Structural characterization of *Arabidopsis thaliana* ethylene signaling molecules and the non-ribosomal peptide synthetase from *Planktothrix agardhii*

DISSERTATION

in fulfilment of the requirements for the degree “Dr. rer. nat.”
of the Faculty of Mathematics and Natural Sciences
at Kiel University

Submitted by
Heidi Kaljunen
Kiel, 2014

First referee:	Prof. Dr. Axel J. Scheidig
Second referee:	Prof. Dr. Frank Sönnichsen

Date of the oral examination:	12.12.2014
Approved for publication:	12.12.2014

Signed: Prof. Dr. Wolfgang J. Duschl, Dean

Erklärung

Ich versichere, dass ich meine Dissertation

Structural characterization of *Arabidopsis thaliana* ethylene signaling molecules and the non-ribosomal peptide synthetase from *Planktothrix agardhii*

selbständig, ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe.

Die Dissertation wurde in der jetzigen oder in einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Kiel, 12. Dezember 2014

Table of Contents

ABSTRACT

Plants employ a complex network of signaling pathways to regulate developmental processes and to mediate the responses to both environmental and biological stress factors. Ethylene is one of the key plant hormones involved in controlling this network, which has made it and its signaling pathway a target of intense research for several decades. In the model plant *Arabidopsis thaliana*, the plant hormone is detected by a group of five receptors (ETR1, ERS1, ETR2, ERS2 and EIN4) that resemble the sensor histidine kinases of bacterial two-component system. The main aim in this thesis study was the expression and purification of the full-length ETR1 for structural studies to gain insights into the initial steps in ethylene signaling. The FL ETR1 was successfully expressed in baculovirus expression vector system but the isolation of the receptor from the membrane was hampered. In addition to the FL ETR1, the cytosolic portion of the receptor was studied using Small Angle X-ray Scattering. The resulting SAXS model had the expected dimeric arrangement.

EDR1 from *A. thaliana* is a CTR1-like MAPKKK that is involved in regulating disease resistance responses, cell death and also ethylene-induced senescence. It possesses an N-terminal regulatory domain and C-terminal catalytic domain wit Ser/Thr kinase activity. As EDR1 has been shown to autophosphorylate *in trans*, the mechanism of this was studied using X-ray crystallography. A crystal structure for the catalytically inactive kinase domain of EDR1 (EDR1-D792N) was obtained in the presence of the ATP substrate analog AMP-PNP. The asymmetric unit contained two molecules, one of which surprisingly was in an active-like conformation. Furthermore, the active-like EDR1-D792N molecule was found to form an authentic trans-autophosphorylation complex with the inactive monomer from the adjacent asymmetric unit.

In addition to the plant defense signaling proteins, an adenylation (A) domain from cyanobacterial non-ribosomal peptide synthetase (NRPS) was studied. NRPSs are large multidomain enzymes that are found from a number of fungal and bacterial species and catalyze the ribosome-independent assembly of biologically active peptides with diverse composition and function. The A domain plays a central role in the NRPS system as it recognizes and activates the amino acid, which is incorporated into the growing peptide. The A domain ApnA A_1 from the Anabaenopeptin synthetase of *Planktothrix agardhii* is an interesting member of its class as it has an unusual ability to activate two very distinct amino acids (arginine and tyrosine). Structural studies on this enzyme were performed to elucidate its bi-specificity. Based on the solved ApnA A_1 structures, two active site residues with a crucial role in the substrate binding were identified. The mutation of these residues led to enzyme variants, which were mono-specific for either tyrosine or arginine, or in some instances were able to activate L-tryptophan. Additionally a number of ApnA A_1 mutants were shown to activate unnatural amino acids (4-fluorophenylalanine and 4-azidophenylalanine). A final peptide product with an unnatural amino acid incorporated, could possibly have useful industrial or pharmaceutical applications.

ZUSAMMENFASSUNG

Pflanzen nutzen ein komplexes Signalweiterleitungsnetzwerk zur Regulierung von Entwicklungsprozessen und zur Reaktion auf umweltbedingte und biologische Belastungen. Ethylen ist ein Schlüsselhormon in der Kontrolle dieses Netzwerks, weshalb es zusammen mit seinem Signalweg Objekt jahrzehntelanger intensiver Untersuchungen wurde. Im Modellsystem *Arabidopsis thaliana* wird das Hormon von einer Gruppe aus fünf Rezeptoren (ETR1, ERS1, ETR2, ERS2 und EIN4) erkannt, die den Sensorhistidinkinasen bakterieller zwei-komponenten Systeme ähneln. Hauptziel dieser Arbeit war die Expression und Reinigung des gesamten ETR1 zur strukturellen Untersuchung, um einen Einblick in die ersten Schritte der Ethylensignalisierung zu erhalten. Der vollständige ETR1 wurde erfolgreich im Baculovirussystem exprimiert, die Isolation des Rezeptors aus der Membran gelang allerdings nicht. Zuätzlich zum vollständigen ETR1 wurde die zytosolische Domäne des Rezeptors mit Hilfe der Kleinwinkelstreuung untersucht. Das entsprechende SAXS Modell war wie erwartet ein Dimer.

EDR1 aus *A. thaliana* ist eine CTR1- ähnliche MAPKKK, die Resistenzen, Zelltod und ethyleninduzierte Seneszenz reguliert. Es besteht aus einer N-terminalen Regulationsdomäne und einer katalytischen C-terminalen Domäne mit Ser/Thr Kinaseaktivität. Da gezeigt wurde, dass EDR1 *trans* autophosphoryliert, sollte dieser Mechanismus röntgenkristallographisch untersucht werden. Die Kristallstruktur der katalytischen Domäne von EDR1 (EDR1-D792N) zusammen mit dem ATP-Analogon AMPPNP wurde gelöst. Die asymmetrische Einheit enthält 2 Moleküle, wovon eines überraschenderweise in der aktiven Konformation vorlag. Außerdem formt das aktive

EDR1-D792N einen trans-autophosphorylierenden Komplex mit dem inaktiven Monomer einer benachbarten asymmetrischen Einheit.

