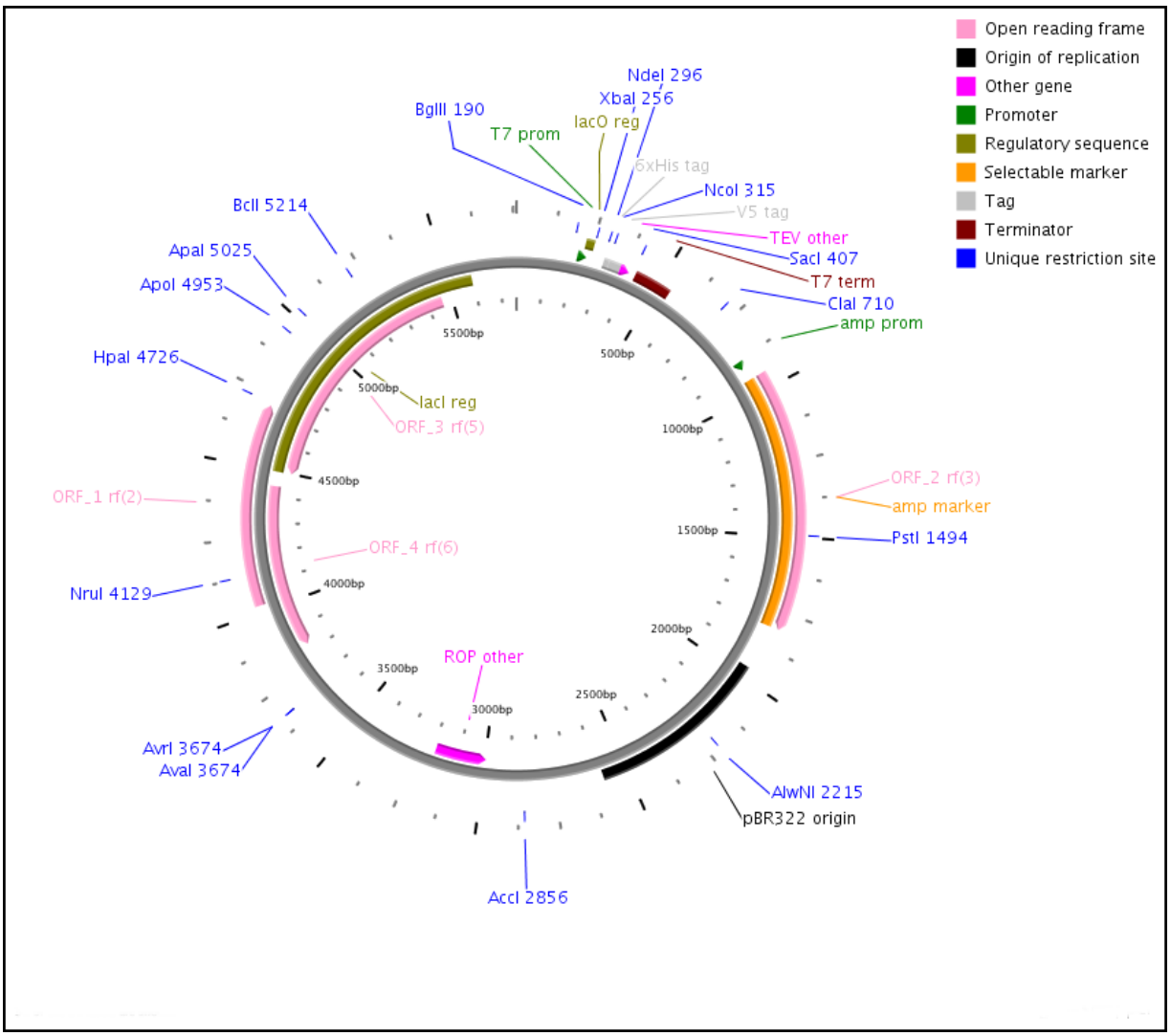


Fig. 21. Map of pET151 / D-TOPO. Image made with PlasMapper.

The coding sequences of the AtWRKY genes were PCR amplified with specific F and R primers (the forward primers were extended by 4-nucleotide CACC fragment at the 5 '). The PCR products were checked by electrophoresis on 1.5% agarose gel and then purified using a PCR Purification Kit (Qiagen). The prepared DNA were cloned into the expression vector pET151 / D-TOPO (Invitrogen) containing the 18-nucleotide DNA fragment encoding a histidine tag (6xHis). Direct cloning (omitting the step of restriction enzyme digestion and ligation) was made possible by the presence of the enzyme topoisomerase covalently bound to a targeting vector. In a next step chemicompetent *E. coli* cells (One Shot TOP10) were transformed. Positive clones were used for plasmid isolation (using a commercial Plasmid Mini Kit, Qiagen). The correctness of the cloned sequences was verified by DNA sequencing.

Bacteria expression strain *E.coli* BL21 Star (DE3) (Invitrogen) was transformed with the plasmid constructs harbouring WRKY coding sequences.

With this method I prepared the following recombinant proteins: AtWRKY6, AtWRKY11, AtWRKY17, AtWRKY18, AtWRKY22, AtWRKY29, AtWRKY30, AtWRKY33, AtWRKY40, AtWRKY50, AtWRKY51, AtWRKY56 and AtWRKY70.

5.2.1.4.2. Ligase Independent Cloning (LIC) into pMCSG7, pMCSG9 and pMCSG48 plasmids

Ligase independent cloning (LIC) is a simple, fast and relatively cheap method to produce expression constructs. This creative technique uses the 3' → 5' exo activity of T4 DNA Polymerase to create very specific overhangs in the expression vector. PCR products with complementary overhangs are created by adding appropriate extensions into the primers and treating them with T4 DNA polymerase as well. Addition of dGTP to the reaction limits the exonuclease processing to the first complementary C residue, and not present in the designed overlap, where the polymerization and exonuclease activities of T4 DNA Polymerase become "balanced". The annealing of the insert and the vector is performed in the absence of ligase by simple mixing of the DNA fragments. Joined fragments have 4 nicks that are repaired by *E.coli* during transformation. This process is very efficient because only the desired products can form. LIC method [44, 98] was applied for obtaining most of the constructs. Those pMCSG-LIC vectors (provided by Midwest Center for Structural Genomics, Argonne, IL, USA) allows to clone the target gene using the same primer pair into a plasmid with different fusion tags:

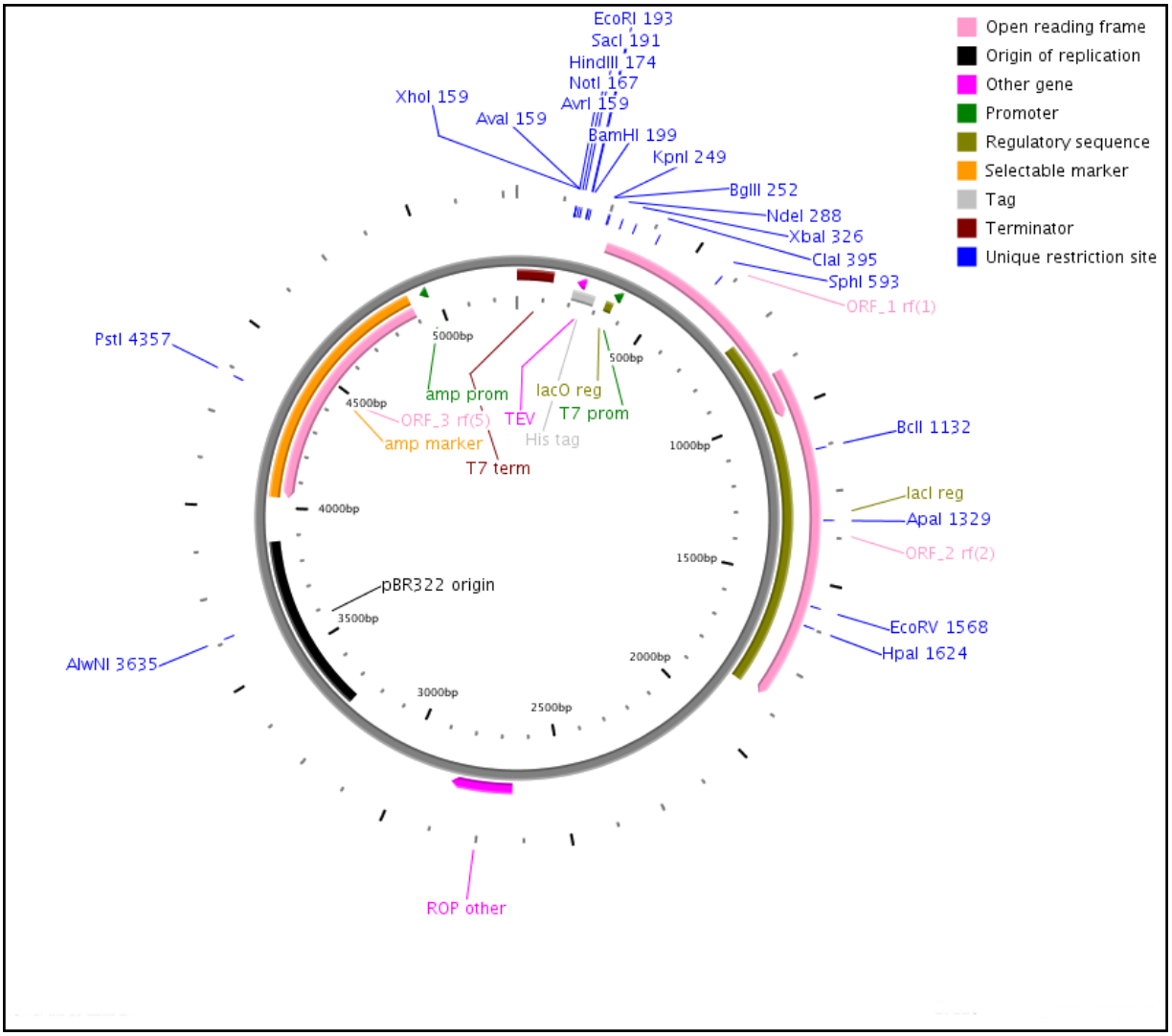


Fig. 22. Map of pMCSG7 vector that posses N-HisTag. Image made with PlasMapper.

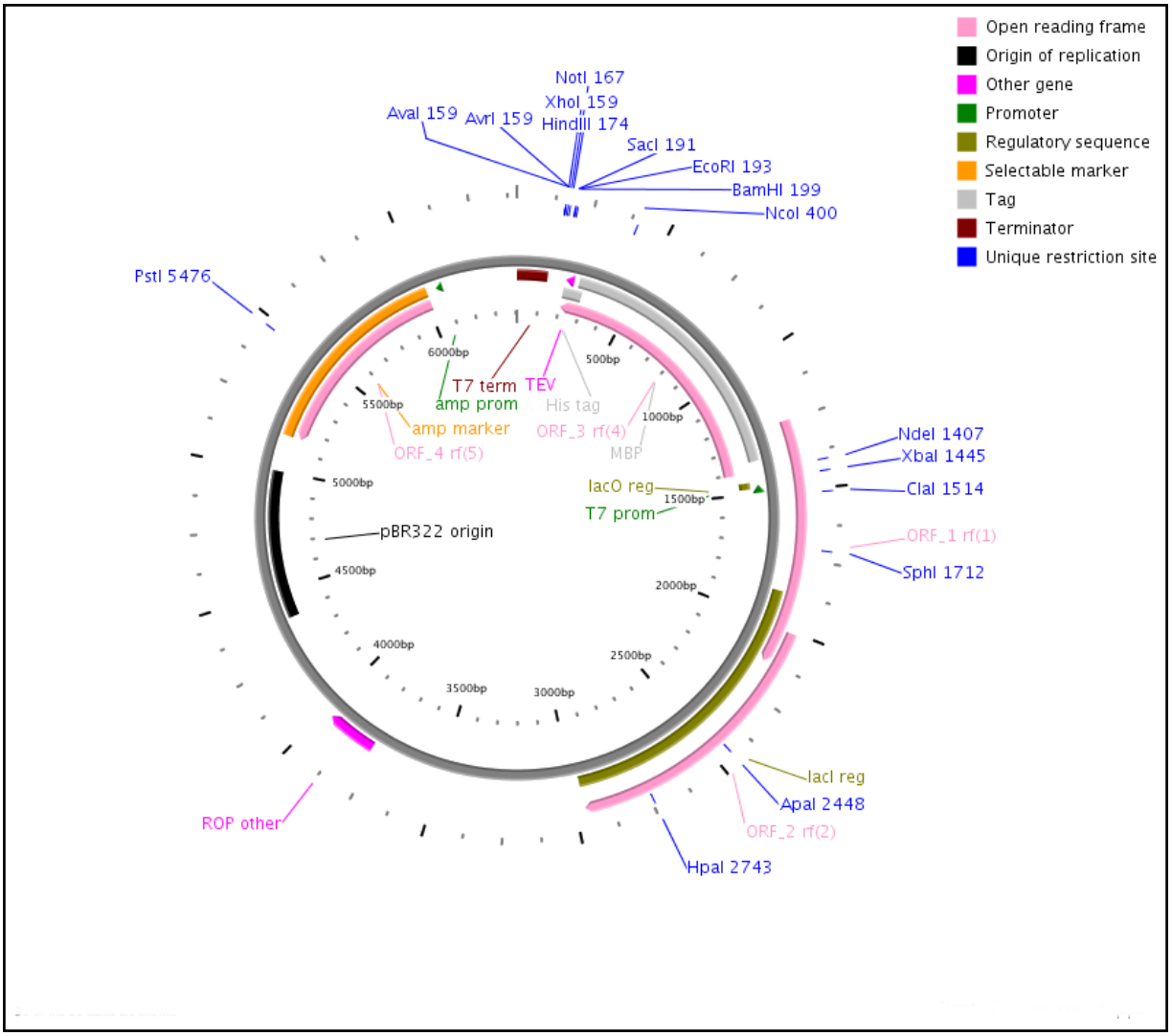


Fig. 23. Map of pMCSG9 vector. Image made with PlasMapper.

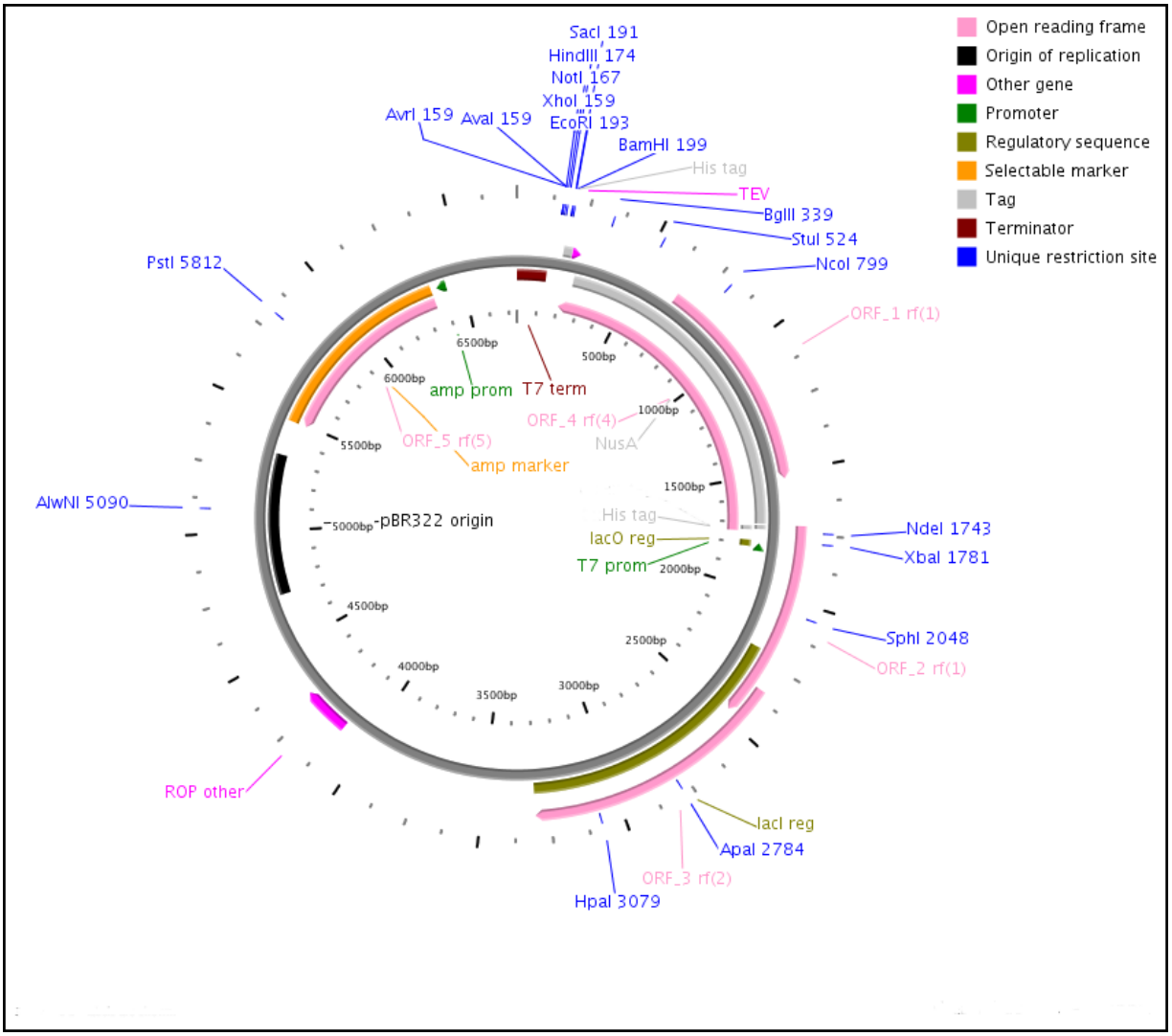


Fig. 24. Map of pMCSG48 vector. Image made with PlasMapper.

Inserts were PCR amplified and vectors are made linear either by restriction enzyme digestion or by PCR. To create an insert with complementary overhangs to the pMCSG-LIC vectors the following primers have to be used:

Forward primer: **TACTTCCAATCCAATGCC** - gene of interest

Reverse primer: **TTATCCACTTCCAATGTTA** - gene of interest (reverse complement)

The **forward primer** should contain: the complementary overhang (shown in **blue**) and the gene of interest should start with the ATG start codon, and should be long enough to overlap with the gene of interest to give a melting temperature of 60°C or more.

The **reverse primer** should contain the complementary overhang (shown in **red**), one or more stop codons (e.g. TAA) if no C-terminal tag is used, and a long enough overlap with the reverse complement strand of the gene of interest to give a melting temperature of 60°C or more. Insert and the vector after treatment with T4 polymerase have sticky 5' ends of fifteen bases in length. The sticky ends are complementary and allow plasmid circularization. A resulting plasmids were used directly for transformation of chemicompetent *E. coli* cells (One Shot TOP10). The transformed cells were spread over a medium supplemented with selective antibiotics that allow to grow only colonies that have incorporated a proper resistance gene. The colonies were subsequently PCR tested for presence of insert. Positive clones were used for plasmid isolation (using a commercial Plasmid Mini Kit, Qiagen). The correctness of the cloned sequences were verified by DNA sequencing to exclude accidental mutations occurred during the entire procedure. Bacterial cells of the expression strain *E.coli* BL21 Star (DE3) (Invitrogen) were transformed with the prepared plasmids.

With this method I prepared recombinant full length AtWRKY6, AtWRKY11, AtWRKY17, AtWRKY18, AtWRKY22, AtWRKY25, AtWRKY29, AtWRKY30, AtWRKY33, AtWRKY38, AtWRKY40, AtWRKY43, AtWRKY50, AtWRKY51, AtWRKY53, AtWRKY56, AtWRKY62 and AtWRKY70 proteins as well as constructs with proteins fragments containing DNA binding domains from AtWRK18 and AtWRKY30 proteins.

5.2.1.4.3. Cloning into pET-32a(+) vector

pET-32a(+)vector (Novagene, USA) is designed for cloning and high-level expression of peptide sequences fused with the 109aa Trx•Tag™ thioredoxin protein [108]. This vector is able to express a fusion protein with a 6-histidine tag at thrombin site and a T7 tag at the N-

terminus. These additional amino acids together with thioredoxin increase the size of expressed protein by 15 kDa.

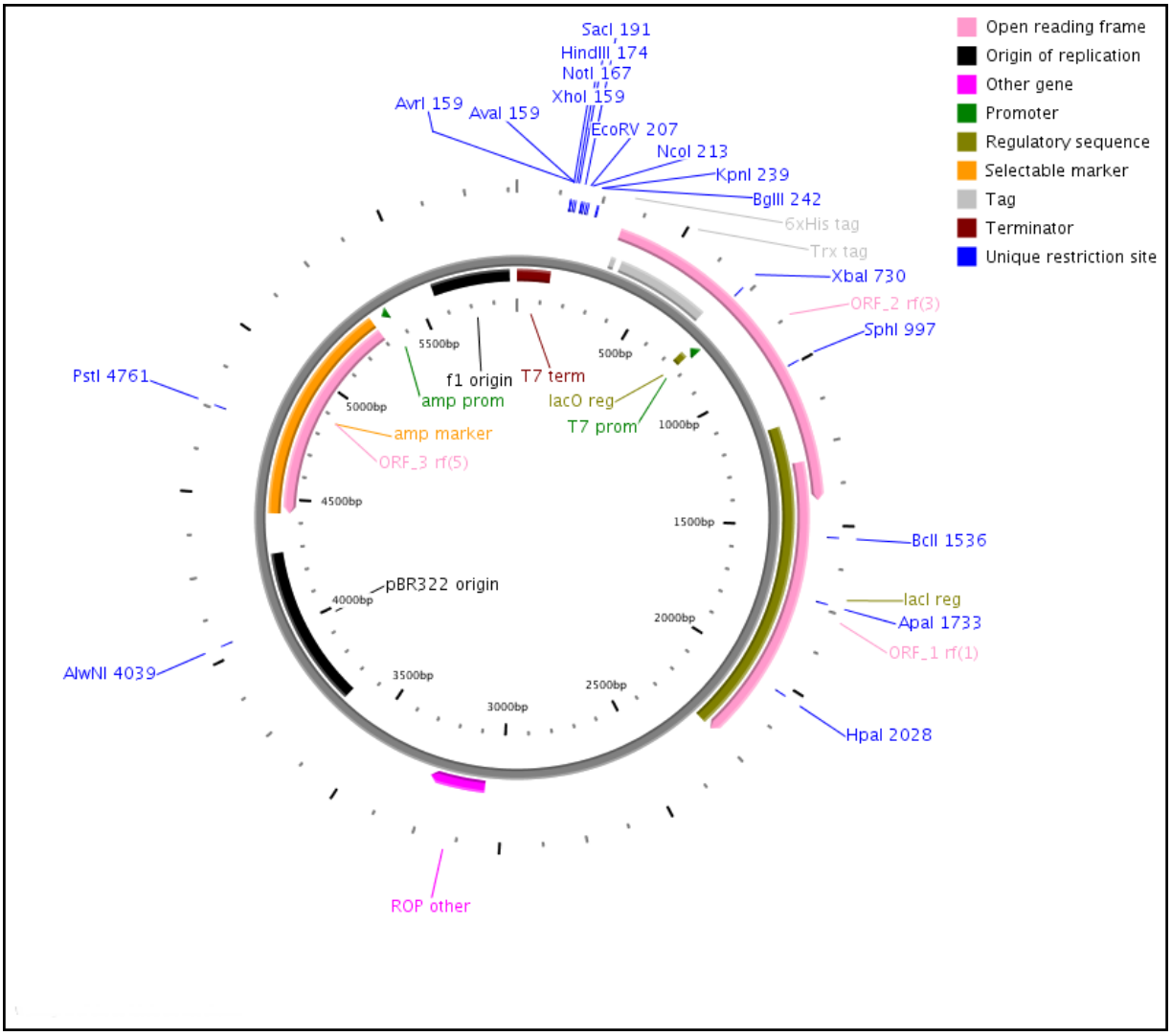


Fig. 25. Map of vector pET-32a(+) designed for cloning and high-level expression of peptide sequences fused with the 109aa Trx•Tag™ thioredoxin protein. Image made with PlasMapper

Specific primers were designed: forward with introduction of *Eco*RI and TEV recognition site at 5'and reverse with *Xho*I recognition site at 5' end. The coding sequences of the AtWRKY genes were amplified by PCR The PCR product and pET32a were digested with *EcoRI* and *Xho*I. Prior ligation, vector was dephosphorylated using calf intestine alkaline phosphatase.

For ligation, the PCR amplified WRKY coding DNA and pET32a were mixed in the 10:1 ratio. and ligated by T4 DNA ligase (Fermentas) at 4°C overnight. The ligated products were initially propagated in Top10 *Escherichia coli* competent cells (Invitrogen, USA). Then

colonies were further analyzed by restriction endonuclease digestion and colony PCR. WRKY genes of the recombinant plasmids were sequenced. The recombinant pET32a-WRKY constructs extracted from Top10 *E. coli* cells were transformed into *E. coli*, BL21 Star (DE3) a host strain (Invitrogen, USA).

With this method I prepared recombinant AtWRKY18, AtWRKY40 and AtWRKY56 proteins.

5.2.1.5. Overexpression of recombinant AtWRKY

Routinely for expression of recombinant AtWRKY proteins, the strains BL21(DE3)Star (Invitrogen) (TOPO cloning, pET32a) or BL21Magic (Midwest Center for Structural Genomics, Argonne, IL, USA) (LIC) were used. For optimization of protein overexpression/production few more *E. coli* strains were used: BL21-CodonPlus(DE3)-RIPL (Agilent Technologies/Stratagene), BL21(DE3)pLysS (Novagen), C41(DE3)pLysS (Lucigen), C43(DE3)pLysS (Lucigen), Arctic Express(DE3)RP (Agilent Technologies), Origami2(DE3)pLysS (Novagen) and RosettaGami2(DE3)pLysS (Novagen). Transformed bacteria were grown with vigorous shaking (37°C at 210 rpm) in a liquid LB or TB medium with selective antibiotics (e. g. carbenicillin- 100 μg/ml, kanamycin- 25 μg/ml, chloramfenicol- 34 μg/ml, gentamycin-20 μg/ml, tetracycline-12.5 μg/ml, respectively) to an optical density of the culture OD_{600}~ 1.0. In some cases, the media were supplemented with additives such as 0.4-1% glycerol or 50 μM $ZnCl_2$. Protein expression was induced with IPTG (final concentration 0.3-1.0 mM). Bacteria were grown for a further 4-18 hours in temperature range 15°C -37°C and were harvested then by centrifugation (5000 rpm /4°C for 30 min).
Zero-time point aliquot (uninduced culture) was used as control. The samples collected after harvesting were analysed for expressed protein by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

5.2.1.6. Purification of soluble and insoluble fraction of recombinant AtWRKY proteins

Routinely the following procedure was used. Pelleted cells were lysed in the buffer containing 50 mM Tris pH 7.5, 0.5 M NaCl, 5% glycerol, 100 μg/ml lysozyme, 2 mM DTT. After incubation on ice for 30-60 min, the solution was sonicated on ice for 4-min with appropriate intervals (20 sec.) for cooling. The extract was treated with 250 Unit of Benzonase (Sigma) on ice for 15 minutes. The lysate was centrifuged at 15000 g for 30 min. at 4°C.

For purification of proteins obtained from TOPO system, the supernatant was subjected to purification using an ÄKTA Purifier system (GE Healthcare). In the first step, the proteins were applied on a HisTrap™ column equilibrated with binding buffer containing 50 mM Tris pH 7.5, 500 mM NaCl, 5% glycerol, 2 mM DTT. After binding, the column was washed with 20 mM imidazole in 50 mM Tris pH 7.5, 500 mM NaCl, 5% glycerol, 2 mM DTT. The protein was eluted from the column using a stepwise elution 20-300 mM gradient of imidazole in 50 mM Tris pH 7.5, 500 mM NaCl, 5% glycerol, 2 mM DTT. The eluted protein was afterwards passed through the HiPrep™ 26/10 column to remove imidazole for binding buffer (50 mM Tris pH 7.5, 500 mM NaCl, 5% glycerol, 2 mM DTT). Imidazole-free protein solution was incubated for 4-6 hours at RT or 37°C with a His-tagged TEV protease (60 μg TEV protease/1 mg His-tag protein) to cleave off the His-tag. Subsequently, the sample was applied on a HisTrap™ column to remove the TEV protease, the His-tag and any undigested fusion protein. The first flow-trough was using Amicon Ultra-4 centrifugal devices with 10 kDa cut-off membrane (Millipore). Fractions of highest absorbance from the peak were collected and analyzed by SDS-PAGE.

For isolation of protein from insoluble fraction (inclusion bodies) the pellet obtained from crude extract was washed twice with solubilizing buffer containing 50 mM Tris pH 7.5, 0.5 M NaCl, 5% glycerol, 2 mM DTT and 1% Triton X-100 and centrifuged at 15,000 g for 30 min. at 4°C. The pellet was solubilized in 10 ml solubilizing buffer containing 50mM Tris pH 7.5, 0.5 M NaCl, 5% Glycerol, 7.2 M urea, 25 mM DTT. The suspension was incubated for 1 hour at 37°C with gentle shaking. Insoluble material was removed by centrifugation at 15000 g for 30 minutes. Protein concentration of the sample was determined and adjusted to 1 mg/ml using solubilizing buffer. To renaturate the target protein, dialysis method was used. Also on column refolding was tested. The latter sample was dialyzed against renaturation buffer containing 50 mM Tris pH 7.5, 500 mM NaCl, 5% Glycerol, 2 mM DTT. The sample was concentrated in 50 ml Amicon stirred cells (Millipore) under nitrogen pressure at 58 PSI by using ultrafiltration membranes with 10 kDa cut-off pore size (Millipore) or using Amicon Ultra 10 filters (Millipore). Concentrated sample was centrifuged at 15,000g for 10 minutes to remove insoluble particles. The samples were analysed by SDS-PAGE.

For purification of proteins obtained from pMCSG vectors the following procedure was used.

The supernatant was subjected to purification using a column packed with 6 ml of Ni-Sepharose HP resin (GE Healthcare) and connected to VacMan (Promega). the column was washed 4 times

with 30 ml of binding buffer (50 mM Tris pH 7.5, 500 mM NaCl, 1 mM TCEP) and the purified protein was eluted with 15 ml of elution buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 300 mM imidazole, 1 mM TCEP). The His-tag was cleaved with TEV protease (60 μg TEV protease/1 mg protein, overnight at 4°C). The excess of imidazole was removed by dialysis (overnight at 4°C). The dialysis was performed simultaneously with TEV cleavage. Subsequently, the sample was applied on column packed with Ni-Sepharose HP resin to remove the fusion tag, His_6-tagged TEV protease and any undigested fusion protein. The flow-through was collected, concentrated to 4 ml volume using Amicon Ultra-4 centrifugal devices with 10 kDa cut-off membrane (Millipore) and applied on a HiLoad Superdex 200 16/60 column (GE Healthcare) equilibrated with a buffer composed of 50 mM Tris-HCl, pH 8.0, 200 mM NaCl and 1 mM TCEP. Peak fractions were collected and analyzed by SDS-PAGE. The procedure was modified to the individual protein. The buffer composition were modified depending on the protein and buffer pH were optimized according to pI values of the proteins. Also additives such as glycerol, arginine, glycine and detergent (DDM) were used if needed (in case of precipitation or aggregation of obtained proteins).

5.2.1.7. Cloning, expression and purification of recombinant AtWRKY50 and AtWRKY18^{DBD}

The coding sequence of the *Arabidopsis thaliana* wrky50 gene was obtained from a cDNA library (property of the Max Plank Institute for Plant Breeding Research, Cologne, Germany) as pDONOR201 vector (Gateway). For PCR amplification of the DNA fragment coding for the full lenght of AtWRKY50 protein, the primers complementary to each end of the ORF with overhangs compatible with cloning vector pMCSG48 were synthesized (Genomed, Poland) as follows:

forward 5'-TACTTCCAATCCAATGCCATGAATGATGCAGACACAAACTTGGGGA-3' and reverse 5'-TTATCCACTTCCAATGTTATTAGTTCATGCTTGAGTGATTGTGGGAA-3'.

For PCR amplification of the DNA fragment coding the DNA binding domain from AtWRKY18 protein, the primers complementary to each end of the DBD coding sequence with overhangs compatible with cloning vector pMCSG7 were synthesized (Genomed, Poland) as follows:

forward 5'-TACTTCCAATCCAATGCCTCGACTGTCTACGTGCCTACTGAAA-3' and reverse 5'-TTATCCACTTCCAATGTTATCAAGCATTTGGACCCAAGTGGTTATG-3'.