Zusätzlich zum pflanzlichen Abwehrmechanismus wurde die Adenylierungsdomäne der cyanobakteriellen nicht-ribosomalen Peptidsynthetase (non-ribosomal peptide synthetase, NRPS) untersucht. NRPS sind große Multidomänenenzyme, die in einer Reihe von Pilzen und bakterieller Spezies vorkommen und den ribosomenunabhängigen Aufbau biologisch aktiver Peptide unterschiedlicher Zusammensetzung und Funktion katalysieren. Die A-Domäne spielt eine zentrale Rolle in den NRPS, da sie die Aminosäure erkennt und aktiviert, die in das wachsende Peptid eingebaut wird. Die A-Domäne ApnA A_1 aus *Planktothrix agardhii* der Anabaenopeptinsynthetase is ein aufschlussreicher Vertreter seiner Klasse, da es die außergewöhnliche Fähigkeit besitzt zwei sehr unterschiedliche Aminosäuren (Arginin und Tyrosin) zu aktivieren. Strukturelle Untersuchungen zur Klärung der doppelten Spezifität wurden durchgeführt. Basierend auf Strukturen der ApnA A_1 wurden 2 Reste des aktiven Zentrums identifiziert, die eine zentrale Rolle bei der Substratbindung spielen. Mutation dieser Reste resultierte in Enzymvarianten, die entweder für Tyrosin oder Arginin mono-spezifisch waren, oder in einigen Fällen in der Lage waren L-Tryptophan zu aktivieren. Außerdem konnten einige ApnA A_1 Mutanten unnatürliche Aminosäuren (4-Fluorophenylalanin und 4-Azidophenylalanin) aktivieren. Ein Peptidprodukt mit einer eingebauten unnatürliche Aminosäuren könnte für industrielle oder pharmazeutische Anwendungen nützlich sein.

1. INTRODUCTION

1.1 Production of antimicrobial compounds in plants and bacteria

All organisms are in constant competition for living space and more importantly nutrients. Plants are one of the central and crucial sources of energy for many heterotrophic organisms. As sessile entities they cannot resort to changing their location when acquiring nutrients or confronted with an attacker. Instead, they have developed other mechanisms to protect their integrity. One of the key defence mechanisms in plants is the production of chemical compounds that in the case of pathogen attack halt its progression in the plant cell or in case of natural enemies like herbivores, make the plant undesirable for consumption. These antimicrobial compounds have a variety of shapes and forms, ranging from small toxic organic or inorganic molecules to larger protein molecules that interfere with the invader's biological systems and have ultimately lethal consequences for the attacker. Some of these compounds are produced as a part of the plant innate immune system and are stored in specialized organs and tissues whereas others are synthesized or expressed only in response to pathogen attack (Bednarek & Osbourn, 2009). The first group includes essential oils (e.g. spearmint oil), alkaloids (e.g. nicotine and atropine) and phenolics (e.g. terpenoids, lignin precursors, catechols, coumarins and tannins) (Bennett & Wallsgrove, 1994). The production of defence molecules of the latter kind requires detection of the pathogen, which provides a signal for the plant to activate their synthesis. These include proteins such as glucanases, chitinases and defensins, which target specific structural features of the pathogen by degrading them but also enzymes that needed for the biosythesis of phytoalexins (Ahuja

et al., 2012) and other often phenylpropanoid- or fatty acid-derived secondary metabolites (Scheel, 1998). Expression of these proteins is under tight control and regulated by several plant hormones such as ethylene, salicylic acid and jasmonic acid.

Despite the abundancy of pathogenic species in the bacterial phyla, the bacteria themselves are also faced with situations where elaborate defence mechanims are needed. Main reasons for bacterial defence responses arise from the environmental stress factors, such as the presence of harmful chemicals, UV radiation or the lack of nutrients but also from the presence of other microorganisms. These factors form a range of cues that drive the bacterial movement. Bacteria, like plants have a number of signaling systems that are employed to detect changes in its enviroment and accordingly trigger a suitable response. As microorganisms, bacteria are one of the most abundant species in the planet and very often are forced to compete for the available nutrients with other bacterial species. Possibly due to this, some bacteria are capable of producing chemicals with antimicrobial properties that function against another microorganism. A selection of bacterial species possesses a special enzymatic system that produces complex amino acid derivatives as secondary metabolites in a ribosome-independent manner. These non-ribosomal peptides, produced by the multienzyme Non-Ribosomal peptide synthetase, form an important group of antimicrobial compounds that not only are useful for the organism itself but have also valuable applications in pharmaceutical industry. The details of the non-ribosomal peptide synthesis and the concominant applications are discussed further in dedicated chapter of this thesis.

1.2. Plant immunity

Plants are exposed on a daily basis to various different stress factors that include both biotic ones, i.e. pathogens and abiotic ones such as drought, radical changes in

temperature, ultraviolet radiation, lack of nutrients and pollution. To overcome these external threats they have developed elaborate defense mechanisms. Unlike vertebrates, plants lack mobile cells that are an essential part of the adaptive immune system and therefore rely on their innate immune system and activation of resistance (R) genes leading eventually to systemic resistance that confers a whole-plant wide protection against the attacker (Jones & Dangl, 2006).

As a part of their first line of defense plants use constitutive defenses, which includes an array of structural barriers, such as cell walls, waxy epidermal cuticles and bark but also preformed antimicrobial metabolites (Nürnberger *et al*., 2004; Pieterse *et al*., 2009). The latter ones are usually formed as secondary metabolites and belong to one of three large chemical classes: terpenoids, phenolics or alkaloids. Plants also employ inducible defenses as basal level protection against the invading pathogen. The activation of these defenses is based on the recognition of specific molecular determinants that originate either from the pathogenic organism itself or from its activity. The former ones are commonly referred as pathogen- or microbe-associated molecular patterns (PAMPs or MAMPs) and include compounds abundant in certain type of pathogens such as flagellin, chitin, glycoproteins and lipopolysaccharides (Nürnberger *et al*., 2004; Pieterse *et al*., 2009). In addition to PAMPs and MAMPs, the plant immune response can be triggered by damage-associated molecules, also referred as damage-associated molecular patterns (DAMPs), which are generated from the degradation of the plant cell wall by the lytic enzymes of the pathogen (Nürnberger *et al*., 2004; Muthamilarasan & Prasad, 2013). For simplicity the triggering of plant immune responses by these microbial elicitors is further described using only the term PAMPs. It should be noted that the basic mechanisms also apply for MAMPs and DAMPs.

PAMPs are perceived by specific receptors called pattern recognition receptors (PRR), which localize at the plant cell surface (Jones & Dangl, 2006; de Wit, 2007; Boller & He, 2009; Muthamilarasan & Prasad, 2013). Upon the detection of PAMPs, an immune response known as PAMP-triggered immunity (PTI) is activated. The PTI associated initial responses include changes in ion fluxes across the plasma membrane, activation of MAPK cascades, production of reactive oxygen species (ROS) and production of reactive nitrogen intermediates (RNI) (Jabs *et al*., 1997; Muthamilarasan & Prasad, 2013). These responses can lead into or occur in parallel with transcriptional activation of pathogenesis-related genes, which include production of lytic enzymes (chitinases, glucanases and proteases) aimed to degrade pathogen-specific structural materials, or synthesis of antimicrobial proteins or antimicrobial secondary metabolites (Nürnberger *et al*., 2004).

A well-known example of PAMP-PRR interaction is the recognition of the 22 amino acid epitope in the N terminus of flagellin (flg22) by the FLS2 (flagellin sensitive 2) (Gomez-Gomez & Boller, 2000). Flagellin is the basic building block of the flagellum, which is needed for bacterial motility and consequently is also important for the pathogenicity of the bacteria (Zipfel & Felix, 2005; Ramos *et al.*, 2004; Tans-Kersten *et al.*, 2001). In addition to the importance of flagellin to bacteria, the fact that its N- and C-terminus are conserved among several bacterial species makes it a prime candidate as an immune response elicitor not only in plants but also in other eukaryotic organisms.