The AtWRKY50 and AtWRKY18^{DBD} coding sequences were PCR amplified using pDONR plasmid or pETD151/TOPO-WRKY18 as templates respectively. The reaction product of AtWRKY50 amplification was cloned then into the pMCSG48 expression vector (Midwest Center for Structural Genomics, Argonne, IL, USA) containing N-terminal His_6 and NusA fusion tags. The product of AtWRKY18^{DBD} amplification was cloned then into the pMCSG7 expression vector (Midwest Center for Structural Genomics, Argonne, IL, USA) containing N-terminal His_6 without additional fusion. The pMCSG48–AtWRKY50 and pMCSG7-AtWRKY18^{DBD} plasmid constructs were obtained by ligase-independent cloning [98]. The resulting recombinant plasmids were verified by DNA sequencing. The constructs were used to transform the BL21 Magic strain of *Escherichia coli*. Each culture, for AtWRKY50 or AtWRKY^{DBD} was carried on as follows: 25 ml LB medium containing 25 μg ml^{-1} kanamycin and 100 μg ml^{-1} carbenicilin was inoculated with a single colony and grown overnight at 37 °C. The overnight culture was used for inoculation of 1 l LB medium supplemented with appropriate antibiotics and grown to an OD_{600} of 0.8. The temperature was decreased to 18 °C and protein expression was induced by addition of isopropyl thio-β-D-galactoside at a final concentration of 0.5 mM. The cells were harvested 18 h after induction.

For AtWRKY50 and AtWRKY18^{DBD} purification procedure was identical.

The cell pellet was resuspended in binding buffer (20 mM imidazole, 500 mM NaCl, 50 mM Tris–HCl pH 7.5, 1 mM tris(2-carboxyethyl) phosphine) containing 200 μg/ml lysozyme. Cells were disrupted by sonication on ice using 4-min bursts with appropriate intervals for cooling. After sonication, 1 μl Benzonase (Sigma) was added to get rid of DNA. To remove cell debris, lysate was centrifuged at 17 000 rev min^{-1} for 30 min at 4 °C. The supernatant was loaded onto a column packed with 7 ml of Ni-Sepharose HP resin (GE Healthcare, Pittsburgh, PA, USA), connected to VacMan (Promega, Madison, WI, USA) and the chromatographic process was accelerated with a vacuum pump and the column was washed five times with 30 ml of binding buffer to remove non-specifically bound proteins. The protein of interest was eluted with buffer containing 300 mM imidazole in 500 mM NaCl, 50 mM Tris–HCl pH 7.5 and 1 mM tris(2-carboxyethyl) phosphine. The His_6 or His_6–NusA tag was cleaved with TEV (tobacco etch virus) protease overnight at 4 °C and the excess of imidazole was removed by dialysis simultaneously. The solution containing cleaved protein was mixed with Ni-Sepharose

HP resin to bind the tags and the His6- tagged TEV protease. The flow-through was collected and concentrated to 5 ml. The sample was applied to a HiLoad Superdex 200 16/60 gel filtration column(GE Healthcare) pre-equilibrated with buffer composed of 50 mM Tris/HCl, pH 8.0, 50 mM NaCl and 1 mM tris(2-carboxyethyl)phosphine. The peak fractions corresponded to a molecular mass of 20 kDa (AtWRKY50) or ~9 kDa ($AtWRKY18^{DBD}$) were analyzed on SDS-PAGE and pure fractions were mixed together. The protein sample was concentrated using Amicon Ultra 10 filters (Millipore) to 10 mg ml^{-1}. Pure protein samples were flash-frozen in liquid nitrogen as 100 µl aliquots and stored at -80 °C. For further analysis the samples were thawed, dialysed or diluted, if needed.

5.2.2. Crystallization of AtWRKY50

5.2.2.1. Crystallization of ligand free AtWRKY50

Prior setting up the crystallization screens, the protein concentration was adjusted to desired value and the protein solution was passed through an Ultrafree-MC Centrifugal Filter Unit (Millipore) with 0.1 µm pore size at 4°C. Protein concentration was determined spectrophotometrically at 280 nm or by the Bradford method (Bradford,1976) with BSA as a standard. The sitting-drop vapor-diffusion screening for initial crystallization conditions was performed using high-throughput Robotic Sitting Drop Vapor Diffusion setup (Mosquito). JCSG, PACT premier, PGA, Morpheus, Midas, Proplex crystallization screens (Molecular Dimensions) were used for the experiments. 0.4µl protein samples were mixed with an equal amount of the reservoir solution and equilibrated against 100 µl reservoir solution, and the crystallization plates were stored at 19°C.

The initial screening was performed also manually using hanging drop method and Structure Screen I and II (Molecular Dimensions) and Crystal Screen I and II (Hampton Research). 1µl protein samples were mixed with an equal amount of the reservoir solution and equilibrated against 500 µl reservoir solution, and the crystallization plates were stored at 19°C. The initial screening gave the following hits/crystallization conditions: 100 mM HEPES, pH 7.5, 800 mM potassium sodium tartrate tetrahydrate and 100 mM imidazole/MES pH 6.5, 30 mM $MgCl_2$, 30 mM $CaCl_2$, 20% PEG 550 MME, 10% PEG 20K. The initial screening was followed by manual optimization in hanging drops. The drop volume, pH of buffer and the concentration of reagents were modified. The crystallization trials were performed also in 4,

19 and 28°C. Crystals grew within 4 days at 19°C. After 2 months they were harvested with 0.1 mm nylon loops (Hampton Research), washed with cryo-protectant solution containing 20% (v/v) glycerol in the reservoir cocktail, and vitrified in liquid nitrogen for synchrotron-radiation data collection.

5.2.2.2. Co-crystallization with DNA

Oligonucleotides, (15 bases long): forward 5'-CGCCT**TGAC**CAGCGC-3' and reverse 5'-GCGCTG**GTCA**AGGCG-3' (Genomed, Poland) were annealed in buffer containing 25 mM Tris/HCl, pH 8.0, 50 mM NaCl, 20 mM $MgCl_2$ and 2 mM tris(2-carboxyethyl) phosphine. Both oligonucleotides were mixed at equimolar concentration and heated at 95 °C for 3min. They were allowed then to cool slowly to room temperature and transferred on ice in order to get double stranded DNA. The protein sample at 10 mg/ml was mixed with double stranded DNA in 1:1.2 molar ratio and incubated 1 h at 4°C. Prior to setting up the crystallization screens, the sample was centrifuged at 4°C.

The sitting-drop vapor-diffusion screening for initial crystallization conditions was performed using high-throughput Robotic Sitting Drop Vapor Diffusion setup (Mosquito). Structure Screen I+II, JCSG, PACT premier, PGA, Morpheus, Midas, Proplex crystallization screens (Molecular Dimensions) were used for these experiments. Additionally Natrix (Hampton Research) dedicated for DNA crystallization was used. 0.3µl or 0.6µl protein samples were mixed with 0.3 µl of the reservoir solution and equilibrated against 100 µl reservoir solution, and the crystallization plates were stored at 19°C.

5.2.2.3. Protein modifications to improve crystallization

5.2.2.3.1. Reductive lysine methylation

The reductive methylation is a chemical modification of free amino groups in which primary amines (i.e. lysine residues and the N-terminus) are modified to tertiary amines. The lysines residues exposed on the surface of protein are usually disordered and increase the surface entropy. Methylation of lysines offers opportunity to change the surface properties of protein and potentially its crystallization ability as well.

The methylation reaction was performed in 50 mM HEPES, pH 7.5, 200 mM NaCl, 1 mM TCEP at protein concentrations of 1 mg/ml or less. 20 µl freshly prepared 1 M dimethylamine-borane complex (ABC; Fluka product 15584) and 40 µl 1 M formaldehyde (made from 37%

stock; Fluka product 33220) were added per 1 ml protein solution, and the protein sample was gently mixed and incubated at 4°C for 2 hr. A further 20 µl ABC and 40 µl formaldehyde were added and the incubation was continued for 2 hr. Following the final addition of 10 µl ABC per 1 ml of protein solution, the reaction was incubated overnight at 4°C. Next day the reaction was stopped by addition of 10 ml of buffer containing 50 mM Tris pH 7.5, 200 mM NaCl, 1 mM TCEP and the sample was centrifuged to remove precipitated protein. Soluble methylated AtWRKY50 was concentrated to 5 ml and the purification by size-exclusion chromatography on S200 Superdex 16/60 Äkta express FPLC system pre-equilibrated with 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM TCEP was performed. The fractions were analysed using SDS-PAGE and the appropriate peak fractions of pure protein were pooled, concentrated and crystallization experiments was set up immediately. Crystallization experiment was performed using commercial screens: Morpheus, JCSG, PACT, Structure Screen I+II (Molecular Dimensions).

5.2.2.3.2. Limited proteolysis.

Protein digestion with certain protease can improve ability to crystallization. Proteases allowed getting rid of flexible fragments, linkers or loops in protein of interest. A fragment or domain may crystallize more readily or form better diffracting crystals than the full length protein. The crystallization may be performed immediately after digestion or if recognition of cleavage site is possible, the good approach is to prepare new plasmid with truncated protein form. There are commercially available two kits Proti-Ace and Proti-Ace 2 (Hampton Research) containing sets of 6 unique proteases each.

Proti-Ace	**Proti-Ace 2**
1 mg/ml α-Chymotrypsin,	1 mg/ml Proteinase K,
1 mg/ml Trypsin,	1 mg/ml Clostripain (Endoproteinase-Arg-C),
1 mg/ml Elastase,	1 mg/ml Pepsin,
1 mg/ml Papain,	1 mg/ml Thermolysin,
1 mg/ml Subtilisin,	1 mg/ml Bromelain,
1 mg/ml Endoproteinase Glu-C	1 mg/ml Actinase,

Each protease was prepared according to the manual. The initial digestion was prepared with all 12 enzymes in small scale. Based on the proteolytic pattern of each enzyme visualised by

SDS-PAGE, 2 proteases were choosen for crystallisation trials in large scale. After analysis of digestion patterns, two enzymes: elastase and papaine was choosen. The digestion with elastase and papaine was performed in large scale using 10 µl of enzyme (1 mg/ml) for each 90 µl of protein (10 mg/ml) as recommended in manual. The protein was incubated with enzyme for 1 h in room temperature and then set of crystallization using commercial screens: Morpheus, JCSG, PACT, Structure Screen I+II (Molecular Dimensions) were performed.

5.2.2.4. Crystallization of AtWRKY18^{DBD}

Prior to crystallization trials, a homogenous solution of AtWRKY18^{DBD} protein was concentrated to approximately 10 mg/ml and kept in 50 mM Tris buffer, pH 7.5, 200 mM NaCl, 2mM TCEP. The initial screening was performed manually using hanging drop method and the initial crystallization conditions were based on variants of the known condition described in literature [51]. Next, several crystallization screening experiments were carried out. Screening included Crystal Screens I and II and PEG/ion screen from Hampton Research and also 6 screens from Molecular Dimensions (JCSG, PACT premier, PGA, Morpheus, Midas, Proplex). For screening, two methods were used: the sitting-drop vapor-diffusion and hanging-drop vapour diffusion.

The sitting-drop vapor-diffusion was performed using high-throughput Robotic Sitting Drop Vapor Diffusion setup (Mosquito) and 0.4µl protein samples (5-15 mg/ml) were mixed with an equal amount of the reservoir solution and equilibrated against 100 µl reservoir solution.

Hanging-drop vapour diffusion was set manually 1-4 µl protein samples were mixed with 1-4 µl of the reservoir solution and equilibrated against 500-1000 µl reservoir solution. The crystallization plates were stored at 19°C.

5.2.3. Recombinant protein analyses

5.2.3.1. Protein concentration measurements

Protein concentration was measured in the NanoDrop spectrophotometer based on the UV absorption of Tyr, Trp, Phe residues and disulphide bonds in the measured sample. The absorbance at 280nm and an estimated extinction coefficient was used to calculate protein concentration. The extinction coefficient of the protein was estimated on the basis of the amino acid sequence using the ProtParam tool [62] on Expasy web page

(http://www.expasy.org). This method was convenient only for high-purity samples. In other case, the protein concentration was measured using Bradford method [15].

5.2.3.2. DNA preparation

To date, all studied plant WRKY transcription factors show high binding preference to the DNA sequence element, 5'-TTGACT-3', known as the W-box [58, 168]. In the binding experiment, we used synthesized oligonucleotide identical in sequence to the region of the parsley PR1-1 promoter containing one W-box [149].

For all experiments, DNA was prepared as follows. Oligonucleotides (15 bases) containing one W-box, the forward 5'-CGCCT**TGAC**CAGCGC-3' and reverse 5'-GCGCTG**GTCA**AGGCG-3' were synthesized (Genomed, Poland) and annealed before use. The conserved WRKY binding motif is underlined. Lyophilized oligonucleotides were resuspended in annealing buffer containing 25 mM Tris/HCl, pH 8.0, 50 mM NaCl, 20 mM $MgCl_2$ and 2 mM tris(2-carboxyethyl) phosphine. Both oligonucleotides were mixed at equimolar concentration and heated at 95°C for 3min. They were allowed then to cool slowly to room temperature and transferred on ice in order to get double stranded DNA.

5.2.3.3. Electrophoretic mobility shift assay (EMSA)

The electrophoretic mobility shift assay (EMSA) technique is based on the observation that protein:DNA complexes migrate more slowly than free linear DNA fragments during non-denaturing polyacrylamide or agarose gel electrophoresis [74, 141]. The classical EMSA protocol consist of few steps. First the protein is purified from the cells. Next step is synthesis and radiolabelling of the DNA probe with phosphorous 32 (^{32}P). Usually label is enzymatically incorporated to the 5′ ends of the DNA probe using ^{32}P-γATP as a substrate for T4 polynucleotide kinase. Purified proteins and radiolabelled DNA probes are incubated in particular binding buffer to promote binding of the proteins to the DNA probe. The DNA–protein complexes are loaded and separated on a non-denaturing polyacrylamide gel to separate the DNA–protein complexes from the free DNA probes. The polyacrylamide gels are then dried down and analyzed via autoradiography. Many labs have moved to alternative EMSA detection systems due to avoid radioactivity. Because dNTPs are available as modified with haptens (e.g., biotin and digoxigenin) or fluorescent dyes, there are numerous nonradioactive methods for performing EMSA. Fluorescent probes can be detected in-gel with

the aid of an appropriate imaging system, but this method has not been very popular to date because of expensive instruments required. Furthermore their sensitivity does not yet compare to radioactive probes.

Here, for EMSA experiment, the non-standard procedure with non-radioactive DNA probe was used. In this visualisation of DNA bands shift involves staining with toluidine blue dye. This method requires more DNA and protein than standard procedure with radioactively or fluorescent labelled oligonucleotides. Method choosen in this experiment is very cheap, easy to perform and do not require sophisticated equipment. This method unfortunately is not so sensitive and exclude quantitative analysis but very well visualize binding. For determination of dissociation constant of protein–DNA complex ITC method was used. For EMSA analysis the 30-nucleotide long DNA fragments:

forward 5'-TTATTCAGCCATCAAAAG**TTGACC**AATAAT-3'

and

reverse 5'-ATTATT**GGTCAA**CTTTTGATGGCTGAATAA-3'

corresponding to the W-box of the parsley PR-1 gene promoter [149] were used. The W-box element is highlighted.

To prepare double stranded DNA, equivalent amounts of the sense and antisense fragments of the respective oligonucleotides (Genomed, Poland) were annealed in 40 mM Tris–HCl pH 7.5, 20 mM $MgCl_2$, and 50 mM NaCl starting from 95°C and allowing to cool slowly to room temperature. Binding reactions were performed in buffer containing 50 mM Tris pH 8.0, 20 mM NaCl, 0,1% TritonX-100 and 1 mM $MgCl_2$. Reaction mixtures which included fixed amount of dsDNA (1.3 μg) and set of AtWRKY50 protein dilutions (0.27-10 μg) or $AtWRKY18^{DBD}$ dilutions (0.15-10 μg) were incubated at 25°C for 25 min. After addition of ficoll, samples were applied immediately on the gel. Gel electrophoresis was carried out at 4°C in running buffer (TB with 0.04% TritonX-100). Gels were stained in 1% tolonium chloride (toluidine blue) solution and destined in water.

5.2.3.4. Isothermal titration calorimetry (ITC)

In order to determine thermodynamic parameters resulting from the interaction of AtWRKY50 protein with the selected DNA, the physicochemical analysis using ITC (Isothermal Titration Calorimetry, Microcal) was performed. Isothermal Titration Calorimetry (ITC) is a technique used in quantitative determination of biomolecular

interactions. It involves direct measurements of heat that is released or absorbed during formation of biological complexes. ITC can simultaneously determine all binding parameters in a single experiment and the main advantage is that it measures the affinity of binding partners in their native states. Therefore any modifications of binding partners are not required. Measurements are held at a constant temperature previously selected for the compounds used in calibration test. Measuring heat transfer during binding allows accurate determination of the reaction parameters such as: enthalpy of the process (ΔH), entropy (ΔS), the dissociation constant of the protein-ligand complex (Kd), and the stoichiometry of the reaction (n) [135], binding constants (K_D), reaction stoichiometry (n) and enthalpy (ΔH). This method provides a complete thermodynamic profile of the molecular interaction.

Isothermal titration calorimetry experiments were performed at 20°C with the use of MicroCal iTC200 calorimeter (GE Healthcare). Changes in heat measured by ITC are highly sensitive to the composition of the mixtures tested, and therefore it is recommended that both the protein and the ligand used for titration were dissolved in the same buffer. The protein was dialyzed against 25 mM Tris pH 7.5 with 50 mM NaCl and 2 mM TCEP. DNA duplex solution was prepared in the dialysis buffer. The concentration of protein in measuring cell was determined by Bradford assay [2] and was 41 µM (or 88 µM) whereas the concentration of DNA duplex in syringe was 297 µM (or 550 µM) . DNA was injected in 2 µl aliquots. Raw ITC data were analyzed with Origin software to obtain following parameters: stoichiometry (N), dissociation constant K_d and changes in the enthaply and entropy during association. One set of sites model was fitted to data without first 7 experimental points, since during the titration of DNA duplex into the buffer alone the hyperbolically decreasing heat effect, likely related to DNA dilution, was observed. The experiment was repeated three times.

5.2.3.5. Secondary structure prediction

5.2.3.5.1 Circular dichroism

Circular dichroism (CD) is a versatile technique in structural biology, with wide range of applications. The most widely used CD applications is secondary structure determination of protein but in addition, it can be used to study protein interactions, structural changes, ligand

binding as well as folding properties. Circular Dichroism relies on the differential absorption of left (L) and right (R) circularly polarised radiation by chromophores which either possess intrinsic chirality or are placed in chiral environments. If, after passage through the sample being examined, the L and R components are not absorbed or are absorbed to equal extents, the recombination of L and R would regenerate radiation polarised in the original plane. However, if L and R form are absorbed to different extents, the resulting radiation would be said to possess elliptical polarisation. Proteins possess a number of chromophores which can give rise to CD signals. Different structural elements have characteristic CD spectra. In the far UV region (240-180 nm), which corresponds to peptide bond absorption, the CD spectrum of proteins can reveal important characteristics of their secondary structure and give the content of structural features such as α-helix and β-sheet. For example, α-helical proteins have negative bands at 222 and 208 nm and a positive band at 193 nm. Proteins with well-defined antiparallel β-pleated sheets (β-helices) have negative bands at 218 nm and positive bands at 195 nm, while disordered proteins have very low ellipticity above 210 nm and negative bands near 195 nm. The collagens are a unique class of proteins, which have three chains that wrap together in a triple helix. Each strand has a conformation resembling that of poly-L-proline in a extended helical conformation where all of the bonds are trans to each other (poly-L-proline II). Charged polypeptides, such as poly-L-glutamate or poly-L-lysine at neutral pH (originally thought to be in random coil conformation) have a similar extended poly-L-proline II-like conformation. The CD spectrum in the near UV region (320-260 nm) reflects the environments of the aromatic amino acid side chains and thus gives information about the tertiary structure of the protein. Because the spectra of proteins are dependent on their conformation, CD can be used to estimate the structure of unknown proteins and monitor conformational changes due to temperature, mutations, heat, denaturants or binding interactions. However, it does not give the residue-specific information that can be obtained by X-ray crystallography or NMR structural determinations, the method has the two major advantages: measurement can be made on small amounts of sample in physiological buffers and it allows monitoring any structural alterations that might result from changes in environmental conditions, such as pH, temperature and ionic strength.

The AtWRKY50 protein samples prior CD measurements were dialyzed against 10 mM phosphate buffer pH 7.5 containing 50 mM NaF. All the CD spectra were recorded on a

JASCO J-815 CD spectrometer equipped with a Peltier-thermostated cell holder. A 0.2 cm cell was used. Each CD spectrum was the average of 3 scans at continuous scanning mode, corrected by subtracting a spectrum of the buffer solution in the absence of protein at identical condition. Each scan in the range of 185-350 nm was obtained with scanning speed 50 nm min^{-1}, a 1 nm bandwidth, 0.5 nm data pitch and data integration time of 1 second. The data were processed using Savitsky-Golay smoothing window of 20 points. The spectra were analyzed using CDSSTR method accessible at DichroWeb server (http://www.cryst.bbk.ac.uk/cdweb/html/). CD data are presented in terms of ellipticity in millidegrees (mdeg) or mean residue ellipticity [Θ] in [deg x cm^2 x $dmol^{-1}$].

5.2.3.5.2. Intrinsically Disordered Protein Regions prediction

AtWRKY50 disordered regions were defined by a bioinformatics sequence analyses. Disordered and structured regions were predicted using the average score from an online server- Metadisorder [102]. This tool allows to calculate "consensus" from results returned by other methods. Metadisorder web service consists of four parts: MetaDisorder, MetaDisorder3D, MetaDisorderMD and MetaDisorderMD2. MetaDisorder builds the weighted consensus using 13 primary disorder methods: DisEMBL (3 versions), DISOPRED2, DISpro, Globplot, iPDA, IUPred (2 versions), Pdisorder, Poodle-s, Poodle-l, PrDOS, Spritz (2 versions), and RONN). Moreover this component was proved to be the best method during CASP8. To find similar sequences Metadisorder3d uses fold recognition methods such as: PSI-BLAST, FFAS, HHsearch, Phyre, Pcons, MGenThreader. The protein disorder is inferred using gaps in the alignments and the genetic algorithm. Metadisordermd merged two mentioned above method into one using genetic algorithm to optimize components integration. The last component of Metadisorder server-Metadisordermd2 is a variant of previous mentioned metadisordermd but for genetic algorithm optimization step a different scoring function was used. Predicted amino acid residues with an average prediction score ≥ 0.5 are designated disordered. A residue with an average prediction score ≤ 0.5 was considered structured.

Additional sequence analysis was performed usinge MoRFpred [46]. This tool predict occurrence of protein molecular recognition features [46]. Molecular recognition features (MoRFs, also known as molecular recognition elements, MoREs) are short binding regions located within longer intrinsically disordered regions that undergo disorder-to-order transitions

upon specific binding MoRFs are implicated in important processes including signaling and regulation. So far, only a limited number of experimentally validated MoRFs is known, which motivates development of computational methods that predict MoRFs from protein chains. Short (5–25 residues) binding regions are often located within longer intrinsically disordered regions. Long disordered binding regions (more than 30 residues) are typically conserved, so they often show up in databases derived from hidden Markov models such as Pfam or SMART. Example MoRFs collected from the Protein Data Bank (PDB) have been divided into three subtypes according to their structures in the bound state: alpha-MoRFs form alpha-helices, beta-MoRFs form beta-strands, and iota-MoRFs form irregular secondary structure.

6. Summary

The WRKY proteins are a large superfamily of transcription regulators of plant genes induced upon pathogen infection and during certain stages of plant development. Their hallmark is strong conservation of the DNA binding domain which contains an invariant WRKYGQK sequence and zinc binding motif. However, the overall sequences of individual representatives are highly divergent. So far there were only structural studies of DNA binding domain available.

The main goals of this thesis were structural studies of entire copies of the WRKY transcription factors from *Arabidipsis thaliana*. In presented studies, I developed an efficient method for expression and purification of recombinant AtWRKY50 and AtWRKY18^{DBD} protein. The obtained proteins retaining the biological activity of the DNA binding. The methods presented in this study allow the production of a significant amount of AtWRKY50 and AtWRKY18^{DBD} in bacterial expression system for further functional and structural studies. Obtained recombinant proteins were high quality to carry out crystallization experiments however all attempts to obtain well diffracting crystals of AtWRKY18^{DBD} or AtWRKY50 protein failed, thus solwing high-resilution crystallographic structure was impossible.

The CD spectrum and bioinformatics sequence analyses employed in this studies allowed to deduce that AtWRKY50 lack of well defined secondary structure and is partially disordered. This may explain difficulties in crystallization and failure to gain the main goal of the thesis - solving the crystallographic structure of the protein of interests.
ITC and EMSA analyses provided evidence for activity of recombinant AtWRKY50 protein and AtWRKY18^{DBD} toward DNA binding.

7. Streszczenie

Białka WRKY stanowią dużą rodzinę czynników transkrypcyjnych występujących wyłącznie u roślin. Białka WRKY regulują transkrypcję genów indukowanych podczas infekcji patogenami oraz genów związanych z niektórymi etapami rozwoju roślin. Ich cechą charakterystyczną jest obecność niezmiennej sekwencji WRKYGQK w obrębie domeny wiążącej DNA oraz unikatowy motyw palca cynkowego. Białka te wykazują duzą różnorodność sekwencji poza regionem domeny wiażącej DNA.
Celem niniejszej pracy były badania krystalograficzne czynników transkrypcyjnych WRKY pełnej długości, gdyż do tej pory w bazie danych PDB dostępne były tylko struktury domeny wiążącej DNA uzyskane metodą NMR oraz krystalografii rentgenowskiej.

W prezentowanych badaniach został opracowany skuteczny sposób ekspresji i oczyszczania rekombinowanych białek AtWRKY50 i AtWRKY18^{DBD}. Otrzymane białka zachowują aktywność biologiczną w testach wiązania DNA. Metody przedstawione w niniejszym opracowaniu umożliwiają otrzymanie znacznych ilości AtWRKY50 i AtWRKY18^{DBD} w bakteryjnym systemie ekspresyjnym do dalszych badań funkcjonalnych i strukturalnych. Otrzymane rekombinowane białka spełniały oczekiwane właściwości fizyko-chemiczne niezbędne do przeprowadzania krystalizacji. Niestety wszystkie próby uzyskania dyfrakcji kryształów AtWRKY18^{DBD} oraz białka AtWRKY50 były nieudane.

Widmo CD oraz analizy bioinformatyczne sekwencji wykonane dla białka AtWRKY50 pozwoliły wywnioskować, że nie posiada ono dobrze zdefiniowanej struktury trzeciorzędowej i jest częściowo nieuporządkowane. Może to wyjaśniać trudności w krystalizacji i nieudane próby realizacji głównego celu pracy - rozwiązanie struktury krystalicznej.
Wykonane analizy ITC i EMSA potwierdziły biologiczną aktywność rekombinowanych białek AtWRKY50 i AtWRKY18^{DBD} do wiązania DNA.

8. References

1. Alexandrova, K.S. and B.V. Conger, *Isolation of two somatic embryogenesis-related genes from orchardgrass (Dactylis glomerata).* Plant science, 2002(162): p. 301-307.
2. Andreasson, E., et al., *The MAP kinase substrate MKS1 is a regulator of plant defense responses.* EMBO J, 2005. **24**(14): p. 2579-89.
3. Arakawa, T. and K. Tsumoto, *The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation.* Biochem Biophys Res Commun, 2003. **304**(1): p. 148-52.
4. Arakawa, T., et al., *Biotechnology applications of amino acids in protein purification and formulations.* Amino Acids, 2007. **33**(4): p. 587-605.
5. Asad, S., et al., *Studies on the refolding process of recombinant horseradish peroxidase.* Mol Biotechnol, 2013. **54**(2): p. 484-92.
6. Asai, T., et al., *MAP kinase signalling cascade in Arabidopsis innate immunity.* Nature, 2002. **415**(6875): p. 977-83.
7. Babitha, K.C., et al., *Co-expression of AtbHLH17 and AtWRKY28 confers resistance to abiotic stress in Arabidopsis.* Transgenic Res, 2013. **22**(2): p. 327-41.
8. Bedouelle, H. and P. Duplay, *Production in Escherichia coli and one-step purification of bifunctional hybrid proteins which bind maltose. Export of the Klenow polymerase into the periplasmic space.* Eur J Biochem, 1988. **171**(3): p. 541-9.
9. Bera, S.K., B.C. Ajay, and A.L. Singh, *WRKY and Na+/H+ antiporter genes conferring tolerance to salinity in interspecific derivatives of peanut (Arachis hypogaea L.).* Australian Journal of Crop Science 2013. **7**(8): p. 1173-1180.
10. Bergfors, T., *Seeds to crystals.* J Struct Biol, 2003. **142**(1): p. 66-76.
11. Besseau, S., J. Li, and E.T. Palva, *WRKY54 and WRKY70 co-operate as negative regulators of leaf senescence in Arabidopsis thaliana.* J Exp Bot, 2012. **63**(7): p. 2667-79.
12. Bhattarai, K.K., et al., *WRKY72-type transcription factors contribute to basal immunity in tomato and Arabidopsis as well as gene-for-gene resistance mediated by the tomato R gene Mi-1.* Plant J, 2010. **63**(2): p. 229-40.
13. Birkenbihl, R.P., C. Diezel, and I.E. Somssich, *Arabidopsis WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward Botrytis cinerea infection.* Plant Physiol, 2012. **159**(1): p. 266-85.
14. Birnbaum, K., et al., *A gene expression map of the Arabidopsis root.* Science, 2003. **302**(5652): p. 1956-60.
15. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal Biochem, 1976. **72**: p. 248-54.
16. Brand, L.H., et al., *DPI-ELISA: a fast and versatile method to specify the binding of plant transcription factors to DNA in vitro.* Plant Methods, 2010. **6**: p. 25.
17. Brotman, Y., et al., *Trichoderma-plant root colonization: escaping early plant defense responses and activation of the antioxidant machinery for saline stress tolerance.* PLoS Pathog, 2013. **9**(3): p. e1003221.
18. Cabrita, L.D. and S.P. Bottomley, *Protein expression and refolding--a practical guide to getting the most out of inclusion bodies.* Biotechnol Annu Rev, 2004. **10**: p. 31-50.

19. Chang, I.F., et al., *Proteomic profiling of tandem affinity purified 14-3-3 protein complexes in Arabidopsis thaliana.* Proteomics, 2009. **9**(11): p. 2967-85.
20. Chen, C. and Z. Chen, *Potentiation of developmentally regulated plant defense response by AtWRKY18, a pathogen-induced Arabidopsis transcription factor.* Plant Physiol, 2002. **129**(2): p. 706-16.
21. Chen, H., et al., *Roles of arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress.* BMC Plant Biol, 2010. **10**: p. 281.
22. Chen, L., et al., *WRKY8 transcription factor functions in the TMV-cg defense response by mediating both abscisic acid and ethylene signaling in Arabidopsis.* Proc Natl Acad Sci U S A, 2013. **110**(21): p. E1963-71.
23. Chen, L., L. Zhang, and D. Yu, *Wounding-induced WRKY8 is involved in basal defense in Arabidopsis.* Mol Plant Microbe Interact, 2010. **23**(5): p. 558-65.
24. Chen, W., et al., *Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses.* Plant Cell, 2002. **14**(3): p. 559-74.
25. Chen, Y.F., et al., *The WRKY6 transcription factor modulates PHOSPHATE1 expression in response to low Pi stress in Arabidopsis.* Plant Cell, 2009. **21**(11): p. 3554-66.
26. Cheng, Y., et al., *Structural and functional analysis of VQ motif-containing proteins in Arabidopsis as interacting proteins of WRKY transcription factors.* Plant Physiol, 2012. **159**(2): p. 810-25.
27. Cheong, Y.H., et al., *Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis.* Plant Physiol, 2002. **129**(2): p. 661-77.
28. Chi, Y., et al., *Protein-protein interactions in the regulation of WRKY transcription factors.* Mol Plant, 2013. **6**(2): p. 287-300.
29. Chisholm, S.T., et al., *Host-microbe interactions: shaping the evolution of the plant immune response.* Cell, 2006. **124**(4): p. 803-14.
30. Chujo, T., et al., *OsWRKY28, a PAMP-responsive transrepressor, negatively regulates innate immune responses in rice against rice blast fungus.* Plant Mol Biol, 2013. **82**(1-2): p. 23-37.
31. Ciolkowski, I., et al., *Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function.* Plant Mol Biol, 2008. **68**(1-2): p. 81-92.
32. Consortium, A.I.M., *Evidence for network evolution in an Arabidopsis interactome map.* Science, 2011(333): p. 601-607.
33. Contento, A.L., S.J. Kim, and D.C. Bassham, *Transcriptome profiling of the response of Arabidopsis suspension culture cells to Suc starvation.* Plant Physiol, 2004. **135**(4): p. 2330-47.
34. Cooper, D.R., et al., *Protein crystallization by surface entropy reduction: optimization of the SER strategy.* Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 5): p. 636-45.
35. Dang, F.-F., et al., *CaWRKY40, a WRKY protein of pepper, plays an important role in the regulation of tolerance to heat stress and resistance to Ralstonia solanacearum infection.* Plant Cell Environment, 2013. **36**(4): p. 757-774.
36. Dangl, J.L., R.A. Dietrich, and M.H. Richberg, *Death Don't Have No Mercy: Cell Death Programs in Plant-Microbe Interactions.* Plant Cell, 1996. **8**(10): p. 1793-1807.