Pathogens have also developed means to overcome the defence mechanisms of PTI by producing specific effector molecules (Jones & Dangl, 2006). These effectors act against the PTI-related defence molecules either through direct interaction or indirect interference with the signaling pathway that is required for their production. The pathogen effectors derive their origin from the so-called *avirulence* (*avr*) genes, the

products of which can be recognized by the plant. In the course of evolution, plants have developed a set of *pathogen resistance* (*R*) genes, which recognize the *avr* genes or their corresponding gene products. This recognition is based on either gene-to-gene interaction or expression of proteins that bind and thus inactivate the pathogenic elicitors. Currently it is known that most of the *R* genes encode for membrane-bound receptor proteins with a nucleotide binding (NB) domain and a leucine-rich repeat (LRR) domain and are therefore referred as NB-LRR proteins. The recognition of the pathogenic effector molecule leads into the effector-triggered immunity (ETI), which sets out a complex set of signaling events that culminate in the hypersensitive response (HR) in the form of programmed cell death. The distruction of the infected tissue aims to prevent the spread of the pathogen to other tissues and functions as a signal to activate the global defence, known as systemic acquired resistance (SAR).

The detection of a pathogen does not only trigger responses at the infection site but also in distal tissues that have not yet been affected by the pathogen attack. This is achieved by the expression of specific signaling molecules, such as salicylic acid (SA), that activate the systemic expression of antimicrobial pathogenesis-related (PR) genes in the distal tissues in order to avoid secondary infection (Fu & Dong, 2013). This phenomenon is also known as SAR and provides the plant protection against the reoccurent attack of a pathogen. Besides SAR, plants also employ another similar form of systemic resistance called induced systemic resistance (ISR) to achieve long-term protection against attacking microorganisms. Unlike SAR, ISR is independent of SA-signaling but relies on the jasmonic acid and ethylene signaling pathways instead. The key to the effectiveness and fine-tuning of the both local and global defense mechanisms lies in the crosstalk between the different phytohormone signaling pathways.

1. 3 Ethylene signaling

1.3.1 Phytohormone ethylene

Ethylene (C_2H_4) is a simple unsaturated hydrocarbon molecule that belongs to the diverse group of plant hormones (i.e. phytohormones). The history of ethylene research dates back to the 19th century when it was observed that leaks of illuminating gas induced premature senescence and defoliation in plants (Lin *et al*., 2009). In 1901 the biologically active compound in the illuminating gas was identified as ethylene. This discovery was made by the Russian plant physiologist Neljubov and was based on his experiments on etiolated pea seedlings (Bleecker & Kende, 2000). A few decades later it was shown that plants are also capable of producing ethylene themselves (Gane, 1934). Since its discovery, ethylene and the molecules involved in its signaling pathway have been a great interest to many researchers, not only due to its importance *in planta* but also in agriculture. Currently, ethylene is known to regulate various important developmental aspects and the responses to many environmental factors, making it a prime target for biotechnological applications that aim to produce more long-lasting and pathogen-resistant plants.

As phytohormone ethylene regulates various processes in the plant life cycle from plant growth and senescence of plant organs to response to biological and environmental stress factors as a part of plant immune system (Bleecker & Kende, 2000; Chang & Stadler, 2001; Johnson & Ecker, 1998; Guo & Ecker, 2004; Lin *et al.*, 2009). Ethylene is recognized as a one of the classical phytohormones, which additionally include auxins, abscisic acid, cytokinins and gibberellins (Gaspar *et al.*, 1996). Due to its versitile function, ethylene is produced in virtually all plant tissues, and like other plant hormones, exists by default in low levels in these tissues (Santner *et al*., 2009). Plants are highly

sensitive in ethylene perception and changes in nanomolar range in the ethylene concentration are sufficient to induce a response.

One of the most studied and best-known effects of ethylene is the inducement of fruit ripening. Fruits can be divided into two categories, climateric and non-climateric fruits, based on their mechanism of ripening (Alexander & Grierson, 2002). In non-climateric fruits, such as grape, orange and pineapple, there is very little change in respiration and the levels of ethylene remain low. The opposite is observed for the climacteric fruits, where the onset of ripening is accompanied with sharp increase in ethylene production and a concominant rise in respiration (Bleecker & Kende, 2000). Application of exogenous ethylene in the young fruits in the preclimacteric state has been shown to accelarate the ripening process. Examples of climacteric fruits include several common fruits, like apples, bananas and tomatoes. Due to the key role of ethylene in the ripening of the climacteric fruits, it is of high value and interest also in agricultural applications. The control of ethylene production and the onset of the ripening enable the preservation of fruits in the pre-ripening (preclimacteric) state and prevent them from spoilage during long transportations from the harvesting location to the customer.

Besides fruit ripening, there is a great number of other senescence and growth related processes that are regulated by ethylene in plants. In flowering plants, ethylene-induced responses include flower senescence together with abscission of the leaves and pedals. In some instances, the effect of ethylene in aging of flowering plants is highly similar to the fruit ripening as a significant increase in ethylene production and a concominant rise in respiration is also observed at the onset of flower senescence (Yang & Hoffman, 1984). Ethylene also plays an important role in regulating morphological changes related to growth and in sex determination (Lin *et al*., 2009). Ethylene-induced growth responses include root nodulation, seed germination, the asymmetric growth of

stem and petiole, and the apical hook formation. Ethylene is also important part of plant immunity and regulates responses to pathogen attack in concert with other plant hormones.

In the model plant *Arabidopsis thaliana*, one of the most profound effects of ethylene is observed when dark-grown germinating seeds are exposed to this plant hormone. These seedlings exhibit a phenotype that is characterized by the exaggerated curvature of the apical hook, radial swelling of the hypocotyl and shortening of the hypocotyl and root (Bleecker & Kende, 2000; Chang & Stadler, 2001; Wang *et al*., 2002). This triple response phenotype has provided the means to screen and subsequently identify *A. thaliana* mutants that have altered synthesis of ethylene or sensitivity to it (Wang *et al*., 2002). This genomic tool has been particularly useful in discovering components in ethylene signaling pathway. Based on the response that *A. thaliana* seedlings exhibit upon exposure to ethylene, the identified mutants from the triple response assay can be categorized into three separate groups: ethylene-insensitive mutants, constitutive triple response mutants and tissue-specific ethylene-insensitive mutants (Guo & Ecker, 2004). The mutants of the first group show little or no response when high levels of the plant hormone or its precursor 1-aminocyclopropane-1-carboxylic acid (ACC) are present, and are called *ethylene-resistant* (*etr*), *ethylene-insensitive* (*ein*) or *ACC-insensitive* (*ain*) (Chang & Stadler, 2001; Wang *et al*., 2002). The second group consists of the ethylene-response mutants that show a constitutive triple response even in the absence of the hormone or its precursor. These mutants can be further divided into subgroups depending on whether the phenotype they exhibit can be suppressed by inhibitors of ethylene biosynthesis and perception or not. Mutants of *A. thaliana* that are insensitive to these inhibitors and retain their altered phenotype are referred to as the *constitutive triple response* (*ctr*) mutants (Wang *et al*., 2002). These differ from the plants

with *ethylene-overproducer* (*eto*) phenotype, which revert back to wild type phenotype after application of the inhibitor. *A. thaliana* plants with *eto* morphology have a defect in the regulation of ethylene production (Wang *et al*., 2002). The last group of ethylene-response mutants is the tissue-specific ones, which have more localized effect on ethylene sensing or production. These include examples like *hookless1* (*hls1*) and *ethylene insensitive root1* (*eir1*) (Guzman & Ecker, 1990; Roman *et al*., 1995; Guo & Ecker, 2004)