37. Das, D. and M.M. Georgiadis, *A directed approach to improving the solubility of Moloney murine leukemia virus reverse transcriptase.* Protein Sci, 2001. **10**(10): p. 1936-41.
38. Davis, G.D., et al., *New fusion protein systems designed to give soluble expression in Escherichia coli.* Biotechnol Bioeng, 1999. **65**(4): p. 382-8.
39. de Pater, S., et al., *Characterization of a zinc-dependent transcriptional activator from Arabidopsis.* Nucleic Acids Res, 1996. **24**(23): p. 4624-31.
40. Dellagi, A., et al., *A potato gene encoding a WRKY-like transcription factor is induced in interactions with Erwinia carotovora subsp. atroseptica and Phytophthora infestans and is coregulated with class I endochitinase expression.* Mol Plant Microbe Interact, 2000. **13**(10): p. 1092-101.
41. Deslandes, L., et al., *Resistance to Ralstonia solanacearum in Arabidopsis thaliana is conferred by the recessive RRS1-R gene, a member of a novel family of resistance genes.* Proc Natl Acad Sci U S A, 2002. **99**(4): p. 2404-9.
42. Devaiah, B.N., A.S. Karthikeyan, and K.G. Raghothama, *WRKY75 transcription factor is a modulator of phosphate acquisition and root development in Arabidopsis.* Plant Physiol, 2007. **143**(4): p. 1789-801.
43. di Guan, C., et al., *Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein.* Gene, 1988. **67**(1): p. 21-30.
44. Dieckman, L., et al., *High throughput methods for gene cloning and expression.* Protein Expr Purif, 2002. **25**(1): p. 1-7.
45. Ding, Z.J., et al., *Transcription factor WRKY46 regulates osmotic stress responses and stomatal movement independently in Arabidopsis.* Plant J, 2014. **79**(1): p. 13-27.
46. Disfani, F.M., et al., *MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins.* Bioinformatics, 2012. **28**(12): p. i75-83.
47. Dodds, P.N. and J.P. Rathjen, *Plant immunity: towards an integrated view of plant-pathogen interactions.* Nat Rev Genet, 2010. **11**(8): p. 539-48.
48. Dong, J., C. Chen, and Z. Chen, *Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response.* Plant Mol Biol, 2003. **51**(1): p. 21-37.
49. Dong, X.Y., Y. Huang, and Y. Sun, *Refolding kinetics of denatured-reduced lysozyme in the presence of folding aids.* J Biotechnol, 2004. **114**(1-2): p. 135-42.
50. Du, L. and Z. Chen, *Identification of genes encoding receptor-like protein kinases as possible targets of pathogen- and salicylic acid-induced WRKY DNA-binding proteins in Arabidopsis.* Plant J, 2000. **24**(6): p. 837-47.
51. Duan, M.R., et al., *DNA binding mechanism revealed by high resolution crystal structure of Arabidopsis thaliana WRKY1 protein.* Nucleic Acids Res, 2007. **35**(4): p. 1145-54.
52. Dunker, A.K., et al., *Intrinsically disordered protein.* J Mol Graph Model, 2001. **19**(1): p. 26-59.
53. Dyda, F., et al., *Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases.* Science, 1994. **266**(5193): p. 1981-6.
54. Ejima, D., et al., *High yield refolding and purification process for recombinant human interleukin-6 expressed in Escherichia coli.* Biotechnology and Bioengineering, 2000. **62**(3): p. 301-310.

55. Esnouf, R.M., et al., *Continuous and discontinuous changes in the unit cell of HIV-1 reverse transcriptase crystals on dehydration.* Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 5): p. 938-53.
56. Eulgem, T., et al., *The WRKY superfamily of plant transcription factors.* Trends Plant Sci, 2000. **5**(5): p. 199-206.
57. Eulgem, T., et al., *Early nuclear events in plant defence signalling: rapid gene activation by WRKY transcription factors.* EMBO J, 1999. **18**(17): p. 4689-99.
58. Eulgem, T. and I.E. Somssich, *Networks of WRKY transcription factors in defense signaling.* Curr Opin Plant Biol, 2007. **10**(4): p. 366-71.
59. Gadjev, I., et al., *Transcriptomic footprints disclose specificity of reactive oxygen species signaling in Arabidopsis.* Plant Physiol, 2006. **141**(2): p. 436-45.
60. Gao, Q.M., et al., *Low oleic acid-derived repression of jasmonic acid-inducible defense responses requires the WRKY50 and WRKY51 proteins.* Plant Physiol, 2011. **155**(1): p. 464-76.
61. Gao, X., et al., *Bifurcation of Arabidopsis NLR immune signaling via Ca(2)(+)-dependent protein kinases.* PLoS Pathog, 2013. **9**(1): p. e1003127.
62. Gasteiger, E., et al., The Proteomics Protocols Handbook, 2005: p. 571-607.
63. Giege, R.M., V., *Crystallogenesis of proteins.* Trends in Biotechnology, 1989. **7**(10): p. 277-282.
64. Goldschmidt, L., et al., *Toward rational protein crystallization: A Web server for the design of crystallizable protein variants.* Protein Sci, 2007. **16**(8): p. 1569-76.
65. Golovanov, A.P., et al., *A simple method for improving protein solubility and long-term stability.* J Am Chem Soc, 2004. **126**(29): p. 8933-9.
66. Grunewald, W., et al., *Tightly controlled WRKY23 expression mediates Arabidopsis embryo development.* EMBO Rep, 2013. **14**(12): p. 1136-42.
67. Grunewald, W., et al., *A role for AtWRKY23 in feeding site establishment of plant-parasitic nematodes.* Plant Physiol, 2008. **148**(1): p. 358-68.
68. Guillaumie, S., et al., *The grapevine transcription factor WRKY2 influences the lignin pathway and xylem development in tobacco.* Plant Mol Biol, 2010. **72**(1-2): p. 215-34.
69. Hammargren, J., et al., *A novel connection between nucleotide and carbohydrate metabolism in mitochondria: sugar regulation of the Arabidopsis nucleoside diphosphate kinase 3a gene.* Plant Cell Rep, 2008. **27**(3): p. 529-34.
70. Hammarstrom, M., et al., *Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli.* Protein Sci, 2002. **11**(2): p. 313-21.
71. Han, K.G., S.S. Lee, and C. Kang, *Soluble expression of cloned phage K11 RNA polymerase gene in Escherichia coli at a low temperature.* Protein Expr Purif, 1999. **16**(1): p. 103-8.
72. Hara, K., et al., *Rapid systemic accumulation of transcripts encoding a tobacco WRKY transcription factor upon wounding.* Mol Gen Genet, 2000. **263**(1): p. 30-7.
73. Hartley, D.L. and J.F. Kane, *Properties of inclusion bodies from recombinant Escherichia coli.* Biochem Soc Trans, 1988. **16**(2): p. 101-2.
74. Hendrickson, W., *Protein-DNA interactions studied by the gel electrophoresis-DNA binding assay.* Biotechniques, 1985. **3**: p. 198-207.
75. Higashi, K., et al., *Modulation of defense signal transduction by flagellin-induced WRKY41 transcription factor in Arabidopsis thaliana.* Mol Genet Genomics, 2008. **279**(3): p. 303-12.

76. Hinderhofer, K. and U. Zentgraf, *Identification of a transcription factor specifically expressed at the onset of leaf senescence.* Planta, 2001. **213**(3): p. 469-73.
77. Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R., and Stüber, D, *Genetic approach to facilitate purifaction of recombinant proteins with a novel metal chelate adsorbent. .* Biotechnology (N Y), 1988. **6**: p. 1321-1325.
78. http://www.arabidopsis.org/browse/genefamily/WRKY.jsp.
79. Hu, Y., et al., *Arabidopsis transcription factor WRKY8 functions antagonistically with its interacting partner VQ9 to modulate salinity stress tolerance.* Plant J, 2013. **74**(5): p. 730-45.
80. Hu, Y., Q. Dong, and D. Yu, *Arabidopsis WRKY46 coordinates with WRKY70 and WRKY53 in basal resistance against pathogen Pseudomonas syringae.* Plant Sci, 2012. **185-186**: p. 288-97.
81. Inoue, H., et al., *Blast resistance of CC-NB-LRR protein Pb1 is mediated by WRKY45 through protein-protein interaction.* Proc Natl Acad Sci U S A, 2013. **110**(23): p. 9577-82.
82. Ishiguro, S. and K. Nakamura, *Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and beta-amylase from sweet potato.* Mol Gen Genet, 1994. **244**(6): p. 563-71.
83. Ishihama, N. and H. Yoshioka, *Post-translational regulation of WRKY transcription factors in plant immunity.* Curr Opin Plant Biol, 2012. **15**(4): p. 431-7.
84. Ito, L., K. Shiraki, and H. Yamaguchi, *Comparative analysis of amino acids and amino-acid derivatives in protein crystallization.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2010. **66**(Pt 6): p. 744-9.
85. Izaguirre, M.M., et al., *Convergent responses to stress. Solar ultraviolet-B radiation and Manduca sexta herbivory elicit overlapping transcriptional responses in field-grown plants of Nicotiana longiflora.* Plant Physiol, 2003. **132**(4): p. 1755-67.
86. Jenkins, T.M., et al., *Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues.* Proc Natl Acad Sci U S A, 1995. **92**(13): p. 6057-61.
87. Jiang, W. and D. Yu, *Arabidopsis WRKY2 transcription factor may be involved in osmotic stress response.* Acta Botanica Yunnanica, 2009. **31**: p. 427-432.
88. Jiang, W. and D. Yu, *Arabidopsis WRKY2 transcription factor mediates seed germination and postgermination arrest of development by abscisic acid.* BMC Plant Biol, 2009. **9**: p. 96.
89. Jiang, Y. and M.K. Deyholos, *Comprehensive transcriptional profiling of NaCl-stressed Arabidopsis roots reveals novel classes of responsive genes.* BMC Plant Biol, 2006. **6**: p. 25.
90. Jiang, Y. and M.K. Deyholos, *Functional characterization of Arabidopsis NaCl-inducible WRKY25 and WRKY33 transcription factors in abiotic stresses.* Plant Mol Biol, 2009. **69**(1-2): p. 91-105.
91. Jin, J., et al., *PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors.* Nucleic Acids Res, 2014. **42**(Database issue): p. D1182-7.
92. Johnson, C.S., B. Kolevski, and D.R. Smyth, *TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor.* Plant Cell, 2002. **14**(6): p. 1359-75.

93. Journot-Catalino, N., et al., *The transcription factors WRKY11 and WRKY17 act as negative regulators of basal resistance in Arabidopsis thaliana.* Plant Cell, 2006. **18**(11): p. 3289-302.
94. Kalde, M., et al., *Members of the Arabidopsis WRKY group III transcription factors are part of different plant defense signaling pathways.* Mol Plant Microbe Interact, 2003. **16**(4): p. 295-305.
95. Kasajima, I., et al., *WRKY6 is involved in the response to boron deficiency in Arabidopsis thaliana.* Physiol Plant, 2010. **139**(1): p. 80-92.
96. Kim, K.C., B. Fan, and Z. Chen, *Pathogen-induced Arabidopsis WRKY7 is a transcriptional repressor and enhances plant susceptibility to Pseudomonas syringae.* Plant Physiol, 2006. **142**(3): p. 1180-92.
97. Kim, K.C., et al., *Arabidopsis WRKY38 and WRKY62 transcription factors interact with histone deacetylase 19 in basal defense.* Plant Cell, 2008. **20**(9): p. 2357-71.
98. Kim, Y., et al., *High-throughput protein purification and quality assessment for crystallization.* Methods, 2011. **55**(1): p. 12-28.
99. Knoth, C., et al., *Arabidopsis WRKY70 is required for full RPP4-mediated disease resistance and basal defense against Hyaloperonospora parasitica.* Mol Plant Microbe Interact, 2007. **20**(2): p. 120-8.
100. Konig, P. and T.J. Richmond, *The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility.* J Mol Biol, 1993. **233**(1): p. 139-54.
101. Kotta-Loizou, I., G.N. Tsaousis, and S.J. Hamodrakas, *Analysis of Molecular Recognition Features (MoRFs) in membrane proteins.* Biochim Biophys Acta, 2013. **1834**(4): p. 798-807.
102. Kozlowski, L.P. and J.M. Bujnicki, *MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins.* BMC Bioinformatics, 2012. **13**: p. 111.
103. Kriminski, S., et al., *Flash-cooling and annealing of protein crystals.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 3): p. 459-71.
104. Kryshtafovych, A., et al., *Protein structure prediction center in CASP8.* Proteins, 2009. **77 Suppl 9**: p. 5-9.
105. Lagace, M. and D.P. Matton, *Characterization of a WRKY transcription factor expressed in late torpedo-stage embryos of Solanum chacoense.* Planta, 2004. **219**(1): p. 185-9.
106. Lai, Z., et al., *Arabidopsis sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense.* Plant Cell, 2011. **23**(10): p. 3824-41.
107. Lai, Z., et al., *Roles of Arabidopsis WRKY3 and WRKY4 transcription factors in plant responses to pathogens.* BMC Plant Biol, 2008. **8**: p. 68.
108. LaVallie, E.R., et al., *A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm.* Biotechnology (N Y), 1993. **11**(2): p. 187-93.
109. Leandro, P., et al., *Glycerol increases the yield and activity of human phenylalanine hydroxylase mutant enzymes produced in a prokaryotic expression system.* Mol Genet Metab, 2001. **73**(2): p. 173-8.
110. Lechtken, A., et al., *Overexpression, refolding, and purification of polyhistidine-tagged human retinoic acid related orphan receptor RORalpha4.* Protein Expr Purif, 2006. **49**(1): p. 114-20.

111. Li, J., et al., *WRKY70 modulates the selection of signaling pathways in plant defense.* Plant J, 2006. **46**(3): p. 477-91.
112. Li, S., et al., *Arabidopsis thaliana WRKY25, WRKY26, and WRKY33 coordinate induction of plant thermotolerance.* Planta, 2011. **233**(6): p. 1237-52.
113. Li, S., et al., *Functional analysis of an Arabidopsis transcription factor WRKY25 in heat stress.* Plant Cell Rep, 2009. **28**(4): p. 683-93.
114. Li, S., et al., *Functional characterization of Arabidopsis thaliana WRKY39 in heat stress.* Mol Cells, 2010. **29**(5): p. 475-83.
115. Li, W., et al., *Bacterial expression, refolding, functional characterization, and mass spectrometric identification of full-length human PPAR-γ* Bioscience, Biotechnology and Biochemistry, 2010. **74**(6): p. 1173-1180.
116. Liu, J.J. and A.K. Ekramoddoullah, *Identification and characterization of the WRKY transcription factor family in Pinus monticola.* Genome, 2009. **52**(1): p. 77-88.
117. Liu, Y.D., et al., *A newly proposed mechanism for arginine-assisted protein refolding--not inhibiting soluble oligomers although promoting a correct structure.* Protein Expr Purif, 2007. **51**(2): p. 235-42.
118. Liu, Z.Q., et al., *Cooperation of three WRKY-domain transcription factors WRKY18, WRKY40, and WRKY60 in repressing two ABA-responsive genes ABI4 and ABI5 in Arabidopsis.* J Exp Bot, 2012. **63**(18): p. 6371-92.
119. Lu, H., et al., *Purification, refolding of hybrid hIFNgamma-kringle 5 expressed in Escherichia coli.* Curr Microbiol, 2001. **42**(3): p. 211-6.
120. Luo, M., et al., *MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis.* Proc Natl Acad Sci U S A, 2005. **102**(48): p. 17531-6.
121. Ly, K., L. O'Ryan, and A.K. Mitra, *Overexpression, purification and biophysical characterisation of E. coli MerT.* Protein Expr Purif, 2015. **108**: p. 85-9.
122. Maeo, K., et al., *Role of conserved residues of the WRKY domain in the DNA-binding of tobacco WRKY family proteins.* Biosci Biotechnol Biochem, 2001. **65**(11): p. 2428-36.
123. Maleck, K., et al., *The transcriptome of Arabidopsis thaliana during systemic acquired resistance.* Nat Genet, 2000. **26**(4): p. 403-10.
124. Mao, G., et al., *Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in Arabidopsis.* Plant Cell, 2011. **23**(4): p. 1639-53.
125. Maxwell, K.L., et al., *A simple in vivo assay for increased protein solubility.* Protein Sci, 1999. **8**(9): p. 1908-11.
126. McPherson, A. and B. Cudney, *Searching for silver bullets: an alternative strategy for crystallizing macromolecules.* J Struct Biol, 2006. **156**(3): p. 387-406.
127. Meier, S., et al., *Co-expression and promoter content analyses assign a role in biotic and abiotic stress responses to plant natriuretic peptides.* BMC Plant Biol, 2008. **8**: p. 24.
128. Mishina, T.E. and J. Zeier, *Pathogen-associated molecular pattern recognition rather than development of tissue necrosis contributes to bacterial induction of systemic acquired resistance in Arabidopsis.* Plant J, 2007. **50**(3): p. 500-13.
129. Mishra, S., et al., *Wound induced tanscriptional regulation of benzylisoquinoline pathway and characterization of wound inducible PsWRKY transcription factor from Papaver somniferum.* PLoS One, 2013. **8**(1): p. e52784.

130. Mollania, N., et al., *An efficient in vitro refolding of recombinant bacterial laccase in Escherichia coli.* Enzyme Microb Technol, 2013. **52**(6-7): p. 325-30.
131. Moult, J., et al., *Critical assessment of methods of protein structure prediction (CASP)--round x.* Proteins, 2014. **82 Suppl 2**: p. 1-6.
132. Nardini, M., et al., *The C-terminal domain of the transcriptional corepressor CtBP is intrinsically unstructured.* Protein Sci, 2006. **15**(5): p. 1042-50.
133. Park, C.Y., et al., *WRKY group IId transcription factors interact with calmodulin.* FEBS Lett, 2005. **579**(6): p. 1545-50.
134. Peroutka Iii, R.J., et al., *SUMO fusion technology for enhanced protein expression and purification in prokaryotes and eukaryotes.* Methods Mol Biol, 2011. **705**: p. 15-30.
135. Perozzo, R., G. Folkers, and L. Scapozza, *Thermodynamics of protein-ligand interactions: history, presence, and future aspects.* J Recept Signal Transduct Res, 2004. **24**(1-2): p. 1-52.
136. Pnueli, L., et al., *Molecular and biochemical mechanisms associated with dormancy and drought tolerance in the desert legume Retama raetam.* Plant J, 2002. **31**(3): p. 319-30.
137. Prive, G.G., *Detergents for the stabilization and crystallization of membrane proteins.* Methods, 2007. **41**(4): p. 388-97.
138. Qiu, J.L., et al., *Arabidopsis MAP kinase 4 regulates gene expression through transcription factor release in the nucleus.* EMBO J, 2008. **27**(16): p. 2214-21.
139. Rariy, R.V. and A.M. Klibanov, *Correct protein folding in glycerol.* Proc Natl Acad Sci U S A, 1997. **94**(25): p. 13520-3.
140. Ren, X., et al., *ABO3, a WRKY transcription factor, mediates plant responses to abscisic acid and drought tolerance in Arabidopsis.* Plant J, 2010. **63**(3): p. 417-29.
141. Revzin, A., *Gel electrophoresis assays for DNA-protein interactions.* Biotechniques, 1989. **7**(4): p. 346-55.
142. Rizhsky, L., et al., *The zinc finger protein Zat12 is required for cytosolic ascorbate peroxidase 1 expression during oxidative stress in Arabidopsis.* J Biol Chem, 2004. **279**(12): p. 11736-43.
143. Rizhsky, L., H. Liang, and R. Mittler, *The combined effect of drought stress and heat shock on gene expression in tobacco.* Plant Physiol, 2002. **130**(3): p. 1143-51.
144. Robatzek, S. and I.E. Somssich, *A new member of the Arabidopsis WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defence-related processes.* Plant J, 2001. **28**(2): p. 123-33.
145. Robatzek, S. and I.E. Somssich, *Targets of AtWRKY6 regulation during plant senescence and pathogen defense.* Genes Dev, 2002. **16**(9): p. 1139-49.
146. Romero, P., et al., *Sequence complexity of disordered protein.* Proteins, 2001. **42**(1): p. 38-48.
147. Rushton, P.J., et al., *Members of a new family of DNA-binding proteins bind to a conserved cis-element in the promoters of alpha-Amy2 genes.* Plant Mol Biol, 1995. **29**(4): p. 691-702.
148. Rushton, P.J., et al., *WRKY transcription factors.* Trends Plant Sci, 2010. **15**(5): p. 247-58.
149. Rushton, P.J., et al., *Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes.* EMBO J, 1996. **15**(20): p. 5690-700.

150. Sauter, C., Ng, J. D., Lorber, B., Keith, G., Brion, P., Hosseini, M. W., Lehn, J. M., Giegé, R. , *Additives for the crystallization of proteins and nucleic acids.* J. Cryst. Growth, 1999. **196**: p. 365-376.
151. Scarpeci, T.E., et al., *Overexpression of AtWRKY30 enhances abiotic stress tolerance during early growth stages in Arabidopsis thaliana.* Plant Mol Biol, 2013. **83**(3): p. 265-77.
152. Schon, M., et al., *Analyses of wrky18 wrky40 plants reveal critical roles of SA/EDS1 signaling and indole-glucosinolate biosynthesis for Golovinomyces orontii resistance and a loss-of resistance towards Pseudomonas syringae pv. tomato AvrRPS4.* Mol Plant Microbe Interact, 2013. **26**(7): p. 758-67.
153. Schoonheim, P.J., et al., *14-3-3 adaptor proteins are intermediates in ABA signal transduction during barley seed germination.* Plant J, 2007. **49**(2): p. 289-301.
154. Shang, Y., et al., *The Mg-chelatase H subunit of Arabidopsis antagonizes a group of WRKY transcription repressors to relieve ABA-responsive genes of inhibition.* Plant Cell, 2010. **22**(6): p. 1909-35.
155. Shen, Q.H., et al., *Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses.* Science, 2007. **315**(5815): p. 1098-103.
156. Shim, J.S. and Y.D. Choi, *Direct regulation of WRKY70 by AtMYB44 in plant defense responses.* Plant Signal Behav, 2013. **8**(6): p. e20783.
157. Shim, J.S., et al., *AtMYB44 regulates WRKY70 expression and modulates antagonistic interaction between salicylic acid and jasmonic acid signaling.* Plant J, 2013. **73**(3): p. 483-95.
158. Singh, S.M. and A.K. Panda, *Solubilization and refolding of bacterial inclusion body proteins.* J Biosci Bioeng, 2005. **99**(4): p. 303-10.
159. Skibbe, M., et al., *Induced plant defenses in the natural environment: Nicotiana attenuata WRKY3 and WRKY6 coordinate responses to herbivory.* Plant Cell, 2008. **20**(7): p. 1984-2000.
160. Smith, D.B. and K.S. Johnson, *Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.* Gene, 1988. **67**(1): p. 31-40.
161. Smith, M.C., Furman, T.C., Ingolia, T.D., Pidgeon, C. , *Chelating peptide-immobilized metal ion affinity chromatography. A new concept inaffinity chromatography for recombinant proteins.* J. Biol. Chem, 1988. **263**: p. 7211-7215.
162. Sun, C., et al., *A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter.* Plant Cell, 2003. **15**(9): p. 2076-92.
163. Tahirov, T.H., et al., *High-resolution crystals of methionine aminopeptidase from Pyrococcus furiosus obtained by water-mediated transformation.* J Struct Biol, 1998. **121**(1): p. 68-72.
164. Tiwari, A. and R. Bhat, *Stabilization of yeast hexokinase A by polyol osmolytes: correlation with the physicochemical properties of aqueous solutions.* Biophys Chem, 2006. **124**(2): p. 90-9.
165. Tsuda, K., et al., *Network properties of robust immunity in plants.* PLoS Genet, 2009. **5**(12): p. e1000772.
166. Tsumoto, K., et al., *Practical considerations in refolding proteins from inclusion bodies.* Protein Expr Purif, 2003. **28**(1): p. 1-8.
167. Tsumoto, K., et al., *Role of arginine in protein refolding, solubilization, and purification.* Biotechnol Prog, 2004. **20**(5): p. 1301-8.

168. Ulker, B. and I.E. Somssich, *WRKY transcription factors: from DNA binding towards biological function.* Curr Opin Plant Biol, 2004. **7**(5): p. 491-8.
169. Vagenende, V., M.G. Yap, and B.L. Trout, *Mechanisms of protein stabilization and prevention of protein aggregation by glycerol.* Biochemistry, 2009. **48**(46): p. 11084-96.
170. van Loon, L.C., M. Rep, and C.M. Pieterse, *Significance of inducible defense-related proteins in infected plants.* Annu Rev Phytopathol, 2006. **44**: p. 135-62.
171. Van Loon, L.C., et al., *Recommendations for naming plant pathogenesis-related proteins.* . Plant Molecular Biology Reporter, 1994(12): p. 245-264.
172. van Verk, M.C., et al., *A Novel WRKY transcription factor is required for induction of PR-1a gene expression by salicylic acid and bacterial elicitors.* Plant Physiol, 2008. **146**(4): p. 1983-95.
173. Vandenabeele, S., et al., *A comprehensive analysis of hydrogen peroxide-induced gene expression in tobacco.* Proc Natl Acad Sci U S A, 2003. **100**(26): p. 16113-8.
174. Vucetic, S., et al., *Flavors of protein disorder.* Proteins, 2003. **52**(4): p. 573-84.
175. Wan, J., et al., *A LysM receptor-like kinase plays a critical role in chitin signaling and fungal resistance in Arabidopsis.* Plant Cell, 2008. **20**(2): p. 471-81.
176. Wang, D., N. Amornsiripanitch, and X. Dong, *A genomic approach to identify regulatory nodes in the transcriptional network of systemic acquired resistance in plants.* PLoS Pathog, 2006. **2**(11): p. e123.
177. Wang, H., et al., *Arabidopsis WRKY45 transcription factor activates PHOSPHATE TRANSPORTER1;1 expression in response to phosphate starvation.* Plant Physiol, 2014. **164**(4): p. 2020-9.
178. Wang, H.J., et al., *Transcriptomic adaptations in rice suspension cells under sucrose starvation.* Plant Mol Biol, 2007. **63**(4): p. 441-63.
179. Wang, X., et al., *Arabidopsis transcription factor WRKY33 is involved in drought by directly regulating the expression of CesA8.* American Journal of Plant Sciences, 2013. **4**(6A): p. 21-27.
180. Wang, Z., et al., *An oligo selection procedure for identification of sequence-specific DNA-binding activities associated with the plant defence response.* Plant J, 1998. **16**(4): p. 515-22.
181. Weickert, M.J., et al., *Stabilization of apoglobin by low temperature increases yield of soluble recombinant hemoglobin in Escherichia coli.* Appl Environ Microbiol, 1997. **63**(11): p. 4313-20.
182. Whitmore, L. and B.A. Wallace, *Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases.* Biopolymers, 2008. **89**(5): p. 392-400.
183. Willmott, R.L., et al., *DNase1 footprints suggest the involvement of at least three types of transcription factors in the regulation of alpha-Amy2/A by gibberellin.* Plant Mol Biol, 1998. **38**(5): p. 817-25.
184. Wu, K.L., et al., *The WRKY family of transcription factors in rice and Arabidopsis and their origins.* DNA Res, 2005. **12**(1): p. 9-26.
185. Wyre, C. and T.W. Overton, *Use of a stress-minimisation paradigm in high cell density fed-batch Escherichia coli fermentations to optimise recombinant protein production.* J Ind Microbiol Biotechnol, 2014. **41**(9): p. 1391-404.
186. Xie, Y., et al., *REVOLUTA and WRKY53 connect early and late leaf development in Arabidopsis.* Development, 2014. **141**(24): p. 4772-83.

187. Xie, Z., et al., *Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells.* Plant Physiol, 2005. **137**(1): p. 176-89.
188. Xu, X., et al., *Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors.* Plant Cell, 2006. **18**(5): p. 1310-26.
189. Xu, Y.H., et al., *Characterization of GaWRKY1, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)-delta-cadinene synthase-A.* Plant Physiol, 2004. **135**(1): p. 507-15.
190. Yamasaki, K., et al., *Solution structure of an Arabidopsis WRKY DNA binding domain.* Plant Cell, 2005. **17**(3): p. 944-56.
191. Yamasaki, K., et al., *Structural basis for sequence-specific DNA recognition by an Arabidopsis WRKY transcription factor.* J Biol Chem, 2012. **287**(10): p. 7683-91.
192. Yang, P.Z., et al., *A pathogen- and salicylic acid-induced WRKY DNA-binding activity recognizes the elicitor response element of the tobacco class I chitinase gene promoter.* The Plant Journal, 1999. **18**(2): p. 141-149.
193. Yasuda, M., et al., *Effect of additives on refolding of a denatured protein.* Biotechnol Prog, 1998. **14**(4): p. 601-6.
194. Zhang, H., et al., *PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database.* Nucleic Acids Res, 2011. **39**(Database issue): p. D1114-7.
195. Zhang, Z.L., et al., *A rice WRKY gene encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells.* Plant Physiol, 2004. **134**(4): p. 1500-13.
196. Zheng, Z., et al., *Functional analysis of Arabidopsis WRKY25 transcription factor in plant defense against Pseudomonas syringae.* BMC Plant Biol, 2007. **7**: p. 2.
197. Zheng, Z., et al., *Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens.* Plant J, 2006. **48**(4): p. 592-605.
198. Zhou, X., Y. Jiang, and D. Yu, *WRKY22 transcription factor mediates dark-induced leaf senescence in Arabidopsis.* Mol Cells, 2011. **31**(4): p. 303-13.
199. Zipfel, C., et al., *Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation.* Cell, 2006. **125**(4): p. 749-60.
200. Zipfel, C., et al., *Bacterial disease resistance in Arabidopsis through flagellin perception.* Nature, 2004. **428**(6984): p. 764-7.
201. Zou, C., W. Jiang, and D. Yu, *Male gametophyte-specific WRKY34 transcription factor mediates cold sensitivity of mature pollen in Arabidopsis.* J Exp Bot, 2010. **61**(14): p. 3901-14.

Part II

Structural studies of enzymes involved in phosphate metabolism

1. Introduction

1.1. Role of phosphorus in plants

Phosphorus (P) is one of the most important macronutrient essential for plant growth. It is used in biosynthesis of cellular components and is also involved in many biochemical pathways essential for plant growth, development and metabolism. As a major nutrient, phosphorus is exploited by plants in relatively large amounts. It is found in every living plant cell and is involved in several key plant functions, including energy transfer, photosynthesis, transformation of sugars and starches, protein activation, nutrient movement within the plant and transfer of genetic traits from one generation to the next. Phosphorus enters the plant through root hairs, root tips, and the external layer of root cells being absorbed by plants mostly as orthophosphate ions (Pi, $H_2PO_4^-$ or HPO_4^{2-}). This form is directly available. Nevertheless, phosphate availability depends on few factors. Due to high reactivity, phosphates might be absorbed only in narrow range of soil pH (pH 5-6) [89, 103]. The sorption of phosphate with soil constituents to metal oxides (Al, Fe) in acidic pH, mineralization to calcium or magnesium phosphates in basic environment [34] and its low rates of diffusion make orthophosphate ions the least readily available nutrient in the rhizosphere. Moreover, soluble form is very often converted into organic soil matter poorly available for plants [69]. The availability of organic phosphorus to support plant growth depends on the rate of their degradation to generate free phosphate. There are various enzymes such as phosphatases, nucleases and phytases involved in the degradation processes. Enzymatic hydrolysis of organic phosphorus is an essential step in the biogeochemical phosphorus cycle. Its acquisition is moreover facilitated by mycorrhizal fungi that grow in association with the plant's roots [29]. Mycorrhizal fungi function as an extended and highly efficient root system where their hyphae acquire, concentrate and transport Pi from soil that would otherwise be beyond the reach of the roots.