1.3.2 Ethylene synthesis in plants

The pathway that describes the steps in the synthesis of ethylene in plants is called the methionine or Yang cycle (Yang & Hoffman, 1984) (Figure 1). Ethylene synthesis proceeds through three main reactions (numbered from 1-3 in the Figure 1), in which S-Adenosyl-L-methionine (S-AdoMet) and a cyclic, non-proteinogenic amino acid 1-aminocyclopropane-1-carboxylic acid (ACC) are the key intermediates in turning L-methionine into ethylene (Adams & Yang, 1979). The conversion of methionine to ethylene requires oxygen and ATP, and in addition to ethylene, carbon dioxide (CO2), hydrogencyanide (HCN), water and 5'-methylthioadenosine (MTA) are formed as byproducts. Due to its toxicity, the gaseous hydrogen cyanide is further metabolized into β-cyanoalanine by β-cyanoalanine synthase. The MTA molecule on the other hand is used to preserve the methylthiogroup and to enable the regeneration of methionine in the methionine cycle.

The first step in ethylene biosynthesis is the conversion of L-methionine to S-AdoMet (reaction 1 in Figure 1). Besides its involvement in ethylene synthesis, S-AdoMet is also needed in several other metabolic reactions important to plant growth and development, such as a methyl donor in a number of transmethylation reactions and in the

biosynthesis of spermidine, spermine and biotin (Ravanel *et al.*, 1998). The production of S-AdoMet from L-methionine is an ATP-dependent reaction and catalyzed by the enzyme S-AdoMet synthase (Wang *et al*., 2002).

The formation of S-AdoMet is followed by the rate-limiting step in ethylene biosynthesis that is the conversion of S-AdoMet to ACC (reaction 2 in Figure 1) (Adams & Yang, 1979; Yang and Hoffman, 1984). ACC synthase (ACS) is the enzyme responsible for the synthesis of ACC and requires pyridoxal 5'-phosphate as a cofactor. ACC synthesis proceeds through a γ-elimination reaction, where the MTA-byproduct is further utilized to regenerate methionine (Yang & Hoffman, 1984; Kende, 1993; Bleecker & Kende, 2000). The regeneration of methionine enables the maintenance of ethylene biosynthesis at high rates even when the levels of methionine are low (Yang & Hoffman, 1984; Kende, 1993). ACC is the immediate precursor of ethylene, which arises from the enzymatic activity of ACC oxidase (ACO). The formation of ethylene from ACC is oxygen-dependent and lack of oxygen results in accumulation of ACC and halted synthesis of ethylene. This phenomenon was observed when plants were kept in nitrogen atmosphere and was followed by rapid synthesis of ethylene after exposure of the plant to air.

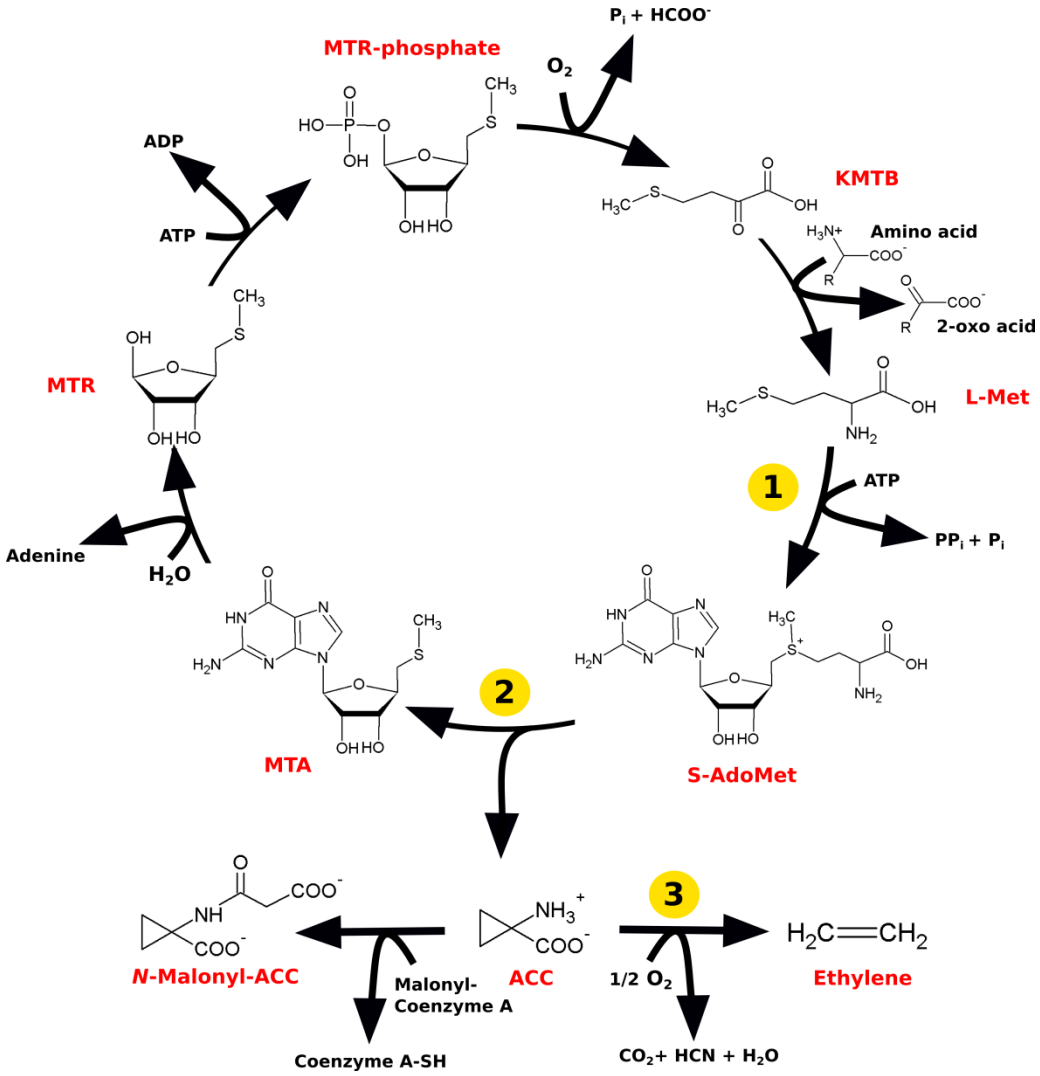

Figure 1. The biosynthesis of ethylene via Yang cycle (Yang & Hoffmann, 1984). The first step in ethylene synthesis is the formation of S-Adenosylmethionine (S-AdoMet) from L-methionine (1). The S-AdoMet is further processed in the rate-limiting step of ethylene synthesis that produces 5'-methylthioadenosine (MTA) and 1-aminocyclopropanecarboxylic acid (ACC) (2). The latter one is the direct precursor of ethylene, which results from the catalytic activity of ACC oxidase in the third and final step of this plant hormone (3). The abbreviations MTR and KMTB stand for 5'-Methylthioribose and α-Keto-γ-methylthio-butyric acid, respectively. Both are intermediate products in the regeneration cycle of L-methionine.

1.3.3 Regulation of ethylene synthesis

All plant tissues are capable of producing ethylene and already nanomolar quantities of ethylene are sufficient to induce a response. Therefore the synthesis of ethylene needs to be tightly regulated. One of the main targets of regulation is the enzyme ACS, which catalyzes the important conversion of S-AdoMet into ACC in the rate-limiting step of ethylene synthesis. The important role of ACS as a regulator of ethylene synthesis is emphasized by the existence of multiple ACS isoforms in a single plant species (Johnson & Ecker, 1998). The variety of ACS isoforms provides the means to control the ethylene synthesis both in protein and gene level. On the protein level this can be observed in the altered rate in ACC production by the different ACS isoforms. This allows the plant to differentiate between steady state and induced synthesis of ethylene. The availability of multiple ACS genes on the other hand enables the differential expression of the encoded protein in response to environmental and developmental signals as well as hormones (Tsuchisaka & Theologis, 2004). It has been shown that in preclimacteric fruits the compound ACC is only present in small amounts and that this is accompanied with low levels of the ACC synthase. Furthermore a significant increase in the amount of both ACC and ACC synthase is detected upon the inducement of ripening process (Adams & Yang, 1977; Adams & Yang, 1979; Kende & Boller, 1981). The phytohormones that have been shown to be involved in the regulation of ACS production include auxin and cytokinin (Johnson & Ecker, 1998). In Arabidopsis these two hormones are involved in promoting the transcription of specific ACS genes that subsequently also enhances ethylene synthesis (Abel *et al*., 1995; Johnson & Ecker, 1998; Vogel *et al*., 1998). Auxin acts by rapidly inducing the expression of ACS4, a member of ACS family in Arabidopsis, whereas cytokinin influences expression of another ACS isoform, ACS5. However, the effect of cytokinin is concentration dependent and only

when the cytokinin levels are low, the ethylene-induced phenotype can be observed (Vogel *et al*., 1998). The ACS levels in the plant tissues are also regulated by a group of BTB (broad-complex, Tramtrack and bric-a-brac) ubiquitin ligases, called Ethylene overproducer 1 (ETO1), ETO1-like1 (EOL1) and EOL2, all of which promote the degradation of ACS proteins (Santner *et al*. 2009). Additional diversity to the ACS-mediated regulation of ethylene synthesis arises from the ability of these enzymes to form dimers (Lin *et al*., 2009). Through both homo- and heterodimerization, the different ACS isoforms are capable of fine-tuning their responses.

There are also a number of chemicals, which inhibit ethylene synthesis when applied exogenously. These include compounds that effect on the membrane structures, such as phosphaditylcholine, Tween 20 and Triton X-100 but also ones that suppress the auxin-induced increase in ethylene synthesis, i.e. cycloheximide, actinomycin D and α-amanitin. Also ethylene itself can function as an inhibitor of its own synthesis. It has been shown that in flavedo tissues of citrus fruit the wounding-induced ethylene production was decreased in the presence of exogenous ethylene (Riov & Yang, 1982). Since the application of ACC was able to recover the loss in the synthesis rate of ethylene, it was suggested that inhibitory effect would be mainly targeted on the activity of ACC synthase (Riov & Yang, 1982).

1.3.4 Arabidopsis thaliana ethylene signaling pathway

The ethylene signaling is initiated by a family of membrane-bound receptors that in *A. thaliana* consists of five members: Ethylene Receptor 1 and 2 (ETR1 and 2), Ethylene Response Sensor 1 and 2 (ERS1 and 2) and Ethylene Insensitive 4 (EIN4). These are further classified into two subfamilies, one consisting of ETR1 and ERS1 (subfamily I), and the other including the remaining three receptors (ETR2, ERS2 and

EIN4; subfamily II) (Hua *et al*., 1998; Chang *et al*., 2004). All five Arabidopsis ethylene receptors have a similar overall architecture, which includes an N-terminal trans-membrane domain and a cytosolic part with a GAF (cGMP-specific phosphodiesterase, adenylyl cyclases and FhlA) domain, a histidine kinase and in some case a receiver domain (Moussatche & Klee, 2004). Furthermore, they are known to share sequence similarity with the prokaryotic two-component systems, which regulate the adaptive responses to environmental stimuli in their host organism and typically consist of a membrane-associated sensor with cytosolic histidine kinase domain and a separate response regulator (Chang *et al*., 1993; Lin *et al*., 2009). Due to the similarity with these bacterial two-component signaling molecules, ethylene signaling has also been postulated to involve similar His-Asp phosphorelay in the signal transfer from cytosol to nucleus as observed in the bacterial system. However, its role as the main signaling mechanism remains speculative since only the receptors in the subfamily I contain all the conserved histidine kinase motifs found in the prokaryotic two-component sensor kinases (Hua *et al*., 1998; Mason & Schaller, 2005). The subfamily II receptors are more diverged in their kinase region and have been shown to possess Ser/Thr kinase activity instead (Hua *et al.*, 1998; Chang *et al.*, 2004, Moussatche & Klee, 2004).

The amphiphilic nature of ethylene enables the unusual location of ethylene receptors at the endoplasmic reticulum (ER) (Chen *et al*., 2002). The functional form of ETR1 is disulfide-linked dimer that requires a copper co-factor for ethylene binding (Schaller *et al*., 1995; Hirayama & Alonso, 2000). This co-factor is provided by a P-type ATPase copper transporter Responsive-to-Antagonist 1 (RAN1), which despite its location at the Golgi apparatus has been shown to directly interact with ETR1 (Hirayama *et al*., 1999). The binding site for both the copper and for the ethylene is embedded in the ER-membrane and locates between the N-terminal transmembrane domains of the

receptor dimer. The ethylene binding presumably causes a conformational change that allows the signal to be transmitted to the cytosolic part of the receptor and further on to the downstream targets.