They are an important component of soil life and soil chemistry. Inside the plant root, P may be stored in the root or transported to the upper portions of the plant. Through various chemical reactions, phosphate is incorporated into organic compounds, including nucleic acids (DNA and RNA), phosphoproteins, phospholipids, sugar phosphates, enzymes, and energy-rich phosphate compounds such as adenosine triphosphate (ATP). Inorganic phosphate (Pi) is the predominant form of P directly absorbed by plant roots and a major transported form of P within the plants [69, 89]. In these organic forms as well as the inorganic phosphate ion, P is moved throughout the plant, where it is available for further

biochemical reactions. Due to its low availability and solubility in soil, it is often a limiting nutrient for their growth. Pi homeostasis is essential for the processes described above to operate at optimum rate for growth and development of the plant. and includes tight regulation of its concentration and transport within cellular compartments [74]. At adequate phosphorus supply of the plants, the vacuole acts as a storage pool of Pi and about 85 – 95% of the total Pi is located in the vacuole [8]. In contrast, in leaves of phosphorus-deficient plants, virtually all Pi is localized in the cytosol and chloroplasts, thus representing the 'metabolic pool' of Pi in the plant [8, 69].

When P is limiting, the most striking symptoms are a reduction in leaf expansion, leaf size, as well as in the number of leaves and development of a dark green leaf color. Shoot growth is more affected than root growth. However, root growth is also reduced by P deficiency, leading to limited water and nutrients uptake. Generally, inadequate P slows the processes of carbohydrate utilization, while carbohydrate production through photosynthesis is continued. When a deficiency occurs the P is translocated from older tissues to active meristematic tissues, resulting in deficiency symptoms appearing on the older plant leaves. Other effects of P deficiency on plant growth besides the premature senescence of leaves include delayed maturity and decreased disease resistance.

Plants have evolved different strategies to overcome limited Pi availability. In response to Pi starvation, plants have developed several physiological, biochemical and molecular adaptations to acquire phosphate (Pi). Plant increase ability to Pi uptake by altering root architecture [63, 82, 101], expression of Pi-regulated genes [11] or by changing their metabolic and developmental processes [86]. There is an increasing number of genes known to be activated under Pi starvation [109]. Altered gene expression are the hallmarks of plant adaptation to Pi deficiency.

1.2. Phosphate homeostasis

Phosphorus plays a vital role in virtually every cellular process that involves energy transfer. High-energy phosphate, held as a part of the chemical structures of adenosine diphosphate (ADP) and adenosine triphosphate (ATP). They act as a source of energy that drives the multitude of chemical reactions within the plant. Biosyntheses of macromolecules in living cells are characteristically accompanied by liberation of pyrophosphate (PPi), a byproduct of ATP hydrolysis. PPi is generated during many cellular processes and various biosynthetic reactions including nucleic acids, protein and polysaccharides biosynthesis, activation of tRNA and fatty-acids, coenzymes synthesis,

signal transduction and transformation of sugars and transport. This compound is quite stable in physiological conditions and specific enzymes are required to eliminate it from the cells.

Soluble inorganic pyrophosphatases are ubiquitous enzymes (EC 3.6.1.1) that catalyse hydrolysis of pyrophosphate (PPi) to two phosphates (Pi) and play key role in controlling intralellular PPi level. PPases are quite active and generally comprises as much as 0.1-0.5% of cell protein, therefore cellular PPi concentration is maintained at micromolar level [104]. Removal of PPi is essential for maintaining the direction of the reaction and presumably to drive anabolism [51]. PPi accumulation in the cells would affect cell functions, inhibits biosynthetic reactions, finally leading to cell death thus the PPases activity is a key importance in maintaining cells viability. Since PPases are central enzymes of pyrophosphate metabolism they provide phosphorus homeostasis.

Because of Pi level fluctuations, its concentration need to be tightly regulated to maintain homeostasis. Stable cytoplasmic level of phosphate in the cells is reached by enzymes belonging to different families and classes including: phosphatases, kinases, pyrophosphatases and apyrases or by subsequent Pi allocation into different cell compartments by specific transporters. Apyrases releases pyrophosphate from nucleoside triphosphates and diphosphates that is substrate for inorganic pyrophosphatases. Phosphatases liberate phosphate that is available for further reactions. Protein phosphatases and kinases are necessary for Pi homeostasis during the acquisition, storage, release, and metabolic integration of Pi [37, 45, 81, 84]. Precisely, protein kinases confer fine regulation to protein phosphorylation, whereas protein phosphatases are able to hydrolyze phosphomonoester metabolites, releasing inorganic phosphate (Pi) from these substrates [27, 38].

1.3. Inorganic pyrophosphatases

Inorganic pyrophosphatases (PPases) from various organisms have been studied and classified based on 3D structure similarity and sequence homology between two major classes: soluble PPases occurred mainly in cytoplasm and membrane-integral H^+ and/or Na^+-translocating PPases [41, 47, 93]. The linkage of those classes is substrate - inorganic pyrophosphate, but they do not show any sequence or structure similarity to each other [44] and also exhibit distinct catalytic mechanism.

Membrane PPases occur mainly in plants, algae and in some bacteria or protozoa. Those enzymes have a completely different architecture and function comparing to soluble

PPases (Fig. 1) as they are primarily ion pumps producing proton and/or sodium gradients. They are coupling the synthesis of PPi to H^+ and/or Na^+ pumping [65, 67] and have definitely broader biological function as they are crucial for survival of plants and bacteria under various stress conditions [21, 64, 66]. They occur in species where energy limitation is frequent, and are important during nutrient deficiency and stress conditions such as drought, anoxia, cold, low-light intensity or high salinity to provide ion gradients when ATP is limited [21, 64, 66]. In eukaryotes, they occur predominantly in the vacuolar membranes of plants [66] and in the acidocalcisomal membranes of protista [68].

Among class of soluble pyrophosphatases, two unrelated families were defined [31, 94]. Family I includes most of the currently known eukaryotic, archaebacterial and eubacterial soluble PPases, with the best characterized representatives from *Saccharomyces cerevisiae* [32, 33] and *Escherichia coli* [31, 88]. In *Arabidopsis thaliana* five soluble inorganic pyrophosphatases classified into family I were identified.

Family II occurs almost exclusively in bacteria. Prokaryotic representatives comprises enzymes from *Bacillus subtilis* [94], *Streptococcus gordonii* [2], few archaeal species, including *Methanococcus jannaschii* and several other bacterial strains [55, 87].

Despite their common function, family II PPases have a completely different three-dimensional structure and fold topology when compared to the family I enzymes (Fig. 1). All PPases from family I characterised so far are oligomeric enzymes but differ in size and number of subunits. Prokaryotic type enzymes occur as either homotetramers or homohexamers with a subunit molecular mass of approximately 20 kDa. Eukaryotic PPases were submitted so far to function as homodimers with 28-35 kDa subunits. Plant PPases are an exception in this respect and they have been reported to function as 25 kDa monomers [77]. Besides the differences in oligomeric structure there are appreciable sequence similarities between all of them [15, 73]. Moreover, all Family I PPases share the same structural fold and conserved active site [15, 32]. The common structural feature is a twisted five-stranded β-barrel core, with differences existing in connecting loops and, for the eukaryotic PPases, the N and C-terminal extensions.

In contrast to family I PPases, which have a single-domain structure, family II PPases have two domains joined by a flexible hinge with the active site [2, 72], formed at the interface between the N and C-terminal domains. Representatives of Family II function as homodimers with 34 kDa subunits.

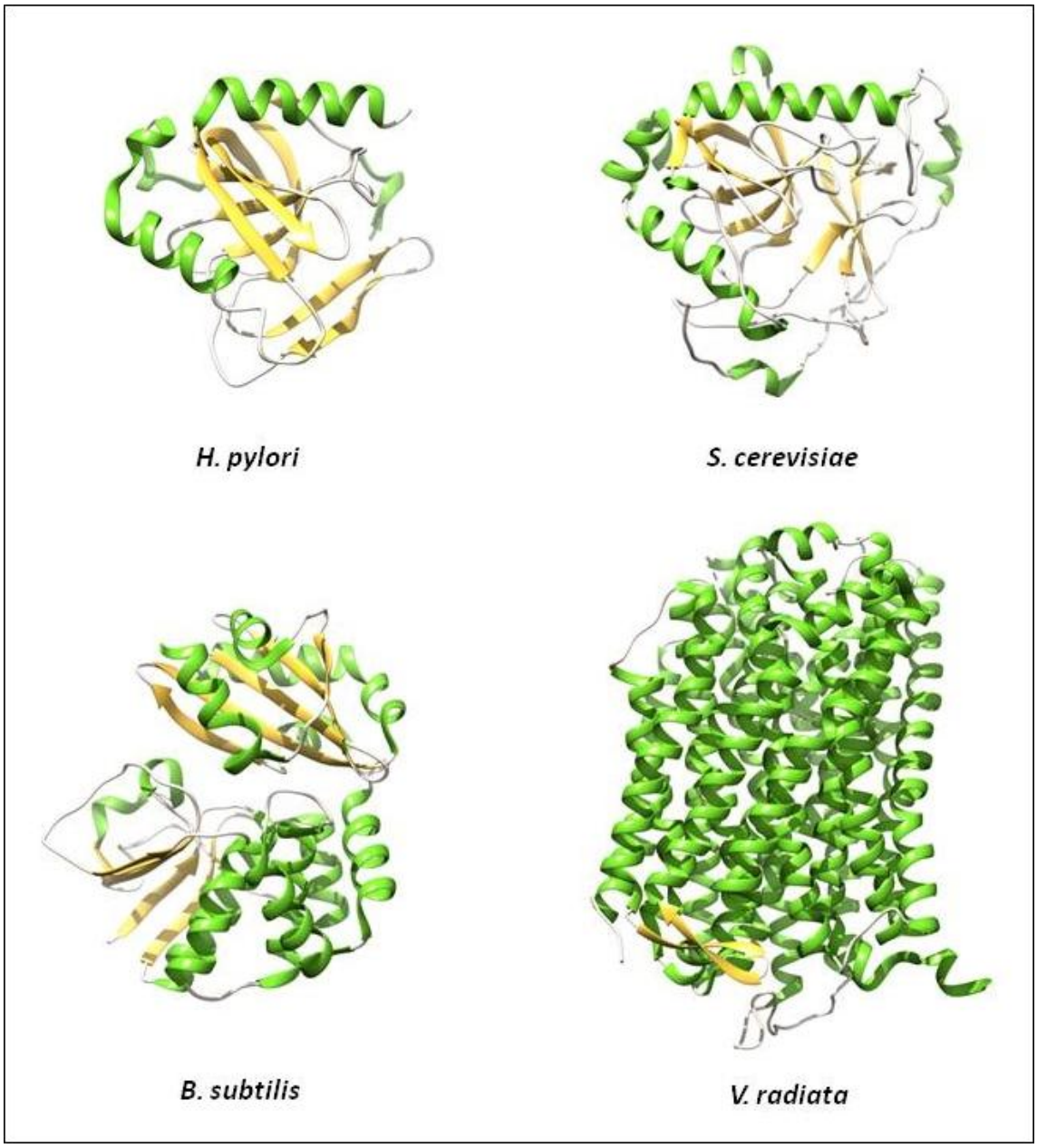


Fig. 1. PPases representatives of different families. Family I: *H. pylori* (PDB code: 1ygz) and *S. cerevisiae* (PDB code: 1e6a), Family II: *B. subtilis* (PDB code: 1wpm), Membrane PPase from *V. radiata* (PDB code: 4a01). The picture presents all structures as monomers that forms (besides membrane PPase) higher ordered structures as described in the text.

1.4. Plant PPases

The presence of soluble PPases in higher plants has been documented [49, 95], but the enzymes remain poorly characterized [77]. So far, two soluble PPases from model plant – *Arabidopsis thaliana* that belongs to Family I: AtPPA1 and AtPPA4 have been cloned, produced as recombinant proteins in *E. coli* fused to gluthatione-S-transferase (GST) and characterized [77]. They have been described as rather compact monomers and their size

was estimated using size exclusion chromatography. Those monomers retained the PPase activity. However, whole *Arabidopsis thaliana* genome coding six soluble inorganic pyrophosphatases. Sequence alignment showed that five of them (AtPPA1-5) were very conserved with about 71-88% identity. Those five occur in cytoplasm or as indicated in past reports might be targeted to various cell compartments [7, 105]. The AtPPA6 protein was only 22% identical with the other five known protein sequences. Moreover, *in vitro* import experiments demonstrated that the AtPPA6 protein could be imported into chloroplasts and localized in the stromal compartment [92]. It was demonstrated by *in vitro* import experiments [92], but taking into account that in plastids, many biosynthetic reactions take place and high amount of PPi is produced, the PPase activity appears to be vital. AtPPA6 was also found to be closely related to the predicted protein sequence from *Chlamydomonas rheinhardtii* annotated as chloroplastidic isoform [22]. An alignment of the *Arabidopsis* sequences and dendrogram shown in Fig. 2 and 3 respectively, demonstrate that AtPPA6 protein have an N-terminal extension in comparison with the other five *Arabidopsis* proteins. The predicted molecular weight of AtPPa6 is higher, and was estimated as 33.4 kDa in comparison with 24.5–25 kDa for the other *Arabidopsis* sequences. In addition to soluble PPases, *Arabidopsis thaliana,* as mentioned earlier, possess one much larger membrane-integral PPase, which works as a reversible proton pump but does not have any sequence similarity to the five others soluble PPases. Membrane PPase is much larger, have completely different architecture. Its primary function is ion pumping to produce proton or sodium gradients during low-energy stress conditions but not pyrophosphate hydrolysis [41, 47, 93].

Expression pattern of *Arabidopsis thaliana* PPases genes has been analyzed in different plat tissues during all developmental stages [77]. Available information on the expression of AtPPA genes indicates that they are expressed in diverse plant photo- and hetero-trophic tissues [77], indicating tissue specificity. The AtPPA1 and AtPPA4 are almost ubiquitous enzymes. They are expressed in all tissues at different levels. Comparison of expression levels of AtPPA1 and AtPPA4 showed that AtPPA4 is expressed to a lower level than AtPPA1 in most tissues except cell suspensions stems, hypocotyls and nodes. However, there is a slightly lower level of AtPPA1expression in ovary, stigma and pollen. In addition, AtPPa6, another monomeric plant pyrophosphatase, reported to be imported into the chloroplasts [92], showed expression patters similar to those of AtPPA1, except for a lower expression level in tissues that have reduced photosynthesis (pollen, senescent leaf, hypocotyls and xylem). In contrast, other pyrophosphatase gene AtPPA3 is expressed

preferentially in stamen, pollen and flower but at a low level in lateral roots and root elongation zones, with expression levels below detection in all other tissues.

Analyses of AtPPases expression at particular developmental stages revealed that AtPPA1 is expressed to a high level at all stages, with a slight reduction during senescence. AtPPA4 and AtPPA6 were found to be expressed also at all developmental stages, but at definitely lower level than AtPPA1, moreover, their expression was scarcely detectable in young plants and reached a maximum during flowering [77].

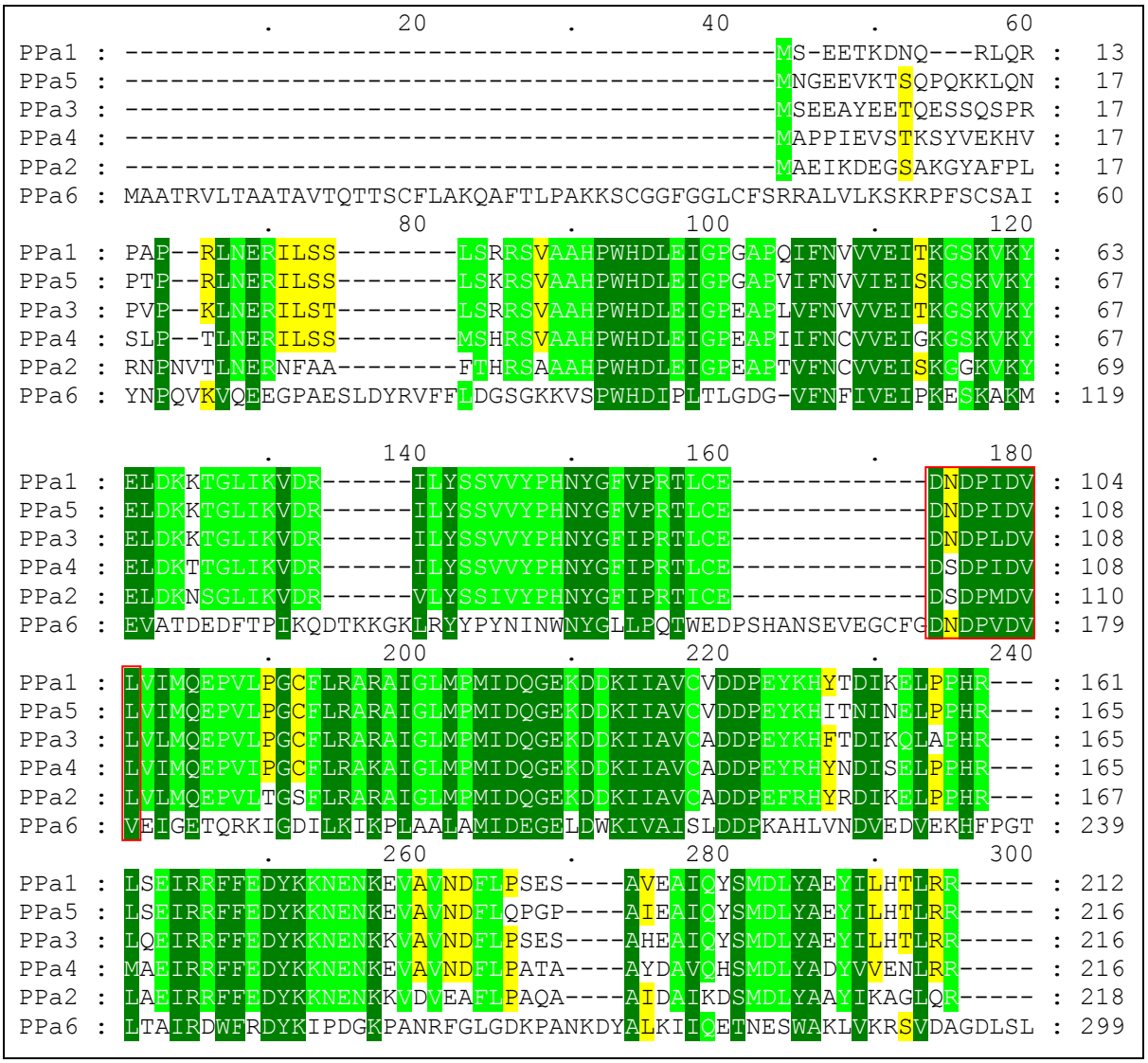


Fig. 2. Multiple amino acid sequence alignment of six inorganic pyrophosphatase (PPase) from *A. thaliana.* The level of conservation is expressed by the darkness of the lettering background. The pyrophosphatase amino acid sequences (locus ID shown in parentheses) from *A. thaliana* have been aligned: AtPPA1 (At1g01050), AtPPA2 (At2g18230), AtPPA3 (At2g46860), AtPPA5 (At4g01480), AtPPA6 (At5g09650)**.** The alignment was calculated in ClustalW [60] and visualized in GenDoc [78].

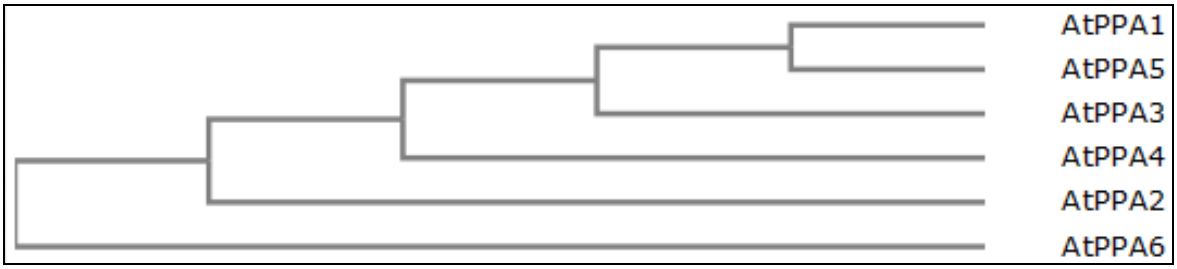


Fig. 3. Phylogram showing the evolutionary distance of AtPPases. Amino acid sequences were aligned using the ClustalW2 program [60]. This alignment was used for generation of the dendrogram using the ClustalW2-Phylogeny online tool. The UPGMA method with uncorrected distances was used.

Sequence alignment from different organisms within family I shows that the sequence identity vary in range of 23-47% and PPases from Family I can be divided into two groups: eukaryotic type including *S. cerevisiae* PPases and prokaryotic type including those of bacterial origin. Surprisingly, plant PPases, belong to the latter type.

There were several structures of soluble PPases representing Family I type deposited in the Protein Data Bank (PDB) but any of them originated from plants. So far, only predicted tertiary structure for *A. thaliana* PPase is available. It was done using homologous modelling. The structures of *S. cerevisiae* PPase and *E. coli* PPase were determined at 2.2-2.3 Å resolution–and used as molecular probes. The modelling tool calculated only the positions of equivalent residues that overlay with probe structures [96]. In this thesis, I present high-resolution crystal structures of *Arabidopsis thaliana* inorganic pyrophosphatase (PDB Code: 4lug) corresponding to the sequence encoded by *ppa*1 gene. This is the first solved crystal structure of pyrophosphatase from higher plants.

1.5. Mechanisms of the PPases activity

The biological function of both soluble PPase Families (I and II) is identical as they mainly hydrolyze the inorganic pyrophosphate (PPi) yielding orthophosphates (Pi). However, sequences and structures of PPases of the Family I and II are unrelated and the mechanism of the PPi hydrolysis is not exactly the same [72]. Moreover, Family II PPases include members that are allosterically regulated [39], absent in the other families.

Catalytic mechanism was deduced according to data from extensive crystallographic studies of eukaryotic and prokaryotic enzymes from *S. cerevisiae, E.coli* (Family I) and *B. subtilis* (Family II) but the precise mechanism of catalysis *via* inorganic pyrophosphate in most organisms still remains uncertain.

Similarly to many enzymes involved in phosphate metabolism, PPases are metal-dependent and divalent metal cations are crucial for their catalytic activity. The natural

metal cofactor of family I PPases is Mg^{2+} and it binds to the enzyme with micromolar affinity [3], whereas family II enzymes are best activated by Mn^{2+} or Co^{2+}, which bind with nanomolar affinity [83]. Moreover, family II PPases bounded to Mn^{2+} are 10-fold more active than family I PPases activated with Mg^{2+} (k_{cat} : 1700-3300 s^{-1} *versus* 110-330 s^{-1}) [55, 83, 111].

PPases from Family I are active also at the presence of other than Mg^{2+} ions. However, Mg^{2+} is most powerful effector and enzymatic activity with other ions as cofactors decreases as follows: $Mg^{2+} > Mn^{2+} > Zn^{2+} > Co^{2+} > Cd^{2}$ [3]. In presence of $Mn^{2+} > Zn^{2+} > Co^{+2}$ but not Mg^{+2} PPases acquire the ability to hydrolyze ADP or ATP [91], thus the substrate specificity might be determined by the metal ions bound to the protein, rather than to substrate.

PPases belonging to Family I posses only one domain forming β-barrel. The active centre is localized between α-helix and the top of the barrel. Novel studies accept that catalysis happens inside an inorganic-metal-phosphate cage. The catalysis proceeds by direct water attack on PPi without formation of phosphorylated enzyme [24]. The protein sidechains have mostly supporting roles in setting metal ions, water and substrate in right positions to allowing catalysis [32]. Catalytic mechanism is similar to other hydrolases supported with magnesium ions and is multistage. First, inorganic pyrophosphatase binds two magnesium ions; at the same time other magnesium ions activate substrate and water molecule that acts as nucleophile. When all metal ions, water molecules and enzyme side chain residues occupy necessary positions, subsequent hydrolysis is possible. The number of bound magnesium ions changes during hydrolysis reaction. The next magnesium ion can bind also to the enzyme-product complex [3]. As mentioned earlier, all known PPases require presence of divalent metal cations, with magnesium conferring the highest activity. It has been shown on the basis of structural studies and sequence analysis that 13-17 conserved amino acids are responsible for activity of all known PPases that belong to the family I. The best studied PPases are *S cerevisiae* and *E. coli* enzymes. The active site is very well conserved, for instance 13–14 of the 17 polar active-site residues are fully conserved in the structure-based sequence alignments of *S cerevisiae* and *E. coli* PPases [46]. Among the conserved residues, the motif D-(S/G/N)-D-P-*ali*-D-*ali*-*ali* where *ali* = C/I/L/M/V (Fig. 2, red frame) has been postulated to be active site [46]. This highly conserved region include three functionally important aspartic acid residues that bind three or four divalent metal cations and help to activate water molecule [12, 33, 88]. Magnesium ions in the active site have multiple functions providing correct conformation for substrate binding. They also

increase electrophilicity of the phosphorus atom, determine the degree of binding and activate water molecule.

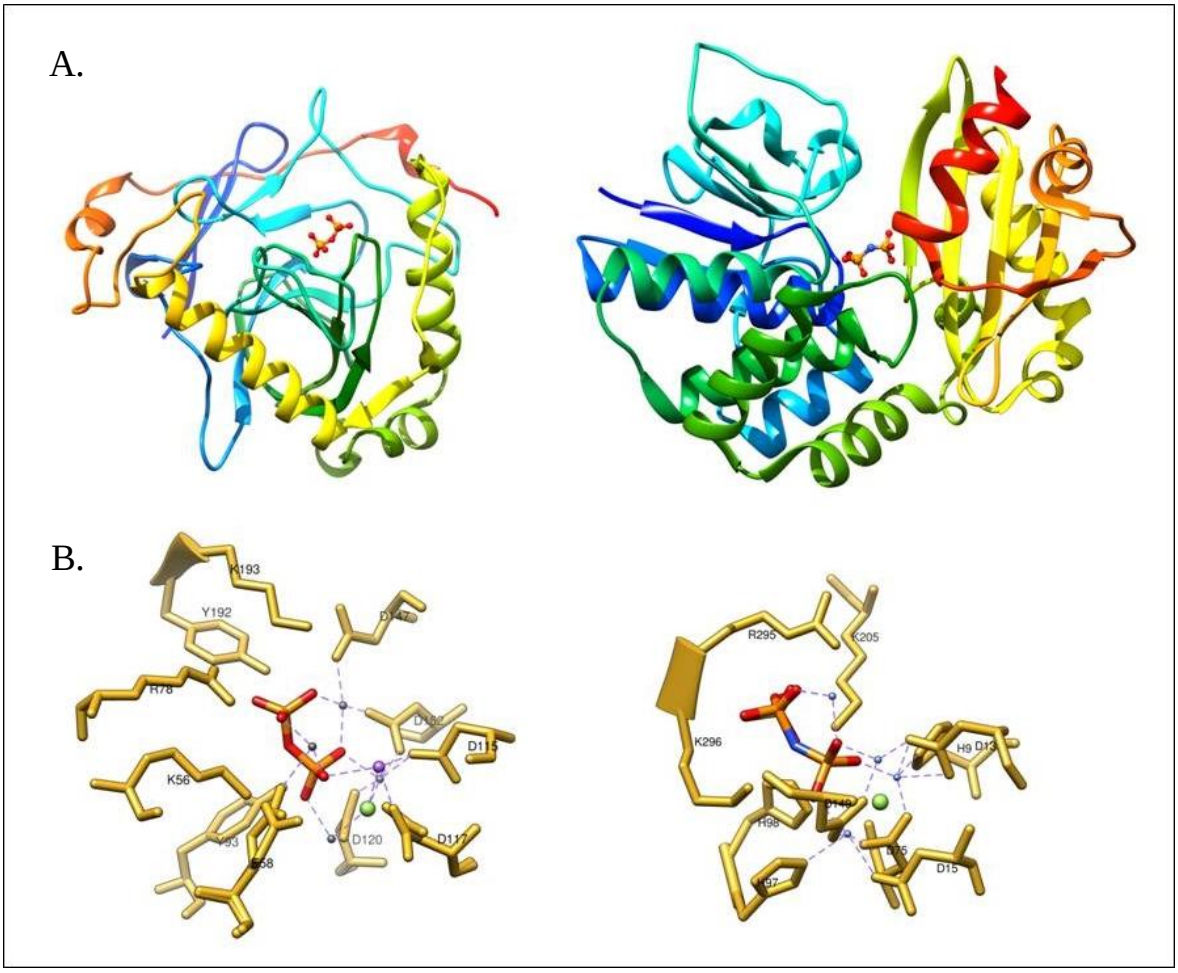


Fig. 4. Comparison of family I and family II PPase structures.
A. schematic models of *S. cerevisiae* family I PPase (PDB:1e6a, *left*) and *B. subtilis* family II PPase (PDB:2haw, *right*). Substrates are shown as balls and sticks: PP_i for family I and PNP for family II. The structures are coloured from blue at the N-terminus to red at the C-terminus.
B. Active sites of family I *S.cerevisiae* PPase (*left*) and family II *B. subtilis* PPase (*right*). Metals are shown as balls, the Mn^{2+} in family I PPase is *gray* and Mg^{2+} in family II PPase is *blue*. Labels for the active site residues are included according to the amino acid one-letter code and sequence numbering.
PP_i and PNP are shown as sticks: *red*, oxygen; *blue*, nitrogen and *orange*, phosphorus. Fluorides are shown as *green* balls.

Family II PPases have a very different three-dimensional structure and topological fold in comparison to the well-characterized enzymes of the type-Family I. They fold into two domains, with the active site at the interface formed between the N and C-terminal domains. The domains are linked by region with a large degree of flexibility enabling a large conformational change in the quaternary structure between the open (without substrate) and closed (with bound substrate) conformations.

The C-terminal domain contains the binding site with high affinity to substrate, whereas the catalytic site is located in the N-terminal domain. The pyrophosphate binding trigger the conformational changes. The C-terminal domain closure onto the N-terminal end brings the substrate into catalytic site which is formed at the domains interface. Structural studies of PPases from *Streptococcus gordonii* and *Bacillus subtilis* revealed that the surfaces involved in forming the active site are largely hydrophilic, consisting almost entirely of acidic and basic amino acid residues.

In the catalytically competent conformation the water molecule bridges the three manganese ions that polarizing P-O bonds in substrate. The metal ions are coordinated by the characteristic DHH amino acid signature (one aspartic acid and two histidine residues). Correctly orientated pyrophosphate is arranged next to the water, ready for nucleophilic attack on the substrate.

However, the formal reaction catalyzed by the two classes of PPase is the same, they are not related in sequence, structure and catalytic mechanism. The differences in mechanism are due to employing other activating metal ions and their coordination. In Family I, Mg^{2+} ions coordinate the water molecule, while in Family II PPases, the Mn^{2+} ions are presented. The known crystal structures showed that the preference for Mn^{2+} over Mg^{2+} in family II PPases is due to presence of histidine residues that bind the metal ions and bidentate carboxylate coordination of metal ion by aspartic acid at the binding site [72]. Recent structural data revealed one more reason related to the different chemical properties of Mg^{2+}(family I) *versus* Mn^{2+} (family II). Upon substrate binding by Family II PPase, the coordination number of the high affinity metal site changes from five to six [20]. The five/six-coordinated geometry is typical for transition metals, such as Mn^{2+} or Co^{2+}, but not for Mg^{2+}, which is usually six-coordinated [28]. Therefore, Mn^{2+} and Co^{2+} fit the active site in family II PPases better than Mg^{2+}.

2. Goal of the thesis

The sequence identity between the *Arabidopsis thaliana* and the other pyrophosphatases (from both prokaryotes and eukaryotes) is below 45%, and the plant enzymes are more similar to the bacterial cytoplasmic pyrophosphatases than to homologs from other organisms. Moreover AtPPA1 was reported to be a monomer in contrast to *E.coli* hexamer and *S. cerevisiae* dimer. The proteins of the same function exhibit structural diversity. Detailed structural studies, comparison of the 3-D structures and sequences from various sources may help to explain the diversity in occurrence of oligomeric forms.

The main goal of my dissertation was to obtain biologically active recombinant AtPPA1 protein and determine its crystal structure.

3. Results and discussion

3.1. Cloning, overexpression and purification of AtPPA1

cDNA of *Arabidopsis thaliana* Columbia-0 wild type plants was used as template to isolate and cloning of target AtPPA1. The specific primers for PCR were designed according to the AtPPA1 sequence available in the TAIR database (http://www.arabidopsis.org). AtPPA1 coding DNA fragment was cloned into pMCSG48 vector using LIC method and the protein was produced as fusion with His-tag followed by NusA and TEV protease cleavage site present between tags and protein sequence. Bacteria were grown at 37°C prior induction, then the temperature was decreased to 18°C and the protein expression was induced with 0.5 mM IPTG. Soluble recombinant AtPPA1 was overexpressed during overnight cultivation of *E. coli* cells.