The cytosolic portion of the ethylene receptors has been shown to directly interact with a putative Mitogen-activated protein kinase kinase kinase (MAPKKK). This downstream component of ethylene-signaling pathway was identified from an *A. thaliana* mutant carrying a recessive mutation that caused constitutive ethylene responses even in the absence of the hormone (Kieber *et al*., 1993). The responsive gene, *constitutive triple response 1* (*ctr1*) encodes a member of a Raf-1 like mitogen-activated protein kinase kinase kinase (MAPKKK) family. CTR1 comprises an N-terminal regulatory domain that interacts with the ethylene receptor and a C-terminal catalytic domain with serine/threonine kinase activity. As a MAPKKK, CTR1 is assumed to take part in a canonical MAPK cascade, which in eukaryotic cells mediates the perception of external stimuli to cellular responses. Therefore the discovery of CTR1 revealed an intruiging feature in ethylene signaling since it would represent an unusual combination of prokaryotic and eukaryotic signaling components.

In the absence of ethylene, both the ethylene receptors and CTR1 are active and repress the downstream signaling in a phosphorylation dependent manner (Figure 2, left panel). In its active state, CTR1 phosphorylates another ER-bound member of ethylene signaling protein family, Ethylene Insensitive 2 (EIN2). EIN2 is a membrane protein with an N-terminal domain that resembles NRAMP (natural resistance-associated macrophage protein) metal ion transporters and a cleavable, cytosolic C-terminal domain (Alonso *et al*. 1999; Bisson *et al*. 2010; Wang *et al*. 2009). EIN2 has shown to be critical in ethylene signaling and to act as a positive regulator of ethylene response. In its phosphorylated form, EIN2 is subjected to ubiquitin-dependent degradation by the 26S proteosome. Upon

binding of ethylene to the receptor, both the receptor and CTR1 become inactive thus preventing the phosphorylation of EIN2 (Figure 2, right panel). This results in targeted proteolytic clevage of the EIN2 C-terminal domain and its re-location to the nucleus where it activates a transcription factor cascade with yet unknown mechanism. An important group of transcription factors that function downstream of EIN2 includes Ethylene Insensitive 3 (EIN3) and EIN3-like (EIL) proteins (Chao *et al*., 1997). These transcription factors have been shown to bind to a specific region in the promoter of *ethylene response factor 1* (*erf1*) and promote the expression of the encoded transcrition factor. ERF1 is a GCC box-binding transcription factor that belongs to the Ethylene-Response-Element-Binding-Protein (EREBP) protein family. The binding of ERF1 to the promoters of the secondary target genes eventually leads to activation and transcription of the ethylene response genes (Zhao & Guo, 2011).

Other proteins involved in ethylene signaling include EIN3-targeting F-box proteins 1 and 2 (EBF1 and EBF2), both of which target EIN3 to degradation in absence of ethylene (Wang *et al.* 2009). It is proposed that a protein called Ethylene Insensitive 5 (EIN5) facilitates the turnover of these two F-box proteins (Olmedo *et al.* 2006). However, the mechanism behind this is currently not known. Also the amount of EIN2 at protein level is controlled by two F-box proteins, EIN2 targeting protein 1 and 2 (ETP1 and 2). In absence of ethylene they target EIN2 to degradation through ubiquitination and interaction with Skp-1/Cullin-1 proteasome complex.

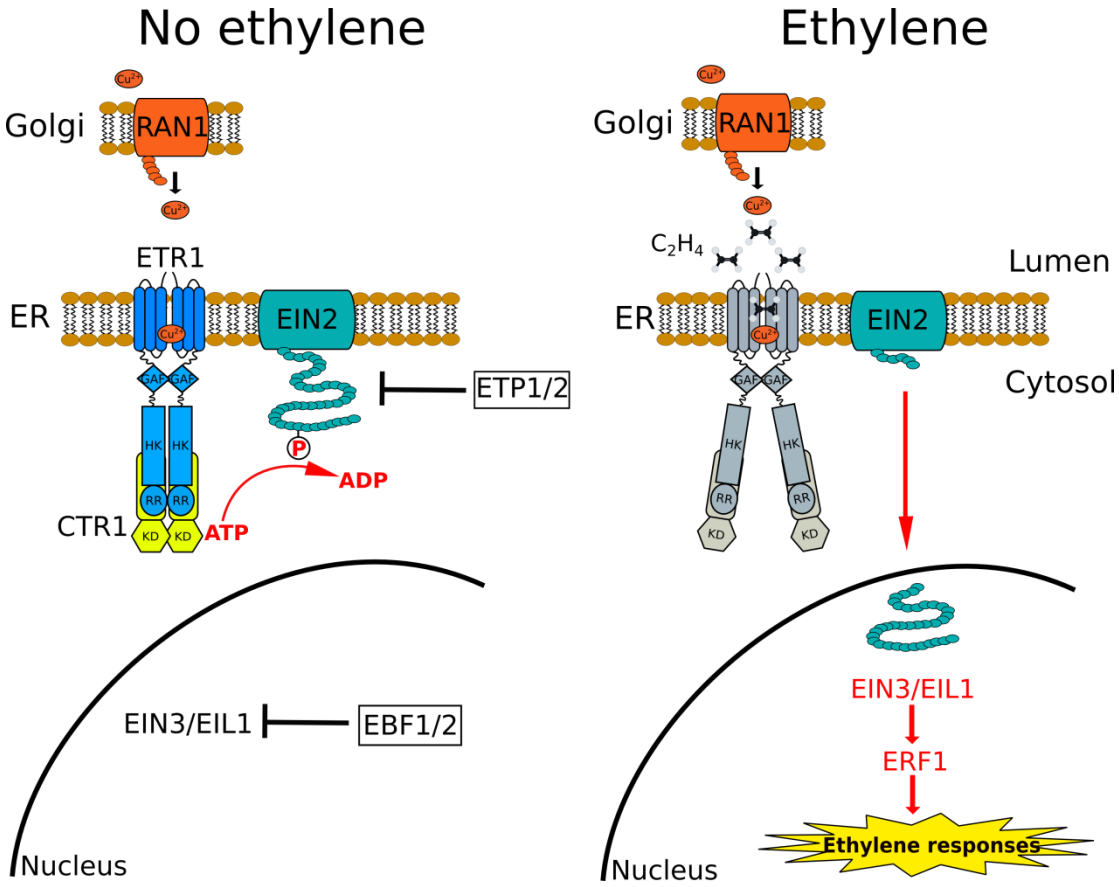

Figure 2. Schematic and simplified representation of the ethylene-signaling pathway in absence and presence of the plant hormone. In the absence of the plant hormone, both the ethylene receptor (represented by ETR1) and its cytosolic interaction partner CTR1 are active, and together suppress the downstream signaling components in a phosphorylation-dependent manner (left panel). Ethylene binding inactivates both the receptor and CTR1, which allows the C-terminal domain of EIN2 to be proteolytically cleaved and relocated to the nucleus. This results in the activation of transcription factors, such as EIN3 and ERF1 and finally to the triggering of ethylene responses.