The protein was purified as described in Materials and methods (section 4.2.1.1). Briefly, the recombinant protein was purified from the bacterial lysate by two consecutive chromatographic steps: chelating chromatography on Ni^{2+}-charged agarose resin followed by cleavage of fusion protein by TEV protease and FPLC size exclusion chromatography.

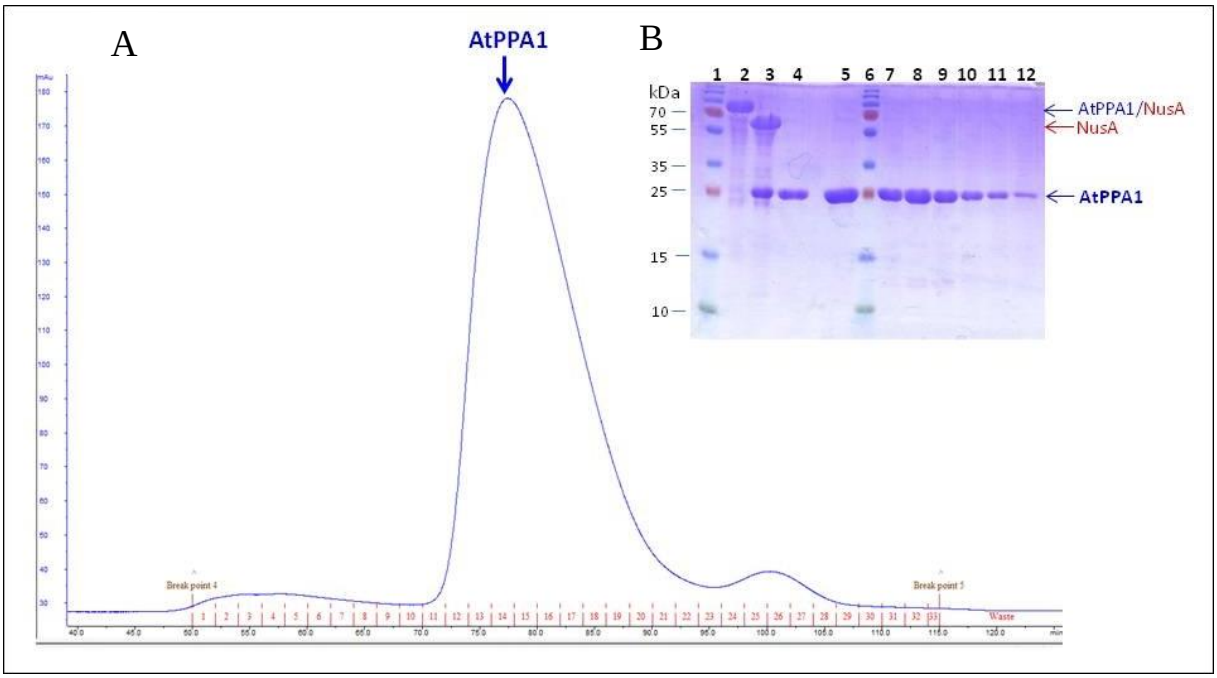


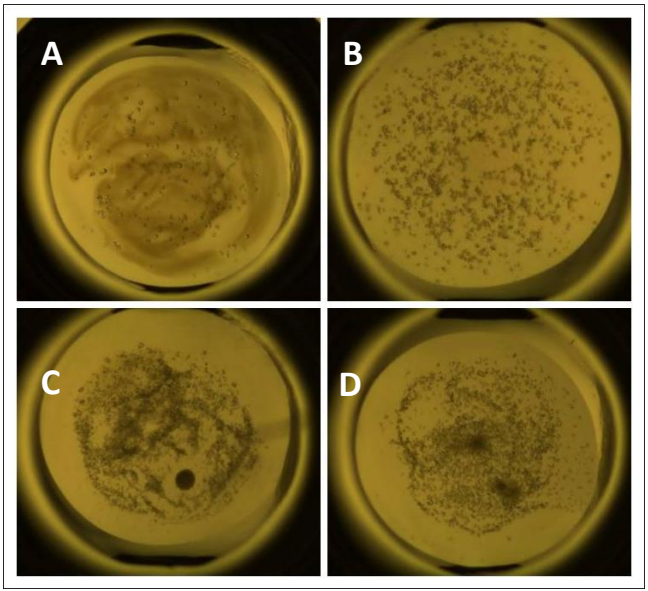
Fig. 5. (A) Size exclusion chromatography on a Superdex 200 FPLC column (GE Healthcare) of the *At*PPA1; (B) SDS-PAGE of AtPPA1 purification steps. Lane: 1, protein ladder; 2, protein after elution from first Ni column; 3, protein after TEV cleavage; 4, protein eluate after second Ni column; 5, AtPPA1 loaded on gel filtration column; 6, protein markers; 7-12, peak fractions after gel filtration.

The final purification step (size exclusion chromatography on Superdex 200 16/60HL column, GE Healthcare) yielded a homogenous protein fraction of AtPPA1 (approx. 10 mg

of pure protein from 1 L culture). The chromatography peak with a maximum UV absorbance was eluted with $V_e \approx 75$ ml. The purification steps were analyzed on SDS-PAGE (Fig. 5.). After all chromatographic steps, the protein was visible on SDS-PAGE as one band of 24 kDa. The pure protein fractions from the peak were concentrated to 7.5 mg/ml and used immediately for crystallization or other analyses.

3.2. Crystallization conditions of AtPPA1

For initial screening of AtPPA1 crystallization conditions, Robotic Sitting Drop Vapor Diffusion setup (Gryphoon, Art Robins Inc.) was applied. Set of three screens: JCSG plus, PACT Premier and Morpheus (Molecular Dimensions) was used for the initial high-throughput experiments. Crystals appeared in 15 conditions and they presented two morphological forms: long and extremely thin needles or irregular-shaped blocks. From those conditions I choose four for manual optimization (Fig. 6):



A) 0.06 M Divalents, 0.1M Buffer system 1 pH 6.5,GOL_P4K 30.0%;

B) 0.1 M Bicine pH 9.0, 10 % w/v PEG 6K;

C) 0.1 M Bicine pH 9.0, 10 % w/v PEG 20,000/ 2% v/v Dioxane;

D) 0.1 M succinic acid pH 7.0, 15 % w/v PEG 3350

Fig. 6. AtPPA1 crystals from robotic plates.

Manual optimization was performed using hanging drop method. Crystallization drops were prepared by mixing the protein solution with the reservoir solution in the ratio 2:1 or 3:1. Crystals appeared within approximately 14 days at 19°C and reached final dimensions after 1 month of growth. After optimization, the irregular crystal blocks that differed in size were obtained (Fig. 7). The crystals were harvested using 0.2-0.4 mm nylon loops (Hampton Research), soaked with cryoprotectant containing 20% glycerol or 20% PEG400 (v/v) in a solution similar to the well solution and finally vitrified in liquid nitrogen for synchrotron-radiation data collection. The best diffracting crystal appeared in the following

condition: 0.1 M succinic acid pH 7.0, 15 % w/v PEG 3350 and 20% glycerol as cryoprotectant.

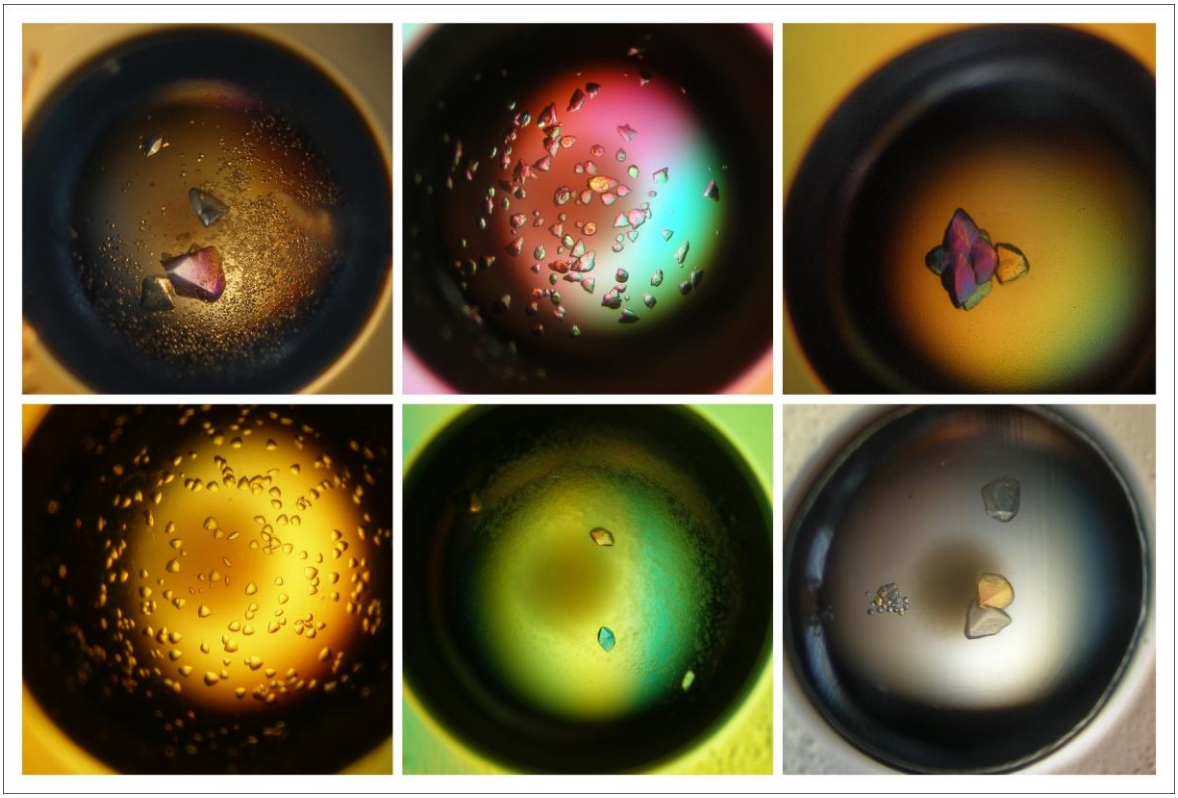


Fig. 7. Crystals of AtPPA1 after optimization.

X-ray diffraction data were collected on the beam line 14.1 at the BESSY synchrotron in Berlin. A total number of 120 diffraction images with 1° oscillation were integrated and scaled using the XDS software [43]. Summary of the X-ray data collection and processing is presented in Table 1.

Table 1. AtPPA1: Data Collection

Wavelength (Å)	0.91841
Space group	H3
Crystal system	trigonal
Unit cell parameters	a=82.02, b=82.02, c=175.08 α=90, β=90, γ=120
Resolution	41.0-1.93 (2.03-1.93)[a]
Unique reflections	33078
Completness	99.7 (98.5)
R_{merge} *(%)*	4.3 (59.8)
$<I/\sigma(I)>$	18.76 (2.04)

[a]Values in parentheses correspond to the highest resolution shell

3.3. Structure solution, refinement and deposition

The crystal structure of AtPPA1 was solved by molecular replacement using Phaser [71]. The crystal structure of a homologous protein from bacteria *Pyrococcus furiosus* (PDB code: 1twl) was used as a search probe. These proteins share 49% sequence identity. ARP/wARP [59] was applied for automatic model building. Coot [19] was used for manual fitting in electron density maps between model refinement and validation cycles carried out in phenix.refine [1]. Hydrogen atoms were added at riding positions. Model was accepted as final with twelve TLS groups [107] defined as suggested by refinement program. The progress of the refinement was monitored and validated using 1005 reflections set aside for R_{free} testing [10]. The final model has an R_{work} of 0.156, R_{free} 0.200 and includes two protein chains in asymmetric unit, 173 water molecules and a two sodium ions. The final model was validated using MolProbity [13] and diffraction precision index was calculated using SFCHECK [102]. A Ramachandran plot generated by PROCHECK [61] showed that the structures have reasonable stereochemistry with no residues in disallowed regions (Fig. 8). The final refinement statistics are listed in Table 2. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession code: 4lug.

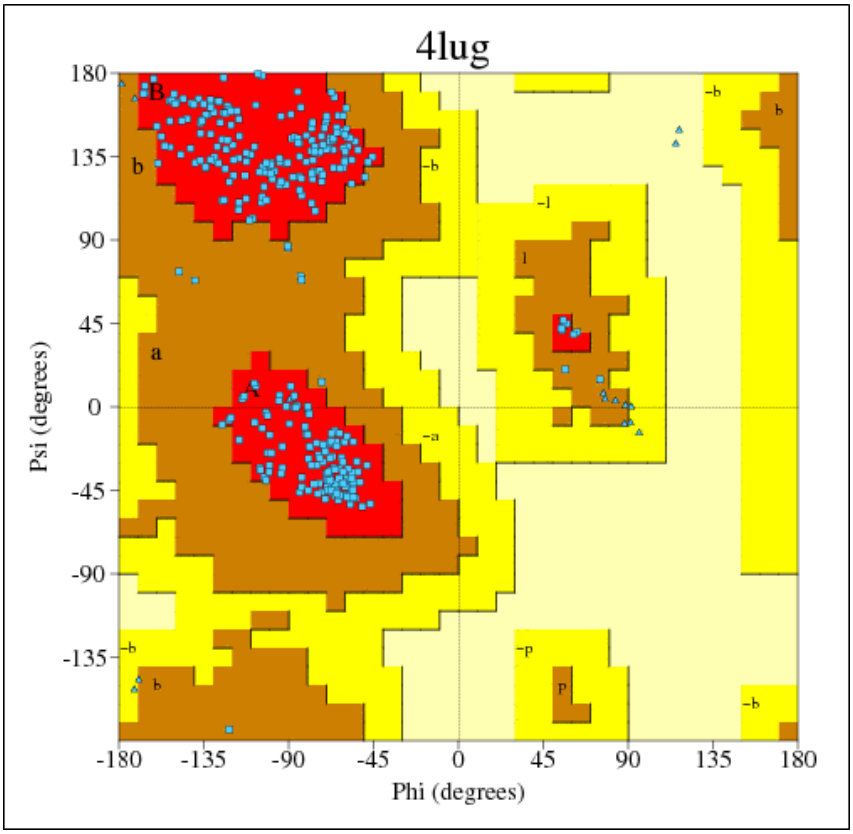


Fig. 8. Ramachandran plot for AtPPA1 (PDB code: 4lug) generated by PROCHECK.

Table 2. AtPPA1: Refinement statistics

Structure solution	Molecular replacement
Model used	1twl
Programs	
structure solution	Phaser
model building	ARP/wARP
manual fitting	Coot
refinement	phenix.refine
No. reflections Work/test	32974/1005
No. of non H atoms	3015
protein/solvent	2842/173
R_{work} (%)	15.6
R_{free} (%)	20.0
R.m.s deviations from ideal geometry	
bond length (Å)	0.019
bond angle (Å)	1.396
Ramachandran statistics (%) favoured/allowed	95.7/4.3

3.4. Overall structure of AtPPA1

The correct solution of the AtPPA1 was found in the rhombohedral H3 space group with two protein chains, labelled A and B, in the asymmetric unit. However the dimer has no biological significance because AtPPA1 forms quaternary structure and appears as a trimer in solution. A biological trimer could be generated by the application of the crystallographic 3-fold axis to the each monomer present in the asymmetric unit. The complete sequence of AtPPA1 is composed of 212 amino-acid residues and its molecular mass is 24.5 kDa. The initial model contained 175 or 176 of the 212 residues in A and B protein chain respectively. The remaining residues of the model have excellent definition in electron density. Only the last 4 or 5 C-terminal residues (chain B and chain A respectively) were disordered and could not be modelled due to weak electron density maps. The solved structure lacking 32 N-terminal residues. Those residues could not be modelled in electron density, only truncated ~20 kDa protein was observed. The latter analyses confirmed that this fragment of AtPPA1 protein was cleaved and was not presented in the crystal. In the structure one sodium cation per chain is presented. Summarising, the current model of *A. thaliana* inorganic pyrophosphatase includes residues 31-207 in chain A, 31-208 in chain B, two sodium cations and 173 water molecules.

In general, structure solution of the AtPPA1 reveals a compact fold, which is similar to the canonical fold of other PPases from family I. PPase is globular protein and belongs to the α+β class of protein folds. AtPPA1 fold is best described as OB (oligonucleotide binding)-fold [76]. In details, AtPPA1 protein fold consists of nine β-strands and two α-helices arranged in a β1-β2-β3-β4-β5-β6-β7-β8-α1-β9-α2 topology (Fig.8). Five β-strands form an antiparallel β-barrel (strands β1, β4, β5, β6, β8) capped on top by helix α1 and on the bottom by a loop between strain β2 and β3. They are surrounded by second helix α2, one helical turn (between α1 and β8) and a β-hairpin (β1-β2). The overall fold of a monomer of AtPPA1 is the same as in the bacterial structures and the core is the same as yeast PPase. The cavity between α1 helix and β-barrel correspond to the active site. The one sodium ion present in active site is coordinated by side-chain carbonyl O atoms (Asp98, Asp103) and 4 or 3 water molecules in chain A and B respectively.

Helix α1 is lean about 60° to helix α2. Both helices are conserved and overlap with the corresponding helices from *E. coli* and *S. cerevisiae* PPases. All secondary structure elements are well preserved in all Family I PPases deposited in PDB.

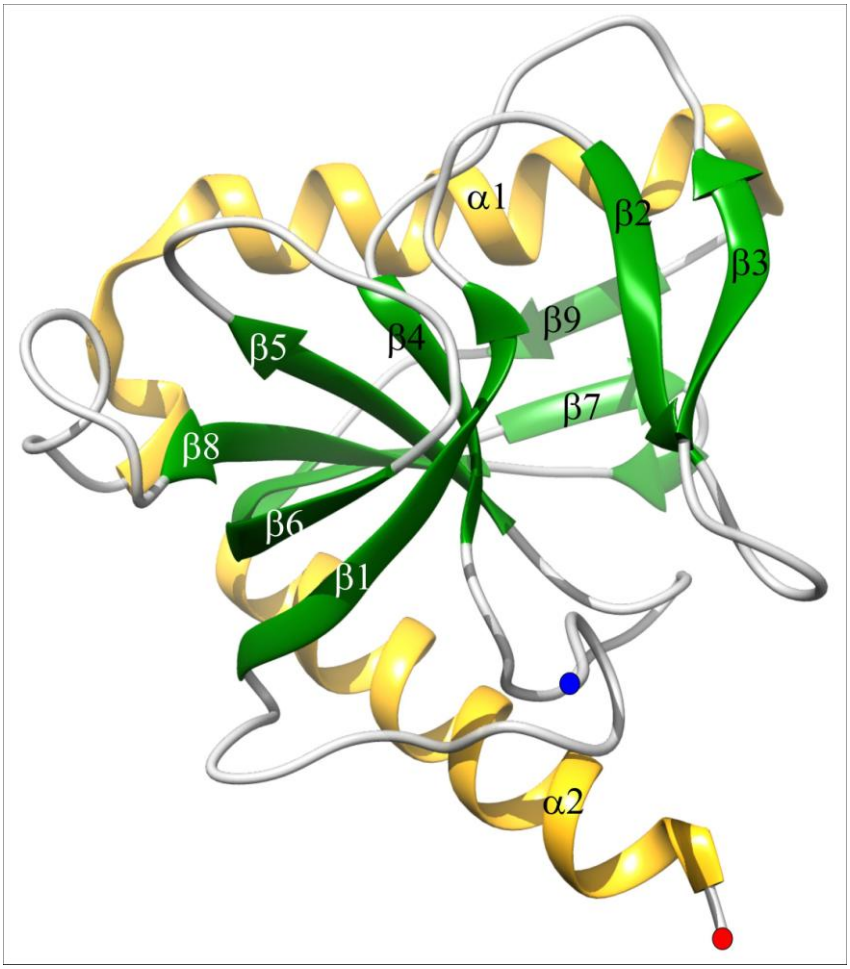


Fig. 9. Overall structure of AtPPA1. The N- and C-termini are is marked by blue and red dots respectively.

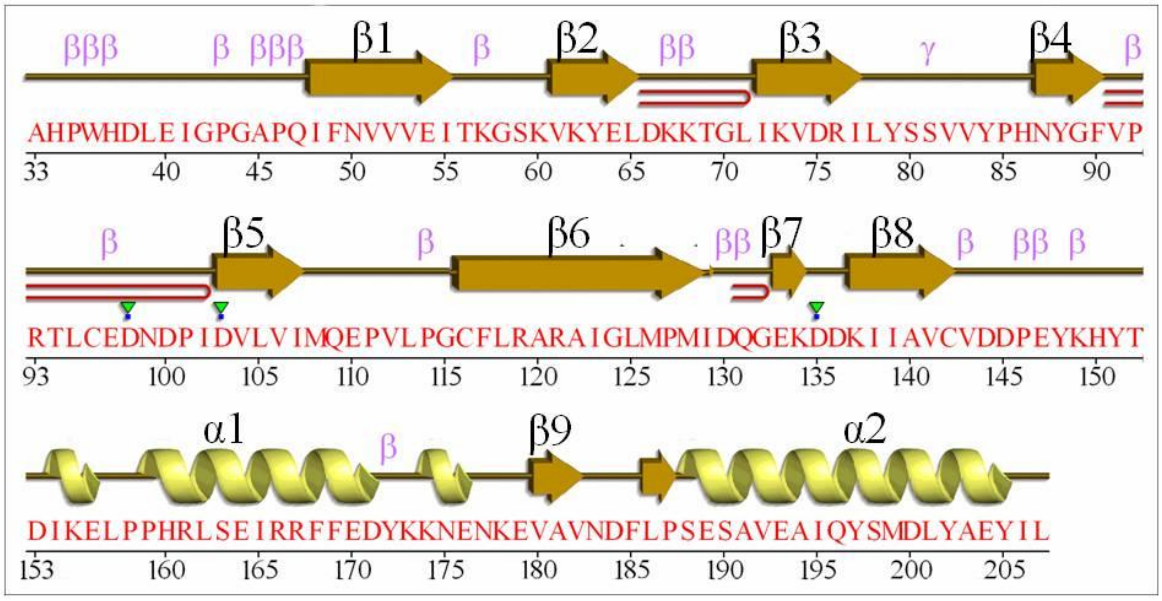


Fig. 10. Secondary structure elements of AtPPA1. The α-helices, β-strands are numbered in order from N- to C-terminus (black letters). β- and ϒ-turns (purple letters) are indicated as well as β-hairpins (red hairpin). The Na^+ binding residues are marked by green triangles with blue dots.

3.5. Metal ions associated with the AtPPA1 protein

In the active site cavity of each monomer, only one Na^+ cation is coordinated by two aspartic acid side-chains and three or four water molecules. The identification of metal cation was based on the octahedral coordination and metal-oxygen distances that were shorter (2.2–2.5 Å) than typical hydrogen bonds. The recognition of the metal cations was also confirmed using the Calcium Bond-Valence Sum (CBVS) method [75]. Detailed information including distances and angles for metal cations coordination are summarized in Table 3.

Table 3. Distances between active site ligands and protein showing hydrogen bonding and metal coordination.

Na/301Na/A	Distance (Å)			Angles (^{o})		
O/459HOH/A	2.13					
O/440HOH/A	2.32	160.15				
OD2/103ASP/A	2.33	95.61	103.18			
O/433HOH/A	2.34	77.29	87.26	121.94		
OD2/98ASP/A	2.59	96.81	69.08	152.60	84.69	
	Na/301Na/A	O/459HOH	O/440HOH	OD2/103ASP	O/433HOH	OD2/98ASP
Na/301Na/B	Distance			Angles		
O/419HOH/B	2.34					
O/424HOH/B	2.40	98.37				
O/498HOH/B	2.42	71.21	166.38			
0/474HOHB	2.47	159.44	82.36	104.25		
OD2/103ASP/B	2.57	77.29	114.67	78.18	84.30	
OD2/98ASP/B	2.63	113.43	71.09	65.40	65.40	148.58
	Na/301Na/B	O/419HOH	O/424HOH	O/498HOH	0/474HOH	OD2/103ASP

The Na^+ binding site roughly corresponds to the M1 site occupied by one of the catalytic Mg^{2+} ions in *E. coli* PPAse (PDB: 1obw) and this position aligns with K^+ binding site in *M. tuberculosis* (1wcf) . In *Arabidopsis thaliana* structure, Na^+ is bound by two aspartic acid residues: Asp98 and Asp103 (Fig. 11) which are analogous to Asp65 and Asp70 residues in *E.coli* that coordinate Mg^{2+} in M1 site (PDB:1obw) and Asp57 and Asp89 residues in *M. tuberculosis* that coordinate K^+.

The only difference is that Na^+ is coordinated by two amino acid residues: Asp98 and Asp103 whereas in *E. coli* the metal cation is coordinated by three residues: Asp65, Asp70 and Asp102.

The presence of Na^+ ions in the structure reflects to high sodium concentration (≈100 mM) in crystallization buffer rather than to any physiological role [26, 56].

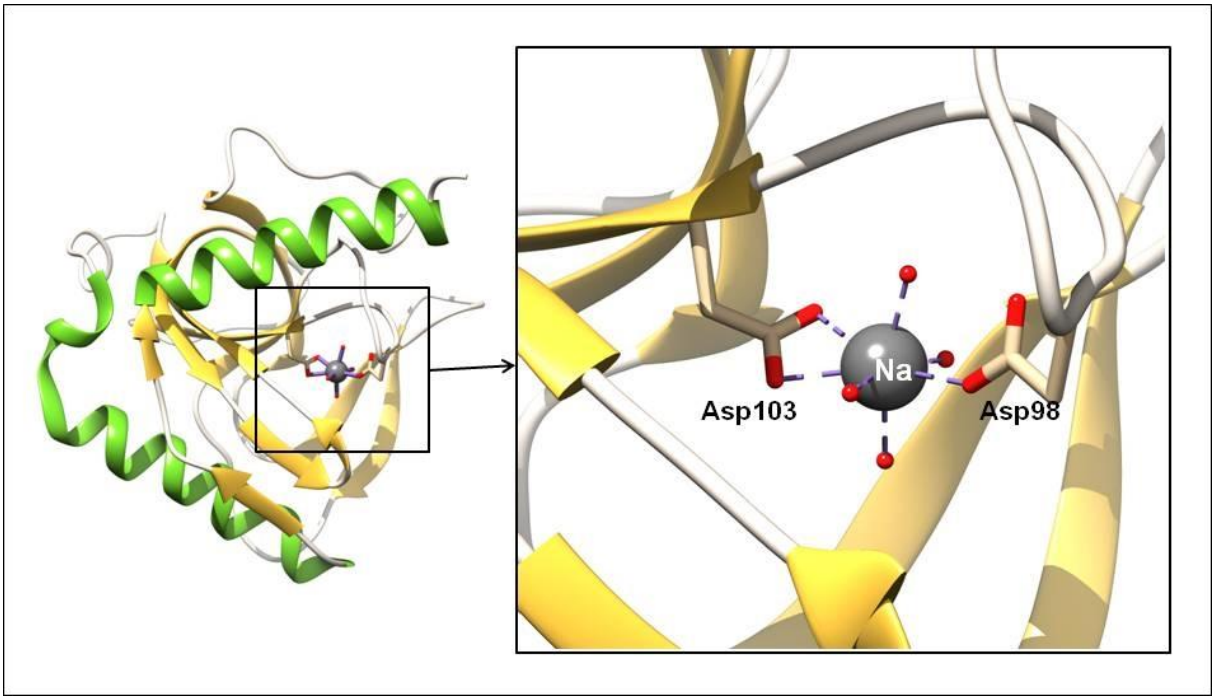


Fig. 11. Schematic representation of metal binding site of *A. thaliana* PPA1 (PDB code: 4lug). The detailed information about metal coordination (distances and angles) are summarized in Table 3.

3.6. N-terminus analysis

In resolved structure (PDB code: 4lug), unexpected lack of N-terminal 32 amino acids is observed. After the structure has been determined, a protein sample was submitted to N-terminal sequencing by chemical degradation to identify cleavage site. Unfortunately sequence analysis of the same crystal used for solving the structure was impossible. Thus, for N-terminus sequencing, three AtPPA1 protein crystals grown in the same condition as that crystal used earlier for solving structure were picked, dissolved in buffer, transferred to the PVDF membrane and sent for sequencing (for details see Materials and Methods, section 4.2.3.1). Interestingly, the analysis returned three variants of N-terminal sequences with cleavage sites: at Leu23, Ser25 or Leu26 (Fig. 12B and C). They differ in length of cleaved peptides composed of: 23, 25 or 26 residues. Obtained sequences corresponded to three crystals used for sequencing of the N-terminus.

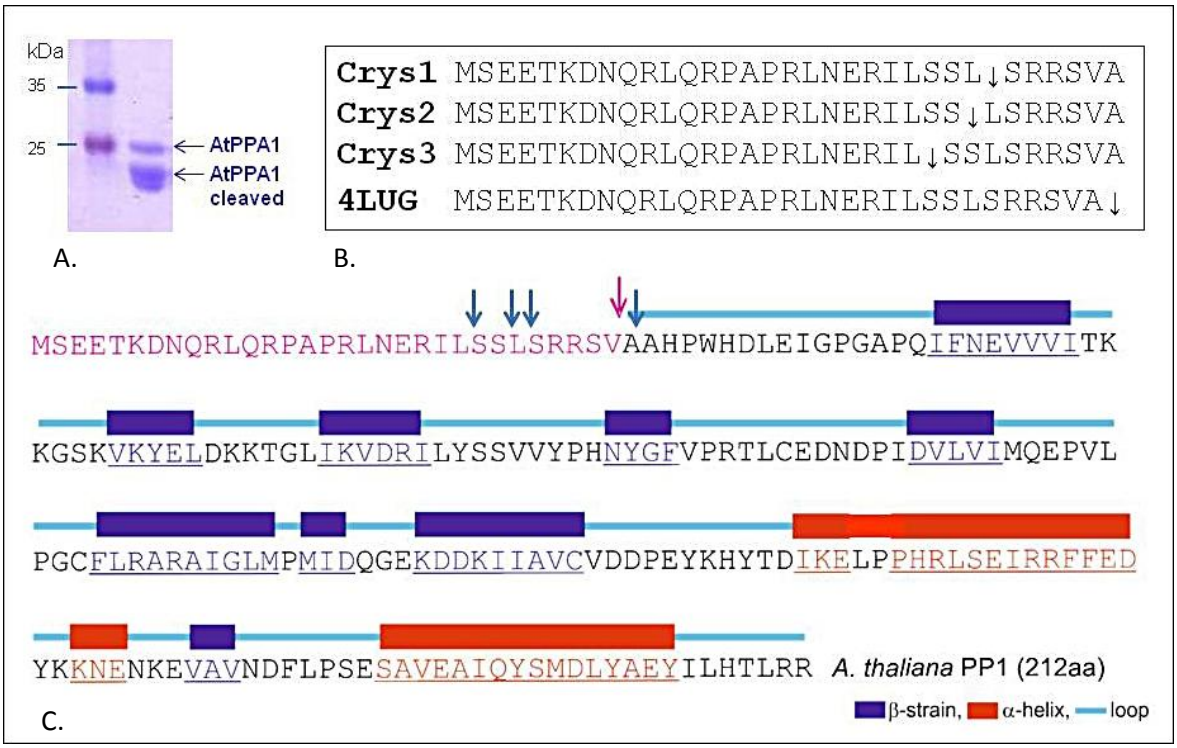


Fig. 12. (A) SDS-PAGE of AtPPA1 showing protein cleavage. (B) Results of N-terminus sequencing. The cleavage site is marked by black arrows (C). The AtPPA1 protein sequence showing the cleavage sites based on sequencing results, crystal structure (blue arrows) and predictions. Red arrow indicate the cleavage site predicted by TargetP online tool.

There is no ambiguity that the autocatalytic cleavage is stepwise. It was confirmed also by electrophoresis (SDS-PAGE) and mass spectrometry. The two bands from gel were cut and sent for MS analyses (Fig. 12A). The analyses of peptides derived from upper band using Mascot server identified peptide covering the region 7-14 N-terminal amino acids.

Truncated protein (the lower band from the gel, Fig. 12A) lacking of this peptide and the identified N-terminal fragment covered residues 30-57. It seems that the proteolysis stops at Ala32 residue and does not go further because on SDS-PAGE the cleaved protein appears at fixed position. Moreover, this missed fragment might be highly flexible and thus it becomes prone to proteolysis. This conclusion was drawn based on Navarro-De la Sancha *et al.* (2007) report, where CD spectrum analysis showed that the first 36 amino acid residues of AtPPA1 lack secondary structure and represent random coil [77]. Interestingly, i*n vitro* self-cleavage is observed after long term storage or during protein crystallisation.

I also noticed that the crystal growth started about 2 weeks after setting crystallization when proteolytic cleavage of the N-terminal peptide was completed. It might be possible that the protein require proteolysis to form crystals because the disordered N-terminal region may probably disturb formation of regular crystal lattice. On the other hand, presence of this N-terminal peptide do not interfere with trimer formation.

For structural studies, I prepared two active site mutants: D98N and D103N. The mutated residues were involved in Na^+ coordination in solved AtPPA1 wild type structure (PDB code: 4lug). Interestingly, I observed, that cleavage of N-terminal fragment is delayed in D98N mutant in comparison to active wild type protein. Additional tests with D103N mutant confirmed auto-proteolytic cleavage, but in case of this mutant, the rate of proteolysis was much slower. Truncated protein usually appeared after few days and the rate of cleavage was protein concentration dependant. Also, the rate of proteolysis correlated well with the rate of crystal growth. More concentrated sample gave faster cleavage. The crystals of D98N mutant grew slower and the proteolysis took more time (Fig. 13). Crystals of the D103N mutant grew even more slowly, were always very small and did not give reasonable diffraction.

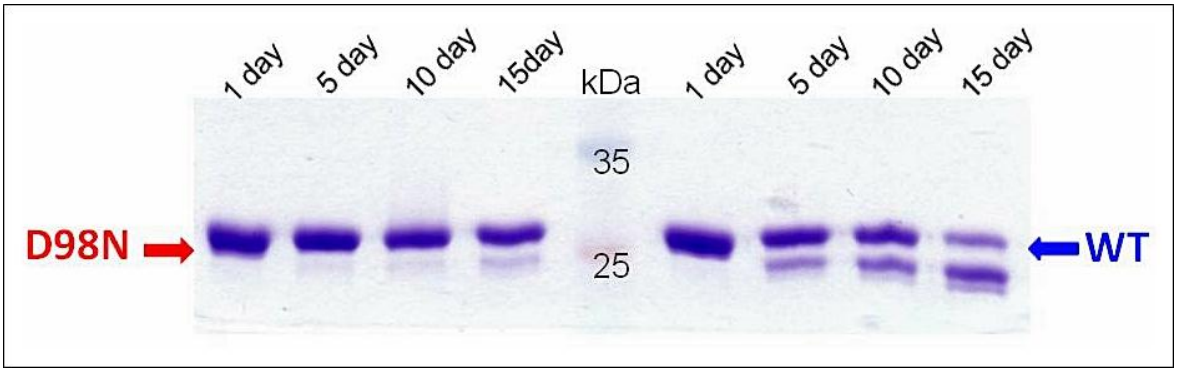


Fig. 13. Cleavage of the N-terminal fragment in D98N mutant and in WT of AtPPA1.

All those observations may suggest that the active site seems to be involved somehow in proteolysis. It is valid, that the cleavage is independent on metal cations because in storage buffer there were neither Mg^{2+} nor other divalent metal ions.

Plant PPase possesses both N-terminal and C-terminal extensions compared with the bacterial protein. Amino acid sequence alignment of *A. thaliana* and *E. coli* soluble pyrophosphatases (see Fig. 18, section 3.10) shows that plant PPase contains N-terminal extension of about 35 residues. A possible function of N-terminal sequences is targeting the proteins to plant organelles.

In plant pyrophosphatase (AtPPA1), presence of the N-terminal extension might be of the biological significance taking part in compartmentation of this protein. It was reported earlier that the N-terminal fragment of AtPPA1 corresponds to putative transit peptide of mitochondrial targeting [77]. Although comparison of known mitochondrial pre-sequences has failed to reveal significant sequence homology, a few universal properties were found. The common feature of mitochondrial targeting peptides is presence of positively charged residues (Arg in particular), lack or rare negatively charged residues and enrichment in Ser, Ala and hydrophobic residues [14, 18, 79, 97]. The amphiphilic region is important for binding to receptors in the outer mitochondrial membrane and the net positive charge may be needed during the import across the membrane [18]. Moreover, the length of the known plant mitochondrial targeting peptides varies from 13 to almost 100 but it was assumed that the transit peptides should be localized among the 40 N-terminal residues as it is an average length [14] and have the ability to form amphiphilic α-helices [18]. It was reported earlier that the N-terminal fragment of AtPPA1 corresponds to putative transit peptide of mitochondrial targeting [77].

To analyze the N-terminal protein sequence I applied bioinformatics online tools *TargetP* [16, 18] and MitoProtII [14] to check a probability of export AtPPA1 protein to mitochondria. Both methods succeed to recognize mitochondrial transit peptide and locate the potential cleavage site between Val31 and Ala32 residues. Helical structure on N-terminus is a property proposed to be signal peptide, however in studied AtPPA1 the peptide was cleaved in the crystal and I do not have structural data for this fragment. The list of properties to fulfill by mitochondrial proteins is quite long but there are a lot of exceptions, thus interpretation of predicted data is almost impossible. For sure, further studies are required to clarify the *in vivo* intracellular destination of studied AtPPA1 protein.

3.7. Oligomeric structure

It was reported earlier that recombinant AtPPA1 lack quaternary structure [77] and the same was reported for PPase from algae *Chlamydomonas reinhardti* [22]. However, my research show that AtPPA1 is a trimer in solution as confirmed by size-exclusion chromatography (Fig. 14), static light scattering measurement (Fig. 15) and is consistent with prediction of PDBe PISA web server [36, 54]. The chromatogram (Fig. 14) showed two well separated peaks: one corresponds to AtPPA1 and the second to the AtWRKY50 (transcription factor extensively studied in Part I of my dissertation used here as protein size marker). AtPPA1 (MW 24.5 kDa) showed significantly higher molecular mass comparing to the AtWRKY50 appeared as a monomer (MW 19.3 kDa). Static light scattering (SLS) measurements were performed to determine particle size of AtPPA1. The estimated molecular mass was 76.75 kDa which corresponds to AtPPA1 trimer (Fig. 15). What is important, the trimeric organisation was determined before cleavage of the N-terminal peptide. This indicates that N-terminal 33 amino acids extension, that is cleaved, do not interfere with trimer formation.

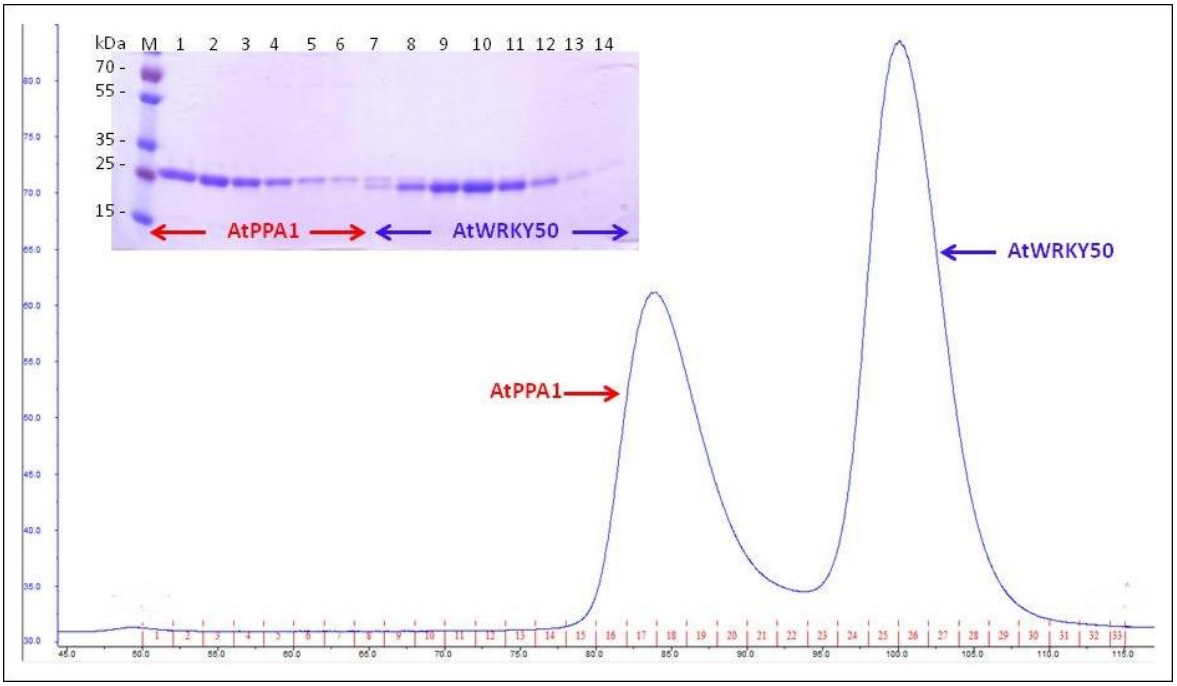


Fig. 14. Size-exclusion chromatography of a mixture containing: AtPPA1 (75 kDa as a homotrimer) and AtWRKY50 (20 kDa as a monomer) shows two distinct peaks. Fractions from both peaks were loaded on SDS–PAGE: monomer of AtPPA1, 25 kDa (peak 1); monomer of AtWRKY50, 20 kDa (peak 2). The molecular masses of the markers (lane 1) are shown in kDa. Lanes 2-15 on SDS-PAGE corresponds to fractions 17-30.

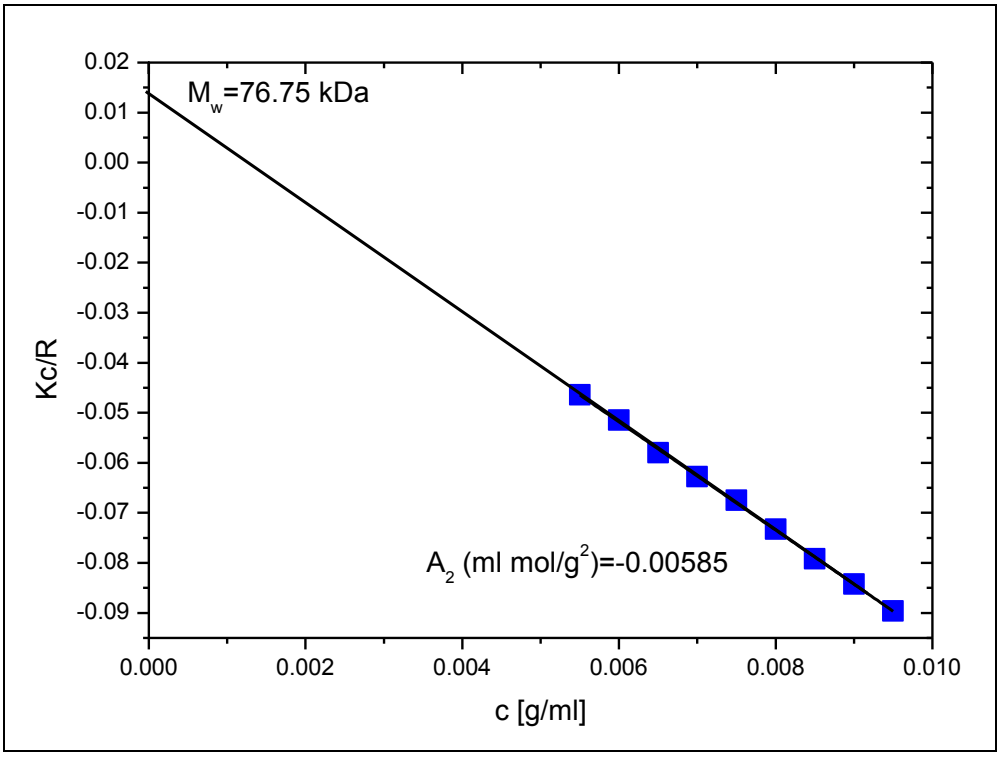


Fig.15. Debye plot of AtPPA1 obtained from SLS measurement.

PPases are oligomeric enzymes that are typically active as homohexamers in bacteria [3, 5, 108] and dimers in eukaryotic organisms [30, 32]. Plant trimeric assembly contrast to other PPases and it is an exception in this respect. Two plant PPases: AtPPA1 and AtPPA4 have been reported earlier to function as monomers [77] in contrary to results of my research showing that AtPPA1 forms trimers. Unfortunately, I did not test the AtPPA4 for oligomeric organisation. Biological assembly of plant PPases has not been studied extensively so far. It is worth mentioning, that plants posses 6 different genes encoding those enzymes and they might differ in oligomeric state, despite their amino acid sequences high conservation.

The previous reports that the AtPPA1 is a monomer [77] might be biological artefact because the protein was produced as fusion with GST and the fusion might interfere with association of monomers to oligomeric structure. However, AtPPA1 in my research was also produced as fusion protein but with NusA instead of GST and in this case the fused protein did not disturb the trimer formation. The AtPPA1 protein trimer cleaved from fusion tag (the activity of whole fusion was not determined) was biologically active and showed high PPase activity. Similarly, the monomer fused with GST was also biologically active. This observations are consistent with earlier reports about bacterial PPases, which showed that the hexameric and also dissociated trimeric or monomeric *E. coli* PPase is active [3]. However, the native hexameric structure is essential for the inactivation of *E.*

coli PPase by phosphoric acid monoesters, the PPase inhibitors and this way for *in vivo* regulation of its activity [98]. In yeast, the intramolecular contacts in dimer are relatively loose and it is also possible to isolate individual subunits that indicate the same PPi hydrolysis rate as dimer [30]. It is worth mentioned that the AtPPA1 monomer when cleaved from GST represents only traces of activity [77] that cannot be recovered in presence of 15% glycerol, 10% PEG3350 or by changing pH from 6.5 to 8.5. The activity was recovered only in presence of 50% ethylene glycol with 3 mM $MgCl_2$ and 0.1 mM EDTA but those conditions were far from physiological [77].

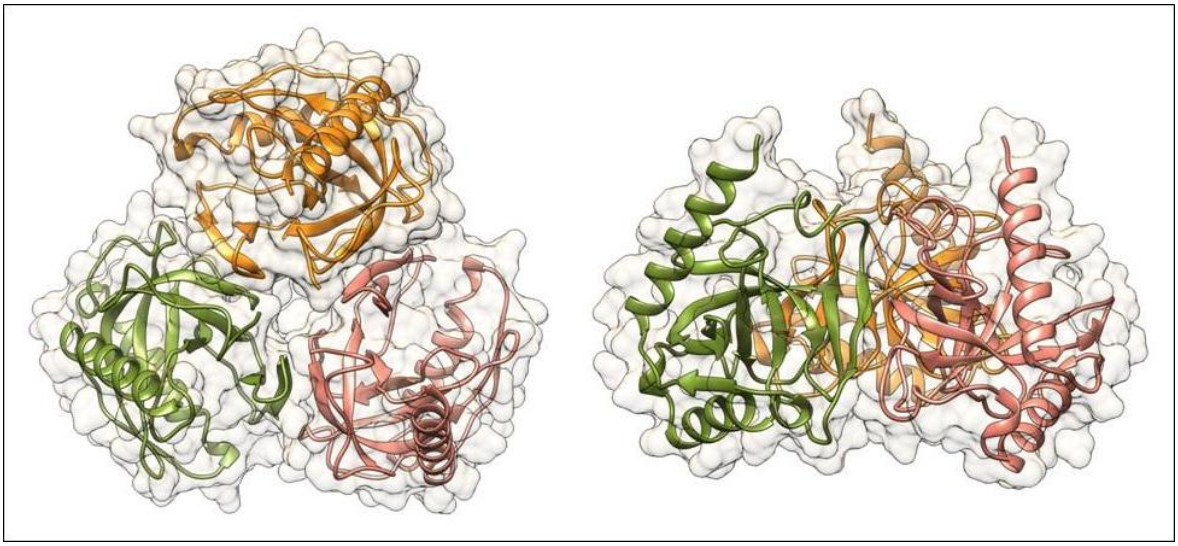


Fig. 16. AtPPA1 trimer: individual subunits are shown in different colors.

Plant AtPPA1 is a trimer (Fig. 16). In the asymmetric unit, dimer composed of chain A and B is presented but each trimer is composed of one type of monomers A or B. The crystal lattice is composed of trimers built from chains A and other trimers built from chains B arranged in layers. The contacts between monomers in the trimer are predominantly provided by hydrogen bonds (Tab. 4) and hydrophobic contacts between strands. Individual subunits of AtPPA1 are globular and rather compact. The internal interactions between monomers are stabilized by two ionic pairs: Phe117-Ile72 and Phe117-Val74 between strains β6 and β3 and by hydrophobic contacts. Other internal contacts include the region Asp144-Lys149 localized on loop between β8 and α1 with Ile77-Tyr79 localized on strain β3.

The surface areas buried upon assembly formation calculated per monomer are as follows: 1540 $Å^2$ in the *Arabidopsis* PPA1 trimer (PDB code: 4lug) and 2877 $Å^2$ in *E.coli* hexamer (PDB code: 2au6). The calculations were done using PDBePISA web server [36, 54].

Table 4. Hydrogen bonds distances (Å) between subunits in AtPPA1 trimer (PDB code: 4lug)

Residue [atom]		**Dist. [Å] Chain A + Chain A**	**Dist. [Å] Chain B + Chain B**
LEU 113 [N]	TYR 63 [OH]	3.06	3.05
ALA 33 [N]	THR 69 [O]	3.62	-
PHE 117 [N]	ILE 72 [O]	2.99	3.06
LYS 149 [NZ]	TYR 79 [OH]	3.16	3.01
GLN 109 [OE1]	LYS 60 [NZ]	3.02	3.30
PHE 117 [O]	VAL 74 [N]	2.71	2.76
ASP 144 [O]	ILE 77 [N]	3.19	3.27

Prokaryotic PPases from family I are generally homohexamers arranged as dimers of trimers. The two trimers are adjacent to each other in such a way that each monomer interacts with four other subunits. In hexamer, the upper trimer is rotated of about 30° with respect to the bottom trimer. The contacts between monomers in trimers are mainly hydrophobic interactions between β-strands while the interface between trimers are provided by the symmetry related α-helices. Another interesting trimer-trimer interaction occurs in *E. coli* PPase which require the Mg^{2+} ions for hexamer stabilization and the binding stoichiometry of Mg^{2+} ion to monomer is 0,5:1 [5]. Each Mg^{2+} ion is octahedrally coordinated to six water molecules that are hydrogen bonded to the Asn24, Ala25, Asp26 residues and the Asn24', Ala25', Asp26' from symmetric subunit [5]. These residues are not well conserved among the prokaryotic PPases and the mechanism of hexamer stabilization might be unique to *E. coli* PPase. Structural studies of PPases from other bacteria such as: *T. thermophilus* [100], *M. tuberculosis* [6] and *S. acidocaldarius* [62] did not identify interface metal ions. Besides residues engaged in Mg^{2+} binding, the trimer-trimer interface in *E. coli* involves helix α1 and triad of His136, His140, Asp143 residues. Those residues are well conserved in other prokaryotic sequences [96]. However, in *T. termophilus* the contacts between trimers are provided by different side chains: Gln130, His134, Thr138, and Leu142, located on one side of the helix. Moreover, the termophilic *T. termophilus* and archebacteria *S. acidocaldarius* appear to form tighter assemblies than

E. coli and *H. pylori* PPases. The buried intramolecular surface is maximized in the hyperthermophilic organisms and suggest that packing is related to their termostability [62]. Besides that, the oligomerization is known to be essential for PPases stability [3]. The hexameric assemblies were found also in the other prokaryotes - Cyanobacteria but interestingly, the PPase from the purple nonsulfur photobacterium *Rhodopseudomonas viridis* exhibits larger molecular mass suggesting dodecamer formation probably due to nonconservative substitutions in most residues involved in the formation of the dimer of trimers in comparison with *E. coli* enzyme [23]. Although these require further studies.

Animal and fungal PPases form homodimers. In *S. cerevisiae* the dimer interface is completely different. It involves aromatic and positively charged residues: Arg51, Trp52, His87 and Trp279. This dimer interfacing residues are well conserved in other eukaryotic PPases except plants. Only His87 in other eukaryotes is replaced by Lys residue.

Summarizing, PPases from various phylogenetic groups represents significant differences in oligomeric organisation associated with conservation of the interfacing residues and sequence similarities

All the non-conservative substitutions and the deletions influence the final quaternary structure and the major differences between compared structures are observed in those regions where sequence alignment results in deletions or insertions. The subunit interface residues of prokaryotic and animal or fungal PPases are not conserved in plant PPases. Plant and algal PPase orthologues show substitutions of the residues: Asn24, His136, His140, Asp143 responsible for hexamer stabilization in *E. coli*. Plant PPase also lacking Arg51, Trp52, His87, Trp279 residues responsible for dimer stabilization in yeast. This might explain unusual trimeric structure of plant AtPPA1.

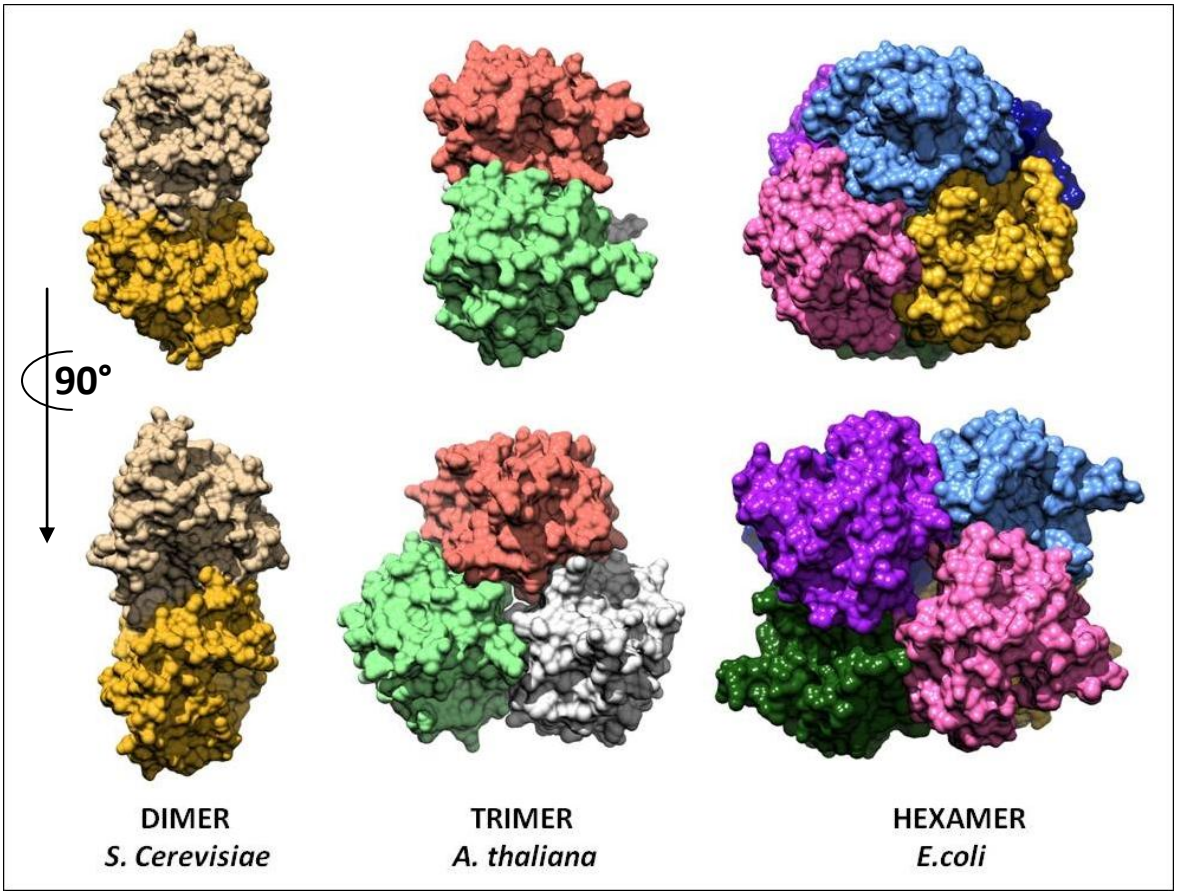


Fig. 17. Oligomeric organization of PPases from different organisms: dimer from *S. cerevisiae* (PDB code: 1E6A), trimer from *A. thaliana* (PDB code: 4lug) and hexamer from E. coli (PDB code:2au6).

3.8. Enzymatic activity of PPases

The hydrolytic activity of recombinant AtPPA1 toward inorganic pyrophosphate was tested to check if obtained protein is biologically active. Activity was confirmed by enzymatic test according to procedure described in Materials and method (section 4.2.4). *Arabidopsis* inorganic pyrophosphatase activity was highest in presence of Mg^{2+} and detectable in presence of Mn^{2+}. These results are in good agreement with previous reports that describe AtPPA1 activity [77]. The protein mutants with mutated Asp (D98N and D103N) that take part in metal cation coordination were inactive. The following divalent cations were also tested: Fe^{2+}, Co^{2+}, Zn^{2+}, Ca^{2+} and Cu^{2+} but the activity was not detectable. The results of the recombinant AtPPA1 activity test are summarized in Table 5.

Table 5. Activity of recombinant AtPPA1-WT and mutated variants (**N-** not determined, **#** the inhibitory effect was tested in presence of Mg^{2+}).

Divalent cation	***WT***	***D98N***	***D103N***
Mg^{2+}	100%	0	0
Mn^{2+}	~10%	0	0
Ca^{2+}	inhibitor	N	N
F^{-}#	inhibitor	N	N

It is worth mentioned that several divalent metal ions, including Mn^{2+}, Fe^{2+}, Zn^{2+}, Cu^{2+} and Co^{2+} support various PPases activity *in vitro* [26, 56], although Mg^{2+} is their physiological cofactor. The enzyme activity estimated with the alternative cofactors is lower than that with Mg^{2+} [56, 106], which is largely due to changes in the product binding and release. The exception are photosynthetic bacteria that may more efficiently replace Mg^{2+} or Mn^{2+} with other divalent cations (see Table 6) [23]. Cyanobacterial PPases (e.g. *Anabena* sp.) (Table 6) are all strictly Mg^{2+}-dependent enzymes. On the contrary, a variability of cation dependence was found among anoxygenic bacteria PPases (e.g. *Rhodopseudomonas viridis*). They show greater heterogeneity in respect to metal cation dependence, as the PPases of many of these bacteria can very efficiently in use divalent cations as cofactors (e.g. Zn^{2+}, Co^{2+} and Fe^{2+}; Table 6). In *Rhodopseudomonas viridis* PPase, the efficiency of those cations is higher than Mg^{2+} but at the same level at the presence of Cu^{2+} and Mn^{2+}, if compared to Mg^{2+} [23]. The high variability in cation dependence was found for the anoxygenic bacteria. It may reflect adaptations to specific metabolic scenarios.

Table 6. Review of divalent cations requirement of soluble inorganic pyrophosphatases activity from different organisms of Family I. The level 100% is assigned to the activity determined with Mg^{2+}.

Divalent cation	*Arabidopsis thaliana1*	*Chlamydomonas reinhardti**	*Escherichia coli*	*Rhodopseudomonas viridis***	*Anabena* sp.****	*Saccharomyces cerevisiae*
Mg^{2+}	100%	100%	100%	100%	100%	100%
Mn^{2+}	<10%	20-30%	10-13%	97%	3%	3%
Fe^{2+}	0	20-30%	4%	134%?	1%	N
Co^{2+}	0	N	10%	125%	3%	1%/20%
Zn^{2+}	0	20-30%	10%	140%	1%	10%
Cu^{2+}	0	20-30%	3%	94%	2%	N
Ca^{2+}	inhibitor	inhibitor	inhibitor	N	N	inhibitor
F^-	inhibitor	N	inhibitor	N	N	inhibitor
No cation	0	0	0	0	0	0
Ref.	[77]	[22]	[23, 32]	[23]	[23]	[26, 111]

Legend:
0 - no activity; N - not determined
**Chlamydomonas reinhardti* - microalga
***Rhodopseudomonas viridis* - purple nonsulfur photobacteria
**** *Anabena* sp. - cyanobacteria

Table 6 summarizes experimental data from different studies. The experimental procedures and reaction conditions were different, thus this is review of the approximate results showing the diversity in ion dependence. Reactivity rates with other cations were compared with the activity with Mg^{2+} that was assumed as 100% in terms of released Pi.

This comparison shows that metal cation requirement is not strictly conserved even in phylogenetic groups despite the conservation of the enzyme active site and the whole topology of PPases from different organisms.

The kinetic parameters for AtPPA1 were not calculated, because comprehensive biochemical characteristics including activity measurements, substrate specificity and inhibitory effects for this enzyme are available in recent publication [77]. The following kinetic parameters characterizing the enzymatic activity of recombinant AtPPA1 towards inorganic pyrophosphate hydrolysis have been reported: $K_{M[MgPPi]} = 0.069 \pm 0.015$ µM and $k_{CAT} = 12.1 \pm 1.0$ 1/s [77]. Single polypeptide chain posses only one pyrophosphate binding cavity. Interestingly in the publication mentioned earlier [77] the saturation kinetic curves were sigmoidal, (based on Hill number) and indicated cooperative character of catalysis, although the reported protein was a monomer. In that publication, the recombinant protein was produced as fusion with GST. The fusion protein was then cleaved and AtPPA1 was

estimated to be a monomer. Researchers interpretation of cooperativity was a low enzyme stability due to the following observations: loss of activity when diluted, after freezing/thawing and during size exclusion chromatography or at low PPi and Mg^{2+} concentrations. Results of my experiments revealed that AtPPA1 is a trimer. Navarro *et. al*. (2007) estimated the protein size using size exclusion chromatography. Since this method is not very precise, the estimation that AtPPA1 is a monomer was wrong. The protein might also form oligomeric structure during storage. It might explain sigmoidal kinetic curves.

The influence of fluoride, a potent inhibitor of PPases, was examined for recombinant AtPPA1 [12, 40, 80]. The enzyme was very sensitive to F^- anion and the inhibitory effect was observed with addition of 0.5 mM NaF. The pyrophosphatase activity was below detection limit under test conditions when added 0.5 mM NaF and this result is in good agreement with earlier reports for bacterial PPases [12, 40]. The AtPPAse activity was also strongly inhibited by Ca^{2+} cations.

It was also reported that AtPPA1 behave as specific pyrophosphatase and did not catalyze detectable release of phosphate from compounds such as: ADP, $NADP^+$, NAD^+, NADH, NADPH, or phosphoribosyl pyrophosphate in presence or without Mg^{2+} [77]. AtPPA1 was inactive towards glycerol-3-phosphate, glucose-6-phosphate, p-nitrophenylphosphate as substrates, neither in the absence nor in the presence of Mg^{2+} [77]. This results are opposed to some previous reports on native plant pyrophosphatases purified from plant extracts, where some Mg^{2+}independent phosphatase activity was detected [70]. However, there is other interesting phenomenon. Family I soluble pyrophosphatases (PPases) exhibit appreciable ATPase activity in the presence of a number of transition metal ions. Unfortunately it was not tested for AtPPA1. PPases from *Saccharomyces cerevisiae*, *Escherichia coli* and rat liver have ability to hydrolyze ATP in presence of Co^{2+}, Zn^{2+} or Mn^{2+} cations. In the presence of its physiological activator Mg^{++}, PPase displays nearly absolute substrate specificity. Besides PPi, also polyphosphates, such as tri- and tetra-polyphosphates, are converted, but the rate of hydrolysis is only 1/60 of that observed for PPi as a substrate [35, 42]. Interestingly, this specificity is lost when transition metal ions such as Co^{2+}, Zn^{2+} and Mn^{2+} are used as cofactors. In the presence of these metal ions, family I PPase displays high catalytic activity against PPi together with the ability to hydrolyse organic tri- and diphosphates, such as ATP and ADP [57, 91]. The inability of Mg^{2+} cations - the physiological cofactors to support the ATPase activity of these enzymes has a physiological significance, since it prevents useless ATP hydrolysis [110].

3.9. Comparison of AtPPA1 with other pyrophosphatases

Arabidopsis PPase subunit (AtPPA1) have an estimated molecular mass of 24.48 kDa, whereas prokaryotic PPases - approx. 20 kDa and eukaryotic PPases - approx. 34 kDa. Thus plant PPase is closer to prokaryotic than to eukaryotic soluble PPases. The sequence identity between the plant and the other pyrophosphatases from both prokaryotes and eukaryotes is below 42% (Tab. 7). Figure 18 show the alignment of the sequences of seven prokaryotic and eukaryotic soluble pyrophosphatases from: *S. acidocaldarius, E. coli, H. pylori, Z. mays, H. sapiens, S. cerevisiae, D. melanogaster* with *A. thaliana* PPA1. Except the N-terminal fragment of about 30 residues, that is lacking in bacteria, the AtPPA1 show significant sequence similarity rather to bacterial pyrophosphatases than to other eukaryotic PPases (Tab. 7).

Table 7. Sequence and structure comparison of AtPPA1 with other prokaryotic and eukaryotic PPases (RMSD of Cα atoms and Q-score were calculated using PDBeFold online server).