1.3.5 Enhanced Disease Resistance 1

The gene for enhanced disease resistance 1 (*edr1*) was first discovered from *A. thaliana* mutant plant that showed resistance to the fungus *Erysiphe cichoracearum*, the causative agent for powdery mildew (Frye & Innes, 1998). EDR1 is a serine/threonine-protein kinase that belongs to the superfamily of Raf-like MAPKKKs (Frye *et al*., 2001).

Its closest homolog is the well-known *A. thaliana* ethylene-signaling molecule CTR1. Like CTR1, EDR1 consists of an N-terminal domain with a putative regulatory function and C-terminal catalytic domain with the kinase activity. EDR1 acts as a negative regulator of disease resistance and programmed cell death in a salisylic acid dependent manner (Frye *et al*., 2001; Tang *et al*., 2005). The *edr1* mutant plant also shows enhanced drought-induced growth inhibition, enhanced ethylene-induced senescence and enhanced sensitivity to the plant hormone abscisic acid (ABA) therefore linking EDR1 to at least two other plant hormone signaling pathway (Frye *et al*., 2001; Tang *et al*., 2005).

Ubiquitin-mediated protein degradation has been suggested as a control mechanism in the EDR1 mediated signaling pathway based on the involvement of F-box protein ORE9 in the drought-induced growth inhibition and ethylene-induced senescence phenotypes (Tang *et al*., 2005). EDR1 is also known to interact with KEG (Keep On Going), a RING finger E3-ubiquitin ligase that negatively regulates the protein levels of the transcription factor ABI5 (Abcisic acid Insensitive 5) (Wawrzynska *et al*., 2008; Gu & Innes 2011). The interaction of EDR1 with KEG was shown to be important for EDR1 function since a missense mutation in *keg* suppresses all EDR1-mediated phenotypes (Wawrzynska *et al*., 2008). It has been proposed that KEG could serve as a substrate for EDR1 and EDR1 might therefore regulate its downstream components by KEG-mediated ubiquitination (Gu & Innes 2011).

Recently, EDR1 was found to negatively regulate the MKK4/MKK5-MPK3/MPK6 kinase cascade, which is required for the basal defense responses in plants (Zhao *et al*., 2014). It was shown that EDR1 directly interacts with MKK4 and MKK5 via its N-terminal domain and additionally suppresses the protein levels of MKK4, MKK5, MPK3 and MPK6.

1.4 Non-Ribosomal peptide synthetase

1.4.1 Non-Ribosomal peptides

A peptide is defined as a compound that consists of two or more amino acids joined covalently together by an amide bond, also known as peptide bond. The main system for the synthesis of naturally occurring peptides or small proteins relies on the translation of the nucleotide sequence in mRNA by the complex assembly of macromolecules, called the ribosome. Prior to reaching the final and functional form, the peptides of ribosomal origin usually go through a proteolytic processing from their gene-encoded precursors. In addition to proteolysis, the produced peptides can be subjected to post-translational modifications, such as glycosylation, dehydration and addition of neighboring sulfhydryl groups to form thioesters. The peptides formed this way have a wide range of important biological functions, both in the source organism and in industrial applications. Examples of important ribosomal peptides include the vertebrate hormones oxytocin, bradykinin and thyrotropin. Also the group of lanthionine-containing peptide antibiotics, also referred as lantibiotics, that are produced mainly by Gram-positive bacteria belong to the category of ribosomal peptides (Sahl *et al.*, 1995). Known lantibiotics include nisin, subtilin and epidermin, which have industrial use as food preservatives (Marahiel *et al.*, 1997). However, there are several antimicrobial/biologically active peptides of microbial origin that are produced in a ribosome-independent manner.

Non-ribosomal peptide synthesis is a unique feature of mainly fungal and bacterial organisms that include members of the Gram-positive *Actinomycetes* and *Bacilli*, cyanobacteria and species of eukaryotic filamentous fungi, such as *Aspergillus* (Mootz *et al.*, 2002). These microorganisms use an alternative system to produce these peptides, a

large multimodular enzyme known as the non-ribosomal peptide synthetase (NRPS). The NRPS enzymes consist of several modules that contain the necessary enzymatic units to recognize, activate and add the residue into the growing peptide chain (Marahiel *et al*., 1997). The order and number of these modules define the sequence and the length of the end product, respectively. The length of the final non-ribosomal peptide can vary between 2 and 48 residues, but on average their size is limited to 4 – 10 residues (Konz & Marahiel, 1999; Gewolb, 2002).

The nonribosomal peptides form a group of molecules with remarkable structural diversity. In addition to the system responsible for their synthesis, the feature that differentiates these peptides from their ribosomal counterparts is the significantly greater selection of starting materials. The synthesis of ribosomal peptides is limited mainly to the classical proteinogenic amino acids and the existance of non-proteinogenic/unnatural amino acids usually arises from post-translational modifications. The non-ribosomal peptides are known also to incorporate many unusual, non-proteinogenic residues, which increase the number of available monomers to over 300 (Konz & Marahiel, 1999). These include D-enantiomers of the common amino acids, N-methylated amino acids, hydroxy acids and fatty acids, among others (Konz & Marahiel, 1999; Kleinkauf, 1979; Walsh, 2008). The structural variety of nonribosomal peptides is further increased by modification of the peptide product through acylation, glycosylation or heterocyclic ring formation (Marahiel *et al*., 1997; Konz & Marahiel, 1999; Walsh, 2004). The end product of nonribosomal peptide synthesis can be linear, cyclic or branched but as a common factor they exhibit a constrained structure. The rigidity of these compounds is most often derived from cyclization of the peptide product through formation of macrocyclic lactams or lactones (Sieber & Marahiel, 2003; Mootz *et al*., 2002). Other ways include oxidative cross-linking and heterocyclization (Sieber & Marahiel, 2003). The restricted

conformation of the peptide product enables the selective recognition by biological targets and also protects the peptide from degradation by proteases (Gewalb, 2002; Sieber & Marahiel, 2003; Walsh, 2004).

The non-ribosomal peptides are usually produced as secondary metabolites in their source organism and are proposed to be involved in coordination of growth and differentiation as signaling molecules, in the breakdown of cellular metabolites, in iron uptake as chelating siderophores, in resistance to oxidative stress, in virulence as effector molecules and in defence as antimicrobial compounds (Lee *et al*., 2005; Williams *et al*., 1989; Gewalb, 2002). Many non-ribosomal peptides also possess biological properties that make them highly useful in medicinal, agricultural and biological research (Finking & Marahiel, 2004; Marahiel *et al*., 1997). For the pharmaceutical industry non-ribosomally produced peptides have important applications as antibiotics, immunosuppressants, antitumor or antifungal drugs (Marahiel, 1997). Examples of nonribosomal peptides with antimicrobial properties include bacitracin from *Bacillus licheniformis*, daptomycin from *Streptomyces roseosporus*, vancomycin from *Amycolatopsis orientalis* and tyrocidine A from *Bacillus subtilis* (Konz *et al*., 1997; Nagarajan, 1991; Wittmann *et al*., 2008; Hu *et al*., 2009). Also the tripeptide δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV), the precursor molecule of the antiobiotics in the penicillin and cephalosporine families, is synthesized by NRPS machinery (Finking & Marahiel, 2004). Cyclosporin A is a powerful immunosuppressant that enabled organ transplantation whereas bleomycin is a potential cancer drug (Kunz & Hall, 1993; Bierer *et al*., 1993; Suzuki *et al*., 1968). The biological applications of the nonribosomal peptides have made the machinery that produces them an intriguing research topic through the possibility to discover compounds with novel or enhanced biological activities. Especially in the modern world where the number of bacterial strains resistant

to known antibiotics is increasing, there is an urgent need for novel antimicrobial compounds.