Organism	Sequence identity (%)*	PDB	RMSD of Cα atoms (Å)	Number of aligned residues	Q-score**
Sulfolobus acidocaldarius	39.31	1qez	1.06	164	0.81
Escherichia coli	28.41	1i40	1.10	167	0.80
Helicobacter pylori	41.04	1ygz	1.29	161	0.73
Zea mays	81.60	-	-	-	-
Homo sapiens	25.00	-	-	-	-
Saccharomyces cerevisiae	23.11	2ihp	1.34	170	0.50
Drosophila melanogaster	21.70	-	-	-	-

*sequence identities were calculated for full length proteins
**Q - score represents the quality function of Cα-alignment. It reduces the effect of RMSD - N_{algn} balance on the estimation of alignments (N_{res1} and N_{res2} stand for the number of residues in the aligned proteins, and empirical parameter R_0 is set to 3 Å: $Q = (N_{algn}*N_{algn})/[(1+(RMSD/R_0)^2)*N_{res1}*N_{res2}]$

All archaeal and bacterial PPase sequences are shorter with a total length of about 170–190 amino acid residues than those from eukaryotes consisting of ca. 210–350 residues. The differences in polypeptide length reflect to a presence of few gaps within plant and bacterial sequences and the presence of an extensions at N- and C-terminus in most plant, animal and fungal sequences. Figure 18 shows that both *Z. mays* and *A. thali*ana PPases have two deletions between 69-70 and 97-98 residues (acc. to *A. thaliana* numbering) with respect to eukaryotic PPases.

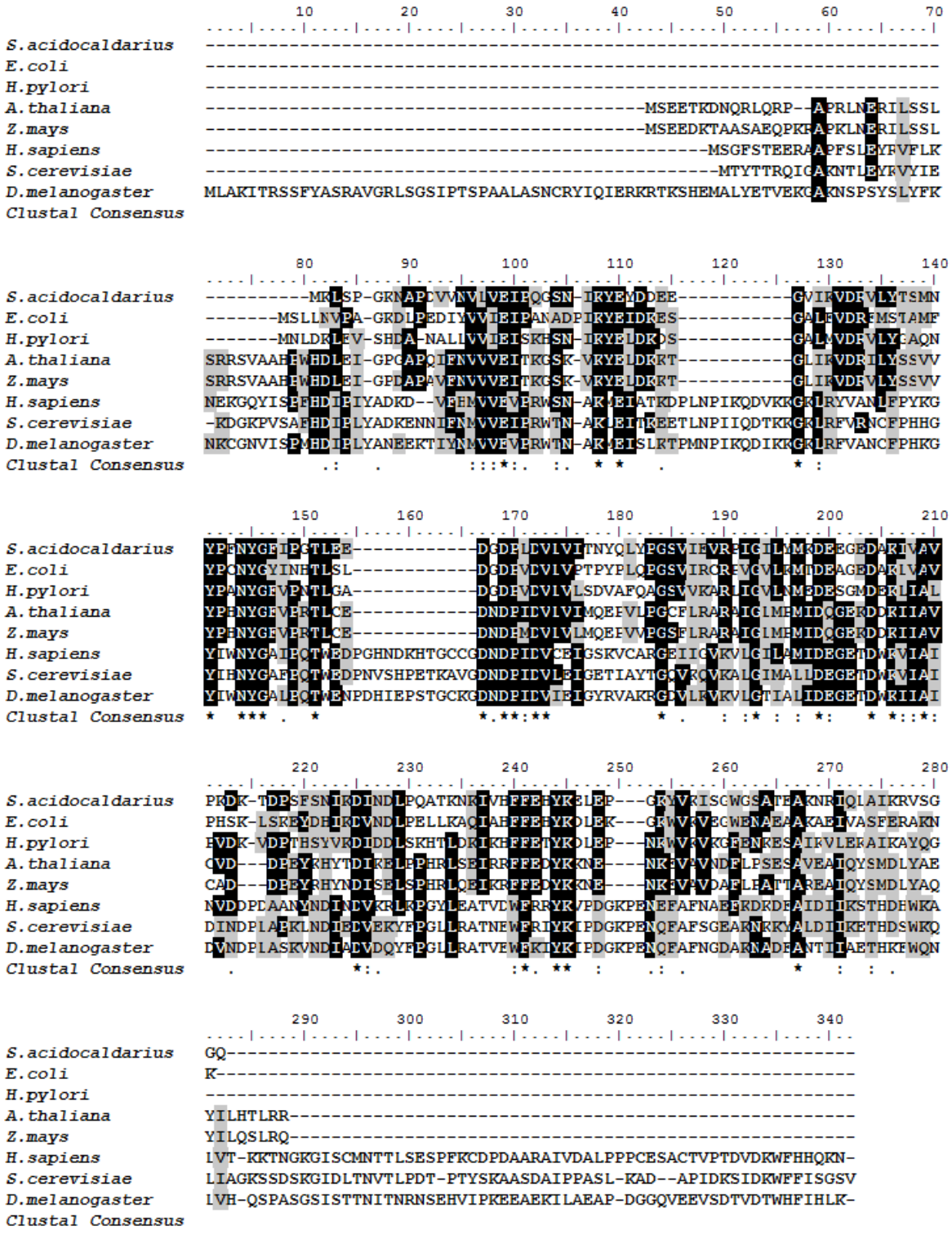


Fig. 18. Multiple sequence alignment of selected PPases from*: S. acidocaldarius* [P50308], *E. coli* [P0A7A9], *H. pylori* [E8QVW7], *A. thaliana* [Q93V56], *Z. mays* [O48556], *H. sapiens* [Q15181], *S. cerevisiae* [P00817] and *D. Melanogaster* [O77460]. (UniProt identifier is shown in brackets) The highly conserved residues are marked with asterisk. The level of conservation is sisualized by level of darkness of the letters background. The alignment was calculated in ClustalW2 [60] and the figure was prepared using the Bioedit editor [25]

It is worth to notice that all prokaryotic soluble PPases possess these deletions. There is also one small 3-4 residues gap between 144-145 position (acc. to *A. thaliana* numbering) that is observed in plant and bacterial PPases. The 3 amino acids gap between 176-177 position is unique for plants. Nevertheless, the main difference between prokaryotic and eukaryotic soluble PPases is due to the fact that the eukaryotic PPases have C-terminal extensions (Fig. 18). This additional segment vary in length from several residues in all plant PPases to about 30 residues in some other eukaryotic enzymes. PPase from *Drosophila melanogaster* have also about 40 residues long extension at N-terminus comparing to the *A. thaliana, Z. mays, Homo sapiens* and *Saccharomyces cerevisiae* PPases. Figure 18 presents only eight sequences but underline the heterogeneity of the soluble PPases from different phylogenetic groups of organisms. Alignment identified 25 conserved residues and 15 of them are part of active site, according to X-ray crystallographic studies in *S. cerevisiae* and *E. coli*. These residues are scattered over the conserved central sequence. In *Arabidopsis*, the first conserved residue occupies position 54 and the last one occupies position 191. A sequence alignment of the *Arabidopsis* polypeptide indicates high sequence identity with the *Zea mays* polypeptide - above 80%. The highest differences between plant and other PPases are present in the C- and N-terminal regions. As I mentioned in previous chapters, a possible function of N-terminal sequences is the targeting of the polypeptides to organelles.

Summing up, the mature protein starts at residue 33, from which the homology with bacterial proteins begin. Taking into account the autoproteolytic activity and bioinformatics predictions, putative cleavage site of AtPPA1 can be found within the SRRSVAAH N-terminal protein fragment. This sequence is also presented in *Z. mays* PPase. The features of the N-terminal extension such as enrichment in Ser residues and lack both positively charged and Gly residues also might suggest that these predictions are correct. Although, based on these observations, it is thus difficult to draw unequivocal conclusions regarding the capacity of the N-terminal extension to target the protein. Considering an increasing number of exceptions to the rules which have been already described [4], no firm conclusion can be inferred from this sequence analysis.

The heterogeneity of PPases is consistent with the phylogenetic analysis [23]. Two well defined groups cluster on the phylogenetic tree: eukaryotic PPases and the prokaryotic (bacterial and archaeal) PPases together with prokaryotic-type homologues of photosynthetic eukaryotes [23]. Interestingly none of the animal PPase sequences has the two prokaryotic-type deletions [23]. As pointed out by the phylogenetic analysis [23], the

cloned *Arabidopsis* soluble PPA1 belongs to the prokaryotic cluster more precisely than to eukaryotic soluble PPases. The similarity of AtPPA1 and other plant PPases to bacterial PPases is surprising taking into account that plants are eukaryotes. The overall three dimensional structure of PPases is well conserved with the exception of loop regions (Fig. 19). The structure of AtPPA1 is similar to other PPases from Family I deposited in PDB to date. The AtPPA1 monomer aligns well with other PPases with RMSD 1-1.4 for Cα atoms. Structural differences are mostly due to variations of the N- and C-termini and are observed also in the loops between β2 and β3, between β3 and β4 and between β8 strain and helix α1. Those regions correspond to deletions or insertions shown on sequence alignment. Phylogenic tree of Family I PPases based on sequence alignment shows that plant AtPPA1 [22] and AtPPA3 (89% identity with AtPPA1) [96] are evolutionary closer to the *H. pylori* and *S. acidocaldarius* than to *E. coli* PPase. The differences are mostly due to variations in sequences presented on alignment as gaps or insertions (Fig. 18). However, considering the structural homology, the overall fold of the PPases monomer appear to be very well conserved. The differences are observed only in packing (tight or loose), N- or C terminal extensions and oligomerisation.

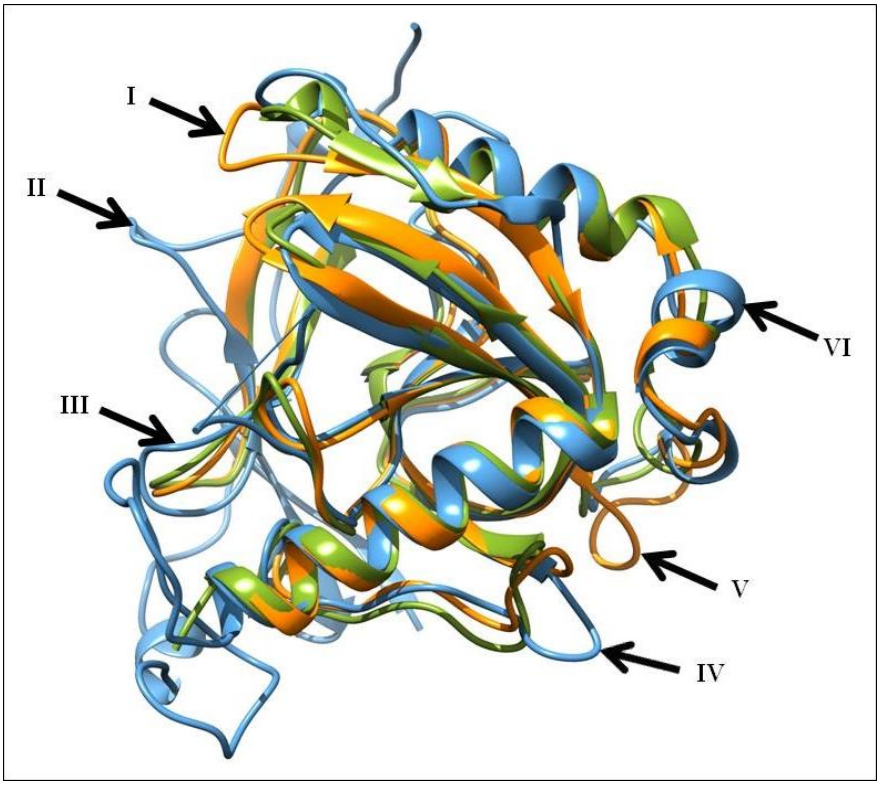


Fig. 19. Comparison of the three-dimensional structures of the Family I soluble inorganic pyrophosphatases from *A. thaliana* (green, PDB:4lug), *S. cerevisiae* (blue, PDB: 2ihp) and *E. coli* (orange, PDB:1i40) representing Family I. They share very similar overall structures, especially the central β-barrel and α-helices. The RMSD of Cα atoms were as follows: 1i40 vs 4lug 1.10 for 167 aa; 2ihp vs 4lug 1.34 for 170 aa. These values for Cα atoms were calculated using PDPeFold server. The arrows indicate the variant loops between: α1- β9 (I), β3-β4 (II), β2-β3 (III), N-terminal loop (IV) and β8-α1 (V-VI).

4. Materials and methods

4.1. Materials

4.1.1. Materials used in the experiments

Other materials and methods used for basics molecular biology methods and protein purification are summarized in Materials and methods section in Part I.

Category	Material	Producer
Crystallization	PACT premier HT-96	Molecular Dimensions
	JCSG-plus HT-96	Molecular Dimensions
Buffer ingredients	Sodium phosphate Monobasic	Sigma-Aldrich
	Sodium phosphate Dibasic	Sigma-Aldrich
	Sodium pyrophosphate	Sigma-Aldrich
	Ammonium molybdate	Sigma-Aldrich
	Iron sulphate	Fluka
Enzymes	DpnI restriction enzyme	NEB
Vector	pMCSG48	Midwest Center for Structural Genomics
Bacterial strains	TOP10	Invitrogen
	BL21Magic	Midwest Center for Structural Genomics
Other	Whatman paper 3mm	Whatmann
	PVDF membrane 0,22µm	Millipore

4.1.2.Oligonucleotodes

Name	Primer sequence (5'-3')	Purpose
AtPPA1_Fw	TACTTCCAATCCAATGCCATGAGTGAAGAAACTAAAGATAACCAGAGG	LIC
AtPPA1_Re	TTATCCACTTCCAATGTTATCAACGCCTCAGGGTGTGGAG	
PPA1_D98N_FW	CGC ACA TTG TGT GAA AAC AAT GAC	PIPE
PPA1_D98N_RE	ATC AAT GGG GTC ATT GTT TTC ACA CA	
PPA1_D103N_FW	GAC AAT GAC CCC ATT AAT GTC TTA	
PPA1_D103N_RE	GAT GAC TAA GAC ATT AAT GGG GTC	
TOPO_T7F	TAATACGACTCACTATAGGG	sequencing
Oligo (dT)	TTTTTTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTCTTTTTTTTTT TTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTT	Reverse transcription

4.1.3. Buffers:

Cell lysis buffer	
Tris-HCl pH 7.5	50 mM
NaCl	500 mM
Imidazole	20 mM
TCEP	1-2 mM
lizozyme	100 μg/ml

HisTrap binding buffer	
Tris-HCl pH 7.5	50 mM
NaCl	500 mM
Imidazole	20 mM
TCEP	1-2 mM

HisTrap elution buffer	
Tris-HCl pH 7.5	50 mM
NaCl	500 mM
Imidazole	300 mM
TCEP	1-2 mM

Dialysis buffer	
Tris-HCl pH 7.5	50 mM
NaCl	500 mM
TCEP	1-2 mM

Gel filtration buffer	
Tris-HCl pH 7.5	20 mM
NaCl	200 mM
TCEP	1-2 mM

Taussky-Shorr Reagent *(100 ml)*	
10% Ammonium Molybdate in 10 N H_2SO_4	10 ml
Ferrous Sulfate, heptahydrate	5 g

4.2. Methods

The coding sequence of inorganic pyrophosphatase AtPPA1 was obtained from fresh *A. thaliana* plant material by isolation of total RNA followed by reverse transcription and PCR amplification with specific primers suitable for *ppa1* sequences. The sequence agreement was confirmed by comparison with the data available from TAIR database.

4.2.1. Molecular biology methods

4.2.1.1. Cloning, expression and purification of AtPPA1

Arabidopsis thaliana plants, ecotype Columbia (Col-0) were grown in a growth chamber at 19–20 °C in compost soil over a 16 h photoperiod. The plant material was harvested, immediately frozen in liquid nitrogen and stored at -80°C. Total RNA was isolated from 6-week old plants leaves using RNeasy Plant Mini Kit (Qiagen). First strand of total cDNA was generated using SuperScript® III Reverse Transcriptase (Invitrogen™) and polyA primer (the detailed procedure was described in section 5.2.1 in PartI). The full length sequence encoding AtPPA1 (locus At1g01050.1, NCBI Accession number CP002684, UniProt Knowledgebase code Q93V56) was amplified by polymerase chain reaction (PCR) with the following pair of primers adapted for ligation independent cloning (LIC):
Forward:
$^{5'}$TACTTCCAATCCAATGCCATGAGTGAAGAAACTAAAGATAACCAGAGG$_{3'}$
Reverse:
$^{5'}$TTATCCACTTCCAATGTTATCAACGCCTCAGGGTGTGGAG$_{3'}$
using total cDNA as template.
The PCR product was cloned into pMCSGT48 expression vector (from the Midwest Center for Structural Genomics, Argonne, IL, USA) witch additional N-terminal 8xHis-NusA fusion tags. The pMCSG48–*Atppa1* construct was obtained by ligation-independent cloning (see section 5.2.1.4.2 in Part I) [48]. The resulting recombinant plasmid was sequenced to confirme correctness of the insert. Proper plasmid was used to transform *E.coli* BL21Magic strain. The recombinant protein was expressed as 8×His-NusA fusion protein. The *E. coli* strain containing the expression plasmid was grown at 37 °C in 1 litre of LB medium containing 25 µg/ml kanamycin and 100 µg/ml carbenicillin until the OD_{600} of the cultures reached 0.8. Then the temperature was decreased to 18 °C and protein

expression was induced by addition of the isopropyl-β-d-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. After 16 has, the cells were collected by centrifugation and the cell pellet was resuspended in binding buffer (20 mM imidazole, 500 mM NaCl, 50 mM Tris–HCl pH 7.5, 1 mM tris(2-carboxyethyl) phosphine). The cells were frozen and stored at -20°C. After freezing and thawing 20 µg/ml lysozyme was added. The lysed bacteria were sonicated on ice using 4-min bursts with appropriate intervals for cooling. After sonication, 1 µl Benzonase (Sigma) was added to get rid of DNA. To remove cell debris, lysate was centrifuged at 17000 x g for 30 min at 4 °C. The supernatant was loaded onto a column packed with 7 ml of Ni^{2+}-chelating Sepharose HP resin (GE Healthcare, Pittsburgh, PA, USA), connected to Vac-Man (Promega, Madison, WI, USA) and the chromatographic process was accelerated with a vacuum pump and the column was washed five times with 30 ml if binding buffer to remove non-specifically bound proteins. The protein of interest was eluted with buffer containing 300 mM M imidazole in 500 mM NaCl, 50 mM Tris–HCl pH 7.5 and 1mM tris(2-carboxyethyl) phosphine. The 8xHis–NusA tag was cleaved with TEV (tobacco etch virus) protease overnight at 4°C and the excess of imidazole was simultaneously removed by dialysis. The solution containing cleaved protein was mixed with Ni^{2+}-Sepharose HP resin to bind the tags and TEV protease. The flow-through was collected and concentrated to 5 ml. The sample was applied onto HiLoad Superdex 200 16/60 gel filtration column (GE Healthcare) pre-equilibrated with buffer composed of 50 mM Tris/HCl, pH 8.0, 50 mM NaCl and 1 mM tris(2-carboxyethyl) phosphine. The size-exclusion chromatography step yielded a homogenous protein fraction and the peak corresponded to a molecular mass of 76 kDa. All homogenous protein fractions were pooled and concentrated to 7.5 mg/ml using Amicon Ultra 10 filters (Millipore). The protein concentration was estimated using the method of Bradford [9] with bovine serum albumin as a standard or UV absorbance at 280 nm. The sample purity was monitored using gel electrophoresis in 15% polyacrylamide gel in denaturing conditions [58]. Pure protein sample was flash-frozen in liquid nitrogen as 100 µl aliquots and stored at -80°C. The samples were thawed, if needed, dialysed or diluted and used for crystallisation.

4.2.1.2. Generation of D98N and D103N mutants of AtPPA1

Site-directed mutagenesis of AtPPA1 was performed using the polymerase incomplete primer extension technique (PIPE) [50]. In this method the mutation is introduced into primer sequences that overlap each other. Two active site mutants were generated. The

aspartic acid codons were changed to asparagine at positions 98 and 103. The pMCSG48–*Atppa1* plasmid carrying the original protein sequence was used as template for PCR reaction. To eliminate the transformation background, template DNA was digested by DpnI according to the manufacturer protocol. The reaction products were used for transformation of the competent *E. coli* cells.

The mutations (in underlined codons) were introduced using specific primers:

PPA1_D98N_FW: CGC ACA TTG TGT GAA AAC AAT GAC

PPA1_D98N_RE: ATC AAT GGG GTC ATT GTT TTC ACA CA

PPA1_D103N_FW: GAC AAT GAC CCC ATT AAT GTC TTA

PPA1_D103N_RE: GAT GAC TAA GAC ATT AAT GGG GTC

The resulting vectors were verified by DNA sequencing. The mutated AtPPA1 proteins were overexpressed and purified as described earlier for the wild-type protein.

4.2.2. Protein X-ray crystallography

X-ray crystallography is a vital method for studying the structure of proteins and other macromolecules. Knowledge found from X-ray diffraction is an indispensable tool in a wide range of research fields from basic biochemistry and biophysics to pharmacy and biotechnology. Structural studies of proteins are very complex and include several steps. Each step may be a bottleneck and limit the next one; that is why this methodology is time consuming and expensive. Crystallographic studies starts from protein crystallization. However crystallization require large amounts of protein of high quality, thus previous steps including protein production and purification are crucial for entire procedure. Very often, obtaining the protein is a limiting factor for further stages of research. Obtaining the particular proteins, such transcription factors, storage proteins, membrane proteins or toxic ones in suitable quantity (usually milligrams) and purity (>90%) is very difficult and often impossible. Different proteins behave in unpredictable ways. They might be difficult to overexpress, unstable and prone to aggregation or hard to purify. The purity and homogeneity can be achieved by appropriate methods of protein purification and handling. The state of the protein sample can be evaluated by biochemical and physical methods like SDS-PAGE (purity), size exclusion chromatography (aggregation) or DLS (polydispersity, heterogeneity).

The crystallographic studies include several main steps:

I. Experimental

- crystallization
- diffraction data collection

II. Computational

- data processing and structure solution
- model building, refinement, structure validation and deposition

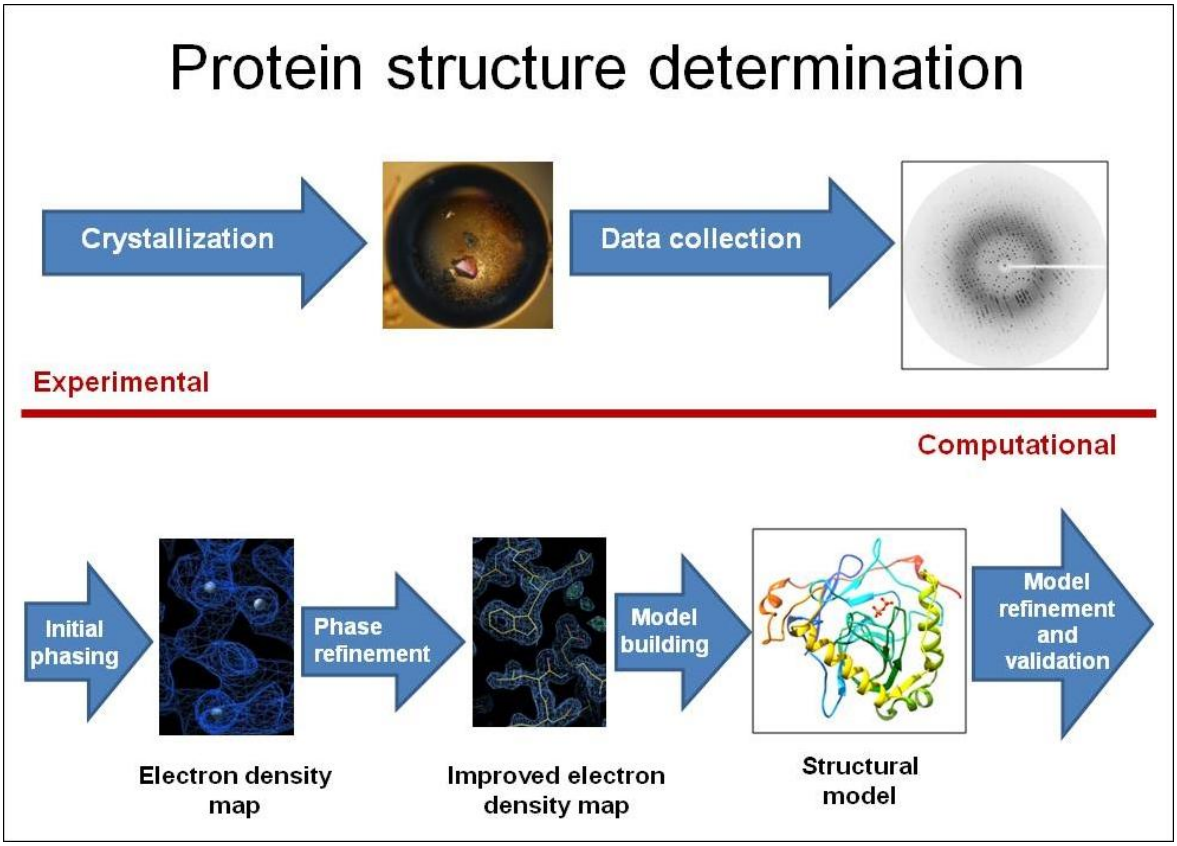


Fig. 20. Steps in protein structure determination

4.2.2.1. Crystallization

The process of crystal forming is controlled by the laws of physical chemistry and thermodynamics. In order to obtain protein crystals the molecules must assemble into a periodic three dimensional lattice. Protein crystals are held together by weak intermolecular interactions, such as hydrogen bonds, salt bridges and hydrophobic interactions, and contain large volumes filled with solvent (27-65%). This is the reason that protein crystals are fragile and require careful handling. Screening for initial protein crystallization conditions requires a large number of experiments exploring various conditions

The protein solution at high concentration (usually 5-20 mg/ml) is mixed with compounds which reduce the solubility of the protein. The common method to prompt protein to form crystals is vapour diffusion when concentration of the precipitant and the protein is gradually increased by water evaporation from the drop into the reservoir increasing the concentration of both the protein and the precipitant. The solution in which protein is dissolved becomes supersaturated and the protein could precipitate or crystallize. The crystallographer's goal is to identify solution conditions that favor the development of large crystals, since larger crystals offer improved resolution of the molecule. However the general problem in protein crystallization is that the crystallization condition for particular protein is unpredictable in advance. It needs to be found empirically and require testing many variables including: type of precipitant, buffer pH, ionic strength, protein concentration, temperature, as far as protein to screen solution ratio. Very often presence of various additional compounds: ligands and additives might be essential.

Due to facilitate testing many crystallization conditions robots and a number of ready-to-use commercial crystallization solutions sets have been developed.

Laboratory procedure for AtPPA1 crystallization

Prior setting up the crystallization screens, the AtPPA1 protein concentration was adjusted to 7.5 mg/ml and the protein solution was passed through an Ultrafree-MC Centrifugal Filter Unit (Millipore) with 0.1 µm pore size at 4°C. Initial screening for crystallization conditions was performed using a Robotic Sitting Drop Vapor Diffusion setup (Mosquito) with Morpheus HT-96, PACT premier HT-96 and JCSG-plus HT-96 reagents from Molecular Dimensions. 0.4µl protein samples were mixed with an equal amount of the reservoir solution and equilibrated against 100 µl reservoir solution, and the crystallization plates were stored at 19°C. After approximately two weeks initial crystals appeared. Four crystallization conditions were chosen for refinement by adjusting the pH as well as the concentration and volume of the protein solution. The best AtPPA1 crystals were obtained with 0.1 M succinic acid pH 7.0 and 15% w/v PEG 3350100. They were grown using the hanging-drop vapour diffusion method at 18°C by mixing 3 µl protein solution with 2 µl precipitating solution on a siliconized cover slide and equilibrating the drop against 0.5 ml precipitant solution. The crystals appeared within two weeks. After 1 month they were harvested with 0.4 mm nylon loops (Hampton Research), washed with cryo-protectant solution containing 20% (v/v) glycerol or 20% PEG400 in the reservoir cocktail, and vitrified in liquid nitrogen prior to synchrotron-radiation data collection.

4.2.2.2. Data collection

Data collection is the last experimental step in X-ray crystallography. Next steps are computations that can be easily repeated thus data collection is a step that requires a lot of decisions made by the crystallographer to choose the right strategy of measurement. Choosing the source of X-ray radiation is very important because protein crystals usually diffract weakly and synchrotron facilities prevail over in house laboratory sources. Important point is also protection of the crystals from damage caused by heat (thermal vibrations) and free radicals generated by the X-rays. Usually it is done by data collection at low temperatures (around 100 K) but the crystals before measurement need special treatment-soaking with cryosolution and freezing in liquid nitrogen. Once the crystal is centered on the goniometer, it is important to collect few preliminary images to assess the crystal quality, elucidate the point group and decide about optimal parameters for data collection strategy: oscillation angle, detector distance, exposure time, the total number of degrees covered in data collection.

Diffraction data are collected in snapshots during the crystal rotation. The oscillation angle (phi) describes the rotation of the crystal during one exposure. For protein crystals typical phi angles are 0.5-1 degrees but it is possible to decrease the angle to of 0.1 which is known as fine slicing. Larger oscillation angle give higher number of reflections on a single image.

Decreasing the oscillation angle to less than 0.5° help to avoid the overlap of reflections in the case of large unit cell dimensions where the reciprocal lattice points are closely spaced. On the other hand it is not recommended for poor diffracting crystals when short oscillation angle and short time of exposure is insufficient to record clear diffraction.

Decision about detector distance should be made considering: resolution of the crystal diffraction, pixel resolution of the detector and the absorption of the X-rays by air. Sometimes for a high-resolution data, it is necessary to collect two or more passes at different resolutions (low and high) with different exposure time and different crystal-to-detector distance to avoid the reflections overloads.

Optimal exposure time greatly influences data quality. In general, the longer crystal exposure time the higher the intensities measured, thus improved the signal to noise ratio. However, setting exposure time two aspects should be considered while too long exposure time could result in reflections overload - higher intensities that the limit of the detectors and the higher radiation damage of the crystal which results in lower data quality.

Very important aspect during diffraction data collection is the completeness of the data. The minimum number of degrees covered in data collection is dictated by the space group symmetry and crystal orientation. Because crystals exposed to X-rays quickly decay, it is important to collect the unique set of reflections in the minimal amount of time. However, significant improvement in signal to noise ratio can be achieved by collecting more data, up to 360°.

Various experiments also dictate the strategy of diffraction and data collection. For instance, multi-wavelength anomalous diffraction (MAD) requires very accurately measured reflection intensities at different wavelengths.

AtPPA1 data collection

X-ray diffraction datasets were collected from single crystals using the oscillation method under cryogenic conditions. The crystals were flash-vitrified in a liquid nitrogen. To prevent ice formation, the crystals were equilibrated with a solution containing a cryoprotective agent: 20% (v/v) glycerol or 20% or PEG400. A total of 120 diffraction images with 1° oscillation for the AtPPA1 crystal were collected on beamline 14.1 at the BESY synchrotron (Berlin, Germany).

4.2.2.3. Computational methods

The data collection is the last step of experimental procedures during solving the crystal structure. The next are based on computational analyses of the collected data and might be repeated in case of failure.

The following stages of data analysis are:

-data processing (visualization, indexing, integrating, averaging and scaling)

-structure solution (converting the numbers into electron density maps and solving the phase problem)

-model building and refinement

-structure validation and deposition

4.2.2.3.1. Data processing

X-ray diffraction data for the AtPPA1 were collected on beamline 14.1 at the BESY synchrotron (Berlin, Germany). A total of 120 diffraction images with 1° oscillation were indexed, integrated and scaled using XDS [43].

4.2.2.3.2. Structure solution, model building and refinement

The AtPPA1 structure was determined by the molecular replacement method, using Phaser [71] within Phenix suite [1]. To find the most appropriate model for molecular replacement, sequence alignment by BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was performed using 73 pdb files available from PDB (http://www.rcsb.org/pdb/). The inorganic pyrophosphatase from *Pyrococcus furiosus* (PDB code: 1twl) showed the highest sequence identity to AtPP1 (49%) was applied as a search model. Automatic structural model building was carried out using the online version of ARP/WARP [59]. The final model was obtained after 79 cycles of manual fitting in electron-density maps using Coot [19]. *Phenix.refine* was used for model refinement. Models were accepted as final, with twelve TLS groups[107] defined as suggested by refinement program.

4.2.2.3.3. Structure validation and deposition

The final model of AtPPA1 was validated using MolProbity [13] and diffraction precision index was calculated using SFCHECK [102]. The identity of the metal cations in both chains were confirmed using the calcium bond-valence sum (CBVS) method [75].

The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession code: 4lug.

4.2.3. Determination of the oligomeric state

Determination of the oligomeric state of AtPPA1 was performed using three metchods: (i) size exclusion chromatography, (ii) static light scattering and (iii) PDBePISA web server [54].

4.2.3.1. Size exclusion chromatography

Size exclusion chromatography was used to separate a mixture of AtPPA1 (75 kDa, trimer) and AtWRKY50 (20 kDa, monomer) used as a molecular mass standard. A mixture of the two proteins (3 mg of each) in 2 ml buffer consisting of 50 mM Tris–HCl pH 7,5, 200 mM NaCl, 1 mM TCEP, was loaded onto a Superdex 200 HiLoad 16/60 column (GE Healthcare) equilibrated with 50 mM Tris–HCl pH 7,5, 200 mM NaCl, 1 mM TCEP. Fractions corresponding to each separated protein were analyzed using SDS–PAGE.

4.2.3.2. Dynamic and static light scattering

Dynamic light scattering (DLS) is a modern method using to determine the size of the particles in the solution as well as to estimate the degree of homogeneity of the protein solution. DLS measures time-dependent fluctuations in the scattering intensity arising from particles undergoing Brownian motion. Brownian motions are random movements of particles suspended in a liquid. An important feature of the Brownian motion is the fact that the larger particles move more slower. Shining a monochromatic light beam, such as a laser, onto a solution with particles in Brownian motion causes a Doppler Shift when the light hits the moving particle, changing the wavelength of the incoming light. This change is related to the size of the particle. The rate of the Brownian motion is defined by a property known as the translational diffusion coefficient in infinitely-dilute solutions and is usually given as the symbol D_0. Measuring the diffusion coefficient and using the autocorrelation function gives a possibility to compute the sphere size distribution and description of the particle's motion in the medium. The benefit of this technique is that there is no practical upper or lower size limit for the materials being investigated. D_0 is often used to calculate the hydrodynamic radius of a sphere through the Stokes–Einstein equation. The hydrodynamic radius (R_H) or Stokes radius of a solute is the radius of a hard sphere that diffuses at the same rate as that solute. Factors that influence the protein hydrodynamic sizes are the molecular weight, shape, or conformation of the molecule and also whether the protein is in his native or folded state.

Static light scattering (SLS) is a technique that measures the intensity of the scattered light at different angles to obtain the average molecular weight M_w of a macromolecule like a protein in solution. An absolute molecular mass of a protein sample in solution may be experimentally determined to an accuracy of better than 5% through exposure to low intensity laser light (690 nm). The intensity of the scattered light measured as a function of angle allows calculation of the root mean square radius, also called the radius of gyration R_g and the second virial coefficient A_2 by measuring the scattering intensity for many samples of various concentrations. Static light scattering (SLS) is capable of measuring molar masses within the range 10^3–10^8 g/mol and is therefore ideal for quality control in protein preparation (e.g. for structural studies) and to the determination of solution oligomeric state (monomer/dimer etc.).

To sum up the Dynamic light scattering measures fluctuation of intensity of scattered light with time and allow us to obtain translation diffusion coefficient (DT) and hydrodynamic radius/Stokes radius (Rh). Static light scattering measures time-averaged intensity of

scattered light and let to obtain three parameters: Molar Mass, Rg (radius of gyration) and A2 - second virial coefficient.

To estimate more precisely molecular weight, static light scattering (SLS) experiments were performed using Zetasizer μV (Malvern Instruments) at a wavelenght of 488 nm. The autocorrelation function of scattered light intensity was automatically calculated.
Debye plot was created by measuring the scattered light at a single angle (90 °) at multiple protein sample concentrations. Debye plot was used to determine the molecular weight by extrapolation to zero concentration and the slope was used to calculate the second virial coefficient.

4.2.3.3. PDBePISA web server

PDBe PISA web server [54] was used for prediction of AtPPA1 multimeric assemblies and calculation of distances between subunits. The program automatically generates table that summarizes all of subunit interactions, showing both the hydrogen bonds and hydrophobic contacts.

4.2.4. N-terminus analyses

4.2.4.1. Protein sequencing

Preparation of a AtPPA1 sample for Edman degradation included gel electrophoresis separation according to Schaegger and Jagow [90] followed by protein transfer to a PVDF membrane.
Three crystals of AtPPA1 were fished from crystallisation drop, dissolved in protein buffer and mixed 1:1 v/v with sample buffer (125 mM Tris-HCl pH 6,8; 4 % SDS; 20 % glycerol; 50 mg DTT; a pinch of Coomasie Brillant Blue G-250 to light blue color). Sample was boiled 5 min in 100°C and developed using Tris-Tricine SDS-PAGE [90]. The gel was run at a constant voltage 30V and 80V in stacking gel and resolving gel respectively. The protein semi-dry transfer was performed immediately on a PVDF membrane (Immobilon-P, 0.45 μm; Millipore) using transfer buffer (10 mM CHAPS, pH11.0; 10% methanol). The membrane was stained for 30 s in 0,1 % Coomassie Brillant Blue R-250, 40% methanol, 1% acetic acid and destained in water. The single protein band corresponding to the molecular weight of 25 kDa was cut and subjected to Edman degradation cycles performed using a fully automated sequencer (Procise 491; Applied Biosystems, USA).
I kindly acknowledge BioCentrum, Krakow, Poland, for N-terminal sequencing analysis.

4.2.4.2. Prediction of signal peptides and cleavage site

For prediction of N-terminal target peptide, two online tools were used: TargetP [18] and MitoProt [14].

TargetP is an online tool that predicts the subcellular location of eukaryotic proteins. It is not clear-cut single location predictor since it also deal with other pre-sequences. The location assignment is based on the investigation of any of the N-terminal pre-sequences: chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP) and "other" localizations. The success rate of predictions was estimated as 85% for plant sequences or 90% for non-plant sequences on redundancy-reduced test sets. The TargetP predictor [17] differentiates between secretory proteins, mitochondrial proteins, chloroplast proteins, and everything else. The method looks for N-terminal sorting signals by feeding the outputs from SignalP, ChloroP, and mitochondrial predictor into a 'decision neural network' that makes the final choice between the different compartments [17]. TargetP also predicts cleavage sites with levels of correctly predicted sites ranging from approximately 40% to 50% for chloroplastic and mitochondrial presequences to above 70% for secretory signal peptides [79]. TargetP is available as a web-server at http://www.cbs.dtu.dk/services/TargetP/.

MitoProt predicts localization of a protein by calculating a number of physicochemical parameters from its amino acid sequence, and then computing a linear discriminant function (LDF) which is compared to a cutoff for mitochondrial/non-mitochondrial localization prediction. MitoProt supplies a series of parameters that permit theoretical evaluation of mitochondrial targeting sequences, as well as calculation of the most hydrophobic fragment of 17 residues in the sequence [14]. Both MitoProt and TargetP suggest a potential cleavage site of the predicted mitochondrial targeting peptides.

4.2.5. Enzymatic activity assay

AtPPA1 and both D98N and D103N mutants activity was determined essentially using assay based on colorimetric reaction. The orthophosphate released during reaction react with ammonium molybdate to form phosphomolybdic acid then the phosphomolybdic acid is reduced by $FeSO_3$ in a weak acid solution and the blue colour appeared can be measured at 660nm. The standard curve was prepared with KH_2PO_4 and was linear from 0.5 to 20 nmoles of phosphate. The enzymatic assay was performed at room temperature. 1 ml of reaction mixtures were prepared. Each contained 0.2 µg purified AtPPA1 protein, 2.5 mM sodium pyrophosphate as a substrate, 2 mM $MgCl_2$ and 50 mM Tris pH 7.5. During

incubation (80 min) the amount of released phosphate was checked every 10 min at following time points, 100 µl of reaction mixture was transferred into 500 µl Taussky-Shorr Reagent [99] to stop the reaction and to estimate released phosphate product. The reagent was prepared freshly by adding 10 ml of 10% ammonium molybdate, tetrahydrate, in 10 N H_2SO_4, to 70 ml deionized water; then added 5 g of ferrous sulfate, heptahydrate, the final volume was brought to 100 ml with deionized water. After addition of Taussky-Shorr reagent, samples containing phosphate developed a blue colour, while negative controls remained clear. Samples were diluted with 400 µl of water. The optical density was measured at 660 nm within 15 min to avoid the background level increment due to nonenzymatic pyrophosphate hydrolysis.

Enzymatic activity was tested qualitatively also with others divalent metal cations: Mn^{2+}, Co^{2+}, Mo^{2+}, Fe^{2+} and Ca^{2+} instead of Mg^{2+}. The inhibitory effect of NaF was determined by adding it at different concentration (0-2mM) to the incubation buffer at presence of Mg^{2+}.

4.2.6. Graphic programs used for structure illustrations and alignments

Molecular graphics were prepared using the program UCSF Chimera package [85]. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco.

Multiple sequence alignments were calculated in ClustalW2 [60]. The figures with highlighted residues were prepared using the Bioedit editor [25] or GenDoc [78]

Pairwise comparison and best Cα-alignment of compared protein structures were done using PDBeFold service at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm), authored by E. Krissinel and K. Henrick [52, 53].

5. Summary

Inorganic pyrophosphatase (PPase) is a ubiquitous cytosolic enzyme which catalyzes the hydrolysis of inorganic pyrophosphate (PPi) to orthophosphate (Pi).

Arabidopsis thaliana inorganic pyrophosphatase (AtPPA1) coding DNA (*ppa1* gene) was cloned into bacterial expression vector and overproduced in *E. coli* cells as a fusion (His-tagged) protein. The recombinant protein was purified from the bacterial lysate by two consecutive chromatographic steps: chelating chromatography on Ni^{2+}-charged resin followed by FPLC size exclusion chromatography. The homogenous protein was submitted for crystallization. X-Ray diffraction data extending to 1.9Å resolution were collected using synchrotron radiation. The structure was solved by molecular replacement using *Pyrococcus furiosus* structure coordinates (PDB code: 1twl) having the highest sequence identity to AtPP1 (49%) and refined to R-factor below 15.6%. The structure coordinates of AtPPA1 have been deposited in PDB with code: 4lug. The structure of AtPP1 represents an OB-fold which overlaps with other structural models for known bacterial and yeast inorganic pyrophosphatases. PPases are oligomeric enzymes that are active as homohexamers, or homotetramers composed of about 20 kDa subunits in prokaryotes. Eukaryotic PPases act as homodimers with 30-35 kDa subunits. Plant PPase is an exception because it function as 75 kDa trimer. Moreover, the analysis of AtPPA1 sequence using PsiPred (signal peptide predictor) revealed that it posses N-terminal putative transit peptide of mitochondrial targeting, and a possible cleavage site at Val31. *In vitro*, cleavage of short (few kDa) fragment is observed during protein storage. Mutant with substitution D98N shows delayed autoproteolysis compared to wild type (WT) protein. Crystal structure refinement and protein sequencing revealed that the N-terminal fragment corresponding to the predicted mitochondrial targeting peptide is cleaved.

6. Streszczenie

Pirofosfataza jest enzymem hydrolizującym nieorganiczny pirofosforan PPi do fosforanu Pi w obecności dwuwartościowych jonów metali. W reakcji z udziałem PPazy w komórce odzyskiwany jest fosforan z pirofosforanu, powstający w wyniku degradacji ATP. Enzym ten jest obecny u wszystkich organizmów, a reakcja hydrolizy pirofosforanu pełni istotną rolę w obiegu fosforu w komórce.

W genomie *Arabidopsis thaliana* zidentyfikowano 5 genów kodujących homologi pirofosfataz cytoplazmatycznych. Do tej pory nie było żadnych danych strukturalnych dotyczących pirofosfataz roślinnych. Głównym celem mojej pracy były badania strukturalne pirofosfatazy AtPPA1. Gen *Atppa1* kodujący ten enzym został wklonowany do bakteryjnego wektora ekspresyjnego i poddany ekspresji w komórkach *E. coli* w celu produkcji białka rekombinowanego ze znacznikiem histydynowym (His-tag). Rekombinowane białko zostało oczyszczone z bakteryjnego lizatu w dwóch kolejnych etapach: chromatografii powinowactwa na oraz filtracji żelowej a następnie poddane krystalizacji. Struktura krystaliczna pirofosfatazy AtPPA1 została rozwiązana z wysoką rozdzielczością (1.9 Å) metodą podstawienia cząsteczkowego i udokładniona (wskaźnik rozbieżności R poniżej 15,6%). Współrzędne struktury AtPPA1 zostały zdeponowane w bazie danych PDB (kod: 4lug).

Ogólna struktura białka AtPPA1 wykazuje duże podobieństwo do znanych wcześniej pirofosfataz z bakterii i drożdży mimo tego, że pirofosfataza z *A. thaliana* wykazuje ok. 40% identyczności (na poziomie sekwencji aminokwasowej) z enzymami bakteryjnymi i tylko ok. 20% identyczności z pirofosfatazami z innych organizmów eukariotycznych nie uwzględniając roślin. Ponadto, pirofosfatazy z różnych grup organizmów wykazują odmienną oligomeryczną organizację: bakteryjne tworzą heksamery, drożdżowe tworzą dimery, natomiast roślinna pirofosfataza AtPPA1 jest trimerem o masie 75 kDa.
Test aktywności enzymatycznej wykazał, ze otrzymane przeze mnie rekombinowane białko AtPPA1 jest biologicznie aktywne. Przygotowane 2 warianty tego białka z mutacjami w obrębie centrum aktywnego (D98N i D103N) nie wykazywały aktywności aktywności enzymatycznej. Dodatkowo zaobserwowałam, że podczas krystalizacji dochodzi do odcięcia 30 N-terminalnych reszt aminokwasowych. W celu zidentyfikowania miejsca cięcia zostało wykonane oznaczenie sekwencji N-końca pirofosfatazy AtPPA1. Interesujące jest to, że w przypadku mutantów, odcinanie N-terminalnego fragmentu białka

zachodzi wolniej. Analiza bioinformatyczna sekwencji AtPPA1 przewidziała, że to białko posiada N-końcowy peptyd sygnalny, z miejscem cięcia po Val31, kierujący to białko do mitochondrium.

7. References:

1. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
2. Ahn, S., et al., *The "open" and "closed" structures of the type-C inorganic pyrophosphatases from Bacillus subtilis and Streptococcus gordonii.* J Mol Biol, 2001. **313**(4): p. 797-811.
3. Avaeva, S.M., *Active site interactions in oligomeric structures of inorganic pyrophosphatases.* Biochemistry (Mosc), 2000. **65**(3): p. 361-72.
4. Baudisch, B., et al., *The exception proves the rule? Dual targeting of nuclear-encoded proteins into endosymbiotic organelles.* New Phytol, 2014. **201**(1): p. 80-90.
5. Baykov, A.A., et al., *Cytoplasmic inorganic pyrophosphatase.* Prog Mol Subcell Biol, 1999. **23**: p. 127-50.
6. Benini, S. and K. Wilson, *Structure of the Mycobacterium tuberculosis soluble inorganic pyrophosphatase Rv3628 at pH 7.0.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2011. **67**(Pt 8): p. 866-70.
7. Bennet, V.L., Ristrophe, D. L., Hamming, J. J., Butler, L. G. , *Maize leaf inorganic pyrophosphatase: isoenzymes, specificity for substrates, inhibitors, and divalent metal ions, and pH optima.* . Biochim. Biophys. Acta, 1973. **293**: p. 232-241.
8. Bieleski, R.L., *Phosphate pools, phosphate transport, and phosphate availability.* . Ann. Rev. Plant. Physiol., 1973. **241**(1): p. 225-252.
9. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal Biochem, 1976. **72**: p. 248-54.
10. Brunger, A.T., *Free R value: a novel statistical quantity for assessing the accuracy of crystal structures.* Nature, 1992. **355**(6359): p. 472-5.
11. Bustos, R., et al., *A central regulatory system largely controls transcriptional activation and repression responses to phosphate starvation in Arabidopsis.* PLoS Genet, 2010. **6**(9): p. e1001102.
12. Chao, T.C., et al., *Kinetic and structural properties of inorganic pyrophosphatase from the pathogenic bacterium Helicobacter pylori.* Proteins, 2006. **65**(3): p. 670-80.
13. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
14. Claros, M.G. and P. Vincens, *Computational method to predict mitochondrially imported proteins and their targeting sequences.* Eur J Biochem, 1996. **241**(3): p. 779-86.
15. Cooperman, B.S., A.A. Baykov, and R. Lahti, *Evolutionary conservation of the active site of soluble inorganic pyrophosphatase.* Trends Biochem Sci, 1992. **17**(7): p. 262-6.
16. Emanuelsson, O., et al., *Locating proteins in the cell using TargetP, SignalP and related tools.* Nat Protoc, 2007. **2**(4): p. 953-71.
17. Emanuelsson, O., et al., *Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.* J Mol Biol, 2000. **300**(4): p. 1005-16.
18. Emanuelsson, O. and G. von Heijne, *Prediction of organellar targeting signals.* Biochim Biophys Acta, 2001. **1541**(1-2): p. 114-9.
19. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

20. Fabrichniy, I.P., et al., *Structural studies of metal ions in family II pyrophosphatases: the requirement for a Janus ion.* Biochemistry, 2004. **43**(45): p. 14403-11.
21. Garcia-Contreras, R., H. Celis, and I. Romero, *Importance of Rhodospirillum rubrum H(+)-pyrophosphatase under low-energy conditions.* J Bacteriol, 2004. **186**(19): p. 6651-5.
22. Gomez-Garcia, M.R., M. Losada, and A. Serrano, *A novel subfamily of monomeric inorganic pyrophosphatases in photosynthetic eukaryotes.* Biochem J, 2006. **395**(1): p. 211-21.
23. Gomez-Garcia, M.R., M. Losada, and A. Serrano, *Comparative biochemical and functional studies of family I soluble inorganic pyrophosphatases from photosynthetic bacteria.* FEBS J, 2007. **274**(15): p. 3948-59.
24. Gonzalez, M.A., et al., *Evidence that catalysis by yeast inorganic pyrophosphatase proceeds by direct phosphoryl transfer to water and not via a phosphoryl enzyme intermediate.* Biochemistry, 1984. **23**(5): p. 797-801.
25. Hall, T.A., *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.* Nucleic Acid Symposium Series, 1999. **41**: p. 95-98.
26. Halonen, P., et al., *Single-turnover kinetics of Saccharomyces cerevisiae inorganic pyrophosphatase.* Biochemistry, 2002. **41**(40): p. 12025-31.
27. Hanks, S.K. and T. Hunter, *Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification.* FASEB J, 1995. **9**(8): p. 576-96.
28. Harding, M.M., *Geometry of metal-ligand interactions in proteins.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 3): p. 401-11.
29. Harrison, M.J., *Molecular and Cellular Aspects of the Arbuscular Mycorrhizal Symbiosis.* Annu Rev Plant Physiol Plant Mol Biol, 1999. **50**: p. 361-389.
30. Harutyunyan, E.H., et al., *X-ray structure of yeast inorganic pyrophosphatase complexed with manganese and phosphate.* Eur J Biochem, 1996. **239**(1): p. 220-8.
31. Harutyunyan, E.H., et al., *Crystal structure of holo inorganic pyrophosphatase from Escherichia coli at 1.9 A resolution. Mechanism of hydrolysis.* Biochemistry, 1997. **36**(25): p. 7754-60.
32. Heikinheimo, P., et al., *The structural basis for pyrophosphatase catalysis.* Structure, 1996. **4**(12): p. 1491-508.
33. Heikinheimo, P., et al., *Toward a quantum-mechanical description of metal-assisted phosphoryl transfer in pyrophosphatase.* Proc Natl Acad Sci U S A, 2001. **98**(6): p. 3121-6.
34. Hinsinger, P., *Bioavailability of soil inorganic P in the rhizosphere as affected by root-induced chemical changes: a review.* . Plant soil, 2001. **237**: p. 173-195.
35. Hohne, W.E. and P. Heitmann, *Tripolyphosphate as a substrate of the inorganic pyrophosphatase from baker's yeast; the role of divalent metal ions.* Acta Biol Med Ger, 1974. **33**(1): p. 1-14.
36. http://www.ebi.ac.uk/pdbe/prot_int/pistart.html, *Protein interfaces, surfaces and assemblies, service PISA at the European Bioinformatics Institute.*
37. Huang, K., et al., *Structure of the Pho85-Pho80 CDK-cyclin complex of the phosphate-responsive signal transduction pathway.* Mol Cell, 2007. **28**(4): p. 614-23.
38. Hunter, T., *Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling.* Cell, 1995. **80**(2): p. 225-36.

39. Jamsen, J., et al., *A CBS domain-containing pyrophosphatase of Moorella thermoacetica is regulated by adenine nucleotides.* Biochem J, 2007. **408**(3): p. 327-33.
40. Jeon, S.J. and K. Ishikawa, *Characterization of the Family I inorganic pyrophosphatase from Pyrococcus horikoshii OT3.* Archaea, 2005. **1**(6): p. 385-9.
41. Jiang, S.S., et al., *Purification and characterization of thylakoid membrane-bound inorganic pyrophosphatase from Spinacia oleracia L.* Arch Biochem Biophys, 1997. **346**(1): p. 105-12.
42. Josse, J., *Constitutive inorganic pyrophosphatase of Escherichia coli. 1. Purification and catalytic properties.* J Biol Chem, 1966. **241**(9): p. 1938-47.
43. Kabsch, W., *Xds.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
44. Kajander, T., J. Kellosalo, and A. Goldman, *Inorganic pyrophosphatases: one substrate, three mechanisms.* FEBS Lett, 2013. **587**(13): p. 1863-9.
45. Kaneko, Y., et al., *Transcriptional and post-transcriptional control of PHO8 expression by PHO regulatory genes in Saccharomyces cerevisiae.* Mol Cell Biol, 1985. **5**(1): p. 248-52.
46. Kankare, J., et al., *The structure of E.coli soluble inorganic pyrophosphatase at 2.7 A resolution.* Protein Eng, 1994. **7**(7): p. 823-30.
47. Kellosalo, J., et al., *The structure and catalytic cycle of a sodium-pumping pyrophosphatase.* Science, 2012. **337**(6093): p. 473-6.
48. Kim, Y., et al., *High-throughput protein purification and quality assessment for crystallization.* Methods, 2011. **55**(1): p. 12-28.
49. Klemme, B. and G. Jacobi, *Separation and characterization of two inorganic pyrophosphatases from spinach leaves.* Planta, 1974. **120**(2): p. 147-53.
50. Klock, H.E. and S.A. Lesley, *The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis.* Methods Mol Biol, 2009. **498**: p. 91-103.
51. Kornberg, A., *Pyrophosphorylases and phosphorylases in biosynthetic reactions.* Adv Enzymol Relat Subj Biochem, 1957. **18**: p. 191-240.
52. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2256-68.
53. Krissinel, E. and K. Henrick, *Multiple Alignment of Protein Structures in Three Dimensions.* Computational Life Sciences, 2005. **3695**: p. 67-78.
54. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. **372**(3): p. 774-97.
55. Kuhn, N.J., et al., *Methanococcus jannaschii ORF mj0608 codes for a class C inorganic pyrophosphatase protected by Co(2+) or Mn(2+) ions against fluoride inhibition.* Arch Biochem Biophys, 2000. **379**(2): p. 292-8.
56. Kunitz, M., *Crystalline inorganic pyrophosphatase isolated from baker's yeast.* J Gen Physiol, 1952. **35**(3): p. 423-50.
57. Kunitz, M., *An improved method for isolation of crystalline pyrophosphatase from baker's yeast.* Arch Biochem Biophys, 1961. **92**: p. 270-2.
58. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.
59. Langer, G., et al., *Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7.* Nat Protoc, 2008. **3**(7): p. 1171-9.
60. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-8.
61. Laskowski, R.A., et al., *PROCHECK - a program to check the stereochemical quality of protein structures.* J. App. Cryst., 1993. **26**: p. 283-291.

62. Leppanen, V.M., et al., *Sulfolobus acidocaldarius inorganic pyrophosphatase: structure, thermostability, and effect of metal ion in an archael pyrophosphatase.* Protein Sci, 1999. **8**(6): p. 1218-31.
63. Lopez-Bucio, J., et al., *Organic acid metabolism in plants: from adaptive physiology to transgenic varieties for cultivation in extreme soils.* Plant Sci, 2000. **160**(1): p. 1-13.
64. Lopez-Marques, R.L., et al., *Differential regulation of soluble and membrane-bound inorganic pyrophosphatases in the photosynthetic bacterium Rhodospirillum rubrum provides insights into pyrophosphate-based stress bioenergetics.* J Bacteriol, 2004. **186**(16): p. 5418-26.
65. Luoto, H.H., et al., *Membrane-integral pyrophosphatase subfamily capable of translocating both Na+ and H+*. Proc Natl Acad Sci U S A, 2013. **110**(4): p. 1255-60.
66. Maeshima, M., *Vacuolar H(+)-pyrophosphatase.* Biochim Biophys Acta, 2000. **1465**(1-2): p. 37-51.
67. Malinen, A.M., et al., *Na+-pyrophosphatase: a novel primary sodium pump.* Biochemistry, 2007. **46**(30): p. 8872-8.
68. Marchesini, N., et al., *Acidocalcisomes and a vacuolar H+-pyrophosphatase in malaria parasites.* Biochem J, 2000. **347 Pt 1**: p. 243-53.
69. Marschner, H., *Mineral nutrition of higher plants*. 1995, London: Academic Press.
70. Maslowski, P., H. Maslowska, and S. Kowalczyk, *Subcellular distribution and properties of alkaline inorganic pyrophosphatase of maize leaves.* Acta Biochim Pol, 1977. **24**(2): p. 117-26.
71. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
72. Merckel, M.C., et al., *Crystal structure of Streptococcus mutans pyrophosphatase: a new fold for an old mechanism.* Structure, 2001. **9**(4): p. 289-97.
73. Meyer, W., et al., *Purification, cloning, and sequencing of archaebacterial pyrophosphatase from the extreme thermoacidophile Sulfolobus acidocaldarius.* Arch Biochem Biophys, 1995. **319**(1): p. 149-56.
74. Mimura, T., *Regulation of phosphate transport and homeostasis in plant cells.* Internat. Rev. Cytol. , 1999. **1911 (1)**: p. 149-200.
75. Muller, P., S. Kopke, and G.M. Sheldrick, *Is the bond-valence method able to identify metal atoms in protein structures?* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 1): p. 32-7.
76. Murzin, A.G., A.M. Lesk, and C. Chothia, *Principles determining the structure of beta-sheet barrels in proteins. II. The observed structures.* J Mol Biol, 1994. **236**(5): p. 1382-400.
77. Navarro-De la Sancha, E., et al., *Characterization of two soluble inorganic pyrophosphatases from Arabidopsis thaliana.* Plant science, 2007. **172**(4): p. 796-807.
78. Nicholas, K.B., Nicholas, H. B., Deerfield, D., *GeneDoc: Analysis and Visualization of Genetic Variation*. 1997.
79. Nielsen, H., et al., *Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.* Protein Eng, 1997. **10**(1): p. 1-6.
80. Oliva, G., et al., *Characterization of the inorganic pyrophosphatase from the pathogenic bacterium Helicobacter pylori.* Arch Microbiol, 2000. **174**(1-2): p. 104-10.
81. Oshima, Y., N. Ogawa, and S. Harashima, *Regulation of phosphatase synthesis in Saccharomyces cerevisiae--a review.* Gene, 1996. **179**(1): p. 171-7.

82. Osmont, K.S., R. Sibout, and C.S. Hardtke, *Hidden branches: developments in root system architecture.* Annu Rev Plant Biol, 2007. **58**: p. 93-113.
83. Parfenyev, A.N., et al., *Quaternary structure and metal ion requirement of family II pyrophosphatases from Bacillus subtilis, Streptococcus gordonii, and Streptococcus mutans.* J Biol Chem, 2001. **276**(27): p. 24511-8.
84. Persson, B.L., et al., *Regulation of phosphate acquisition in Saccharomyces cerevisiae.* Curr Genet, 2003. **43**(4): p. 225-44.
85. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.
86. Raghothama, K.G., Karthikeyan, A. S. , *Phosphate acquisition.* . Plant Soil 2005. **274**: p. 37-49.
87. Rantanen, M.K., et al., *Structure of the Streptococcus agalactiae family II inorganic pyrophosphatase at 2.80 A resolution.* Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 6): p. 738-43.
88. Samygina, V.R., et al., *Reversible inhibition of Escherichia coli inorganic pyrophosphatase by fluoride: trapped catalytic intermediates in cryo-crystallographic studies.* J Mol Biol, 2007. **366**(4): p. 1305-17.
89. Schachtman, D.P., R.J. Reid, and S.M. Ayling, *Phosphorus Uptake by Plants: From Soil to Cell.* Plant Physiol, 1998. **116**(2): p. 447-53.
90. Schagger, H. and G. von Jagow, *Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa.* Anal Biochem, 1987. **166**(2): p. 368-79.
91. Schlesinger, M.J. and M.J. Coon, *Hydrolysis of nucleoside diand triphosphates by crystalline preparations of yeast inorganic pyrophosphatase.* Biochim Biophys Acta, 1960. **41**: p. 30-6.
92. Schulze, S., et al., *Identification of an Arabidopsis inorganic pyrophosphatase capable of being imported into chloroplasts.* FEBS Lett, 2004. **565**(1-3): p. 101-5.
93. Serrano, A., et al., *Proton-pumping inorganic pyrophosphatases in some archaea and other extremophilic prokaryotes.* J Bioenerg Biomembr, 2004. **36**(1): p. 127-33.
94. Shintani, T., et al., *Cloning and expression of a unique inorganic pyrophosphatase from Bacillus subtilis: evidence for a new family of enzymes.* FEBS Lett, 1998. **439**(3): p. 263-6.
95. Simmons, S. and L.G. Butler, *Alkaline inorganic pyrophosphatase of maize leaves.* Biochim Biophys Acta, 1969. **172**(1): p. 150-7.
96. Sivula, T., et al., *Evolutionary aspects of inorganic pyrophosphatase.* FEBS Lett, 1999. **454**(1-2): p. 75-80.
97. Sjoling, S. and E. Glaser, *Mitochondrial targeting peptides in plants.* Trends in plant science, 1998. **3**(4): p. 136-140.
98. Sklyankina, V.A. and S.M. Avaeva, *The quaternary structure of Escherichia coli inorganic pyrophosphatase is essential for phosphorylation.* Eur J Biochem, 1990. **191**(1): p. 195-201.
99. Taussky, H.H. and E. Shorr, *A microcolorimetric method for the determination of inorganic phosphorus.* J Biol Chem, 1953. **202**(2): p. 675-85.
100. Teplyakov, A., et al., *Crystal structure of inorganic pyrophosphatase from Thermus thermophilus.* Protein Sci, 1994. **3**(7): p. 1098-107.
101. Ticconi, C.A. and S. Abel, *Short on phosphate: plant surveillance and countermeasures.* Trends Plant Sci, 2004. **9**(11): p. 548-55.
102. Vaguine, A.A., J. Richelle, and S.J. Wodak, *SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their*

agreement with the atomic model. Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 1): p. 191-205.
103. Vance, C.P., Uhde-Stone C, and A.D. L., *Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource.* New Phytology, 2003. **157**: p. 423-447.
104. Veech, R.L., G.A. Cook, and M.T. King, *Relationship of free cytoplasmic pyrophosphate to liver glucose content and total pyrophosphate to cytoplasmic phosphorylation potential.* FEBS Lett, 1980. **117 Suppl**: p. K65-72.
105. Weiner, H., Stitt, M., Heldt, H. W. , *Subcellular compartmentation of pyrophosphate and alkaline pyrophosphatase in leaves.* Biophys. Biochem. Acta, 1987. **893**: p. 13-21.
106. Welsh, K.M., et al., *Catalytic specificity of yeast inorganic pyrophosphatase for magnesium ion as cofactor. An analysis of divalent metal ion and solvent isotope effects on enzyme function.* Biochemistry, 1983. **22**(9): p. 2243-8.
107. Winn, M.D., M.N. Isupov, and G.N. Murshudov, *Use of TLS parameters to model anisotropic displacements in macromolecular refinement.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 1): p. 122-33.
108. Wu, C.A., et al., *Structure of inorganic pyrophosphatase from Helicobacter pylori.* Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 11): p. 1459-64.
109. Wu, P., et al., *Phosphate starvation triggers distinct alterations of genome expression in Arabidopsis roots and leaves.* Plant Physiol, 2003. **132**(3): p. 1260-71.
110. Zyryanov, A.B., et al., *Mechanism by which metal cofactors control substrate specificity in pyrophosphatase.* Biochem J, 2002. **367**(Pt 3): p. 901-6.
111. Zyryanov, A.B., et al., *Rates of elementary catalytic steps for different metal forms of the family II pyrophosphatase from Streptococcus gordonii.* Biochemistry, 2004. **43**(4): p. 1065-74.