1.4.2 Non-Ribosomal peptide synthetases

The broad spectrum of biological activities displayed by the nonribosomal peptides has made their mechanism of biosynthesis and the enzyme responsible for it, the non-ribosomal peptide synthetase an attractive target for the scientific research. The non-ribosomal synthesis of peptides is mainly encountered in soil-inhabiting microorganisms, such as members of the Gram-positive *Actinomycetes* and *Bacilli* but also in eukaryotic filamentous fungi and in other bacterial species including cyanobacteria (Mootz *et al*., 2002). Several bacterial and fungal operons that encode the large modular enzymes for nonribosomal synthesis have been analyzed using DNA sequencing in order to get information about the molecular architecture of these enzymes (Marahiel, 1997; Marahiel *et al*., 1997). In bacteria, several distict synthetases, encoded in separate gene operons, are usually involved in forming the final peptide product. An example of this is the synthesis of bacitracin in *Bacillus lichenifomis*, where three separate synthetases, BA1 (5 modules, 598 kDa), BA2 (2 modules, 297 kDa) and BA3 (5 modules, 723 kDa) are required to assemble the fully functional peptide. The fungal NRPSs on the other hand are usually formed by one large polypeptide, which is encoded in a single gene. For example the synthesis of the immunosuppressive drug Cyclosporin A is catalyzed by one large enzyme, Cyclosporin synthetase (1600 kDa) from *Tolypocladium niveum* (Marahiel, 1997).

Despite of the high structural diversity of their end products, most of the Nonribosomal peptide synthetases employ a common mode of synthesis known as the multienzyme thiotemplate mechanism (Kleinkauf & von Döhren, 1990; Stein *et al*., 1996;

Marahiel *et al*., 1997). This mechanism describes the formation of peptide bond in the peptide synthetase as a two-step process where the amino acid is first activated through the formation of aminoacyl adenylate intermediate, followed by a transfer of the activated substrate to the 4'-phosphopantheine thiol in a separate domain for the subsequent chain elongation. The basis of the synthesis mechanism lies in the modular organization that replaces the need for a specific template to define the sequence of the peptide product. The nonribosomal peptide synthetases consist of modules, each of which is responsible for recognition and activation of a specific monomeric residue, and the elongation of the growing peptide through condensation of the activated residue with the one of the adjacent module. Each module in turn consists of domains that form functionally independent enzymes carrying out a specific step in the peptide synthesis.

The basic module of NRPS proteins consists of three core domains that are the adenylation (A) domain, thiolation/peptide carrier protein (PCP) domain and the condensation (C) domain (Figure 3) (Strieker *et al.*, 2010). These are found from each elongation module and catalyze the key reactions in the elongation phase of the peptide synthesis. The first and last module of each NRPS, referred to as initiation and termination module, differ from the elongation module in that the former lacks the C domain and the latter contains an additional termination (TE) domain that is responsible for releasing the ready peptide. The growing peptide can also be subjected to variety of modifications to achieve specific chemical or biological properties, thus further increasing the diversity of these compounds. These modifications can take place either during synthesis or after release of the peptide. Specific domains within the certain modules are responsible for the alterations during synthesis, and include examples like the epimerase (E), cyclization (Cy), methyltransferase (MT) and oxidation (Ox) domains (Hur *et al*., 2012). Modifications that occur after the release of the peptide are catalyzed

by separate enzymes and include glycosylation, halogenation and oxidative cross-linking (Hur *et al*., 2012).

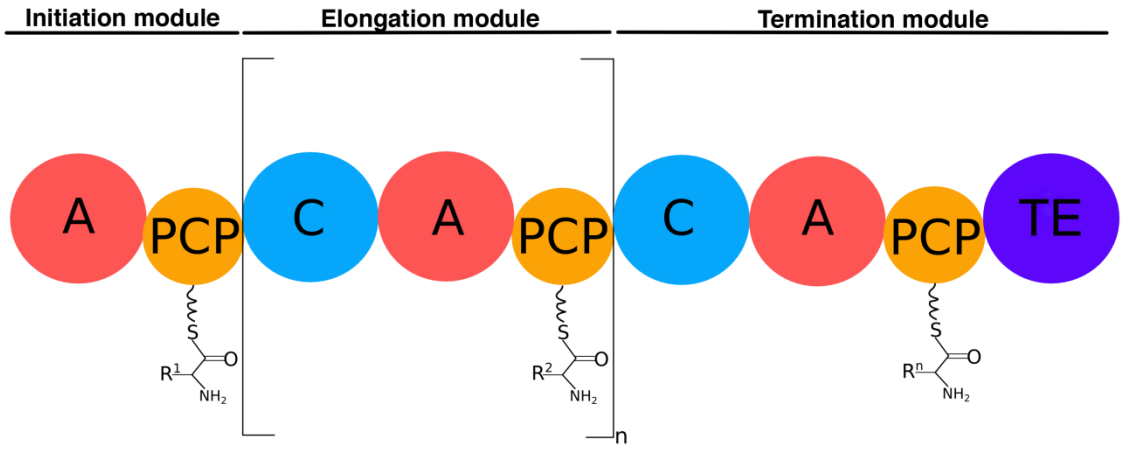

Figure 3. The modular architechture of Non-ribosomal peptide synthetases. One module contains all the necessary enzymatic units for recognition and activation of the amino acid substrate and for its incorporation to the growing peptide chain. These enzymatic units include the adenylation (A) domain, peptide carrier protein (PCP) and condensation (C) domain. Additionally the last module in the NRPS contains a termination (TE) domain that is responsible for the release of the peptide product.

The adenylation (A) domain is the enzymatic unit of about 550 amino acids in length that is responsible for the recognition and activation of its cognate monomeric residue from the pool of available substrates (Konz & Marahiel, 1999). The activation of the target molecule occurs through adenylation, in which the carboxyl group of the substrate reacts with the pyrophosphoryl group of ATP, releasing inorganic pyrophosphate to drive the reaction forward (Figure 4) (Novelli, 1967). Due to its important role in defining the substrate specificity in each module, the A domain is possible the most extensively studied unit in the NRPS system. Engineering the substrate specificity of A domains provides the means to design new peptide products (Marahiel *et al*., 1997).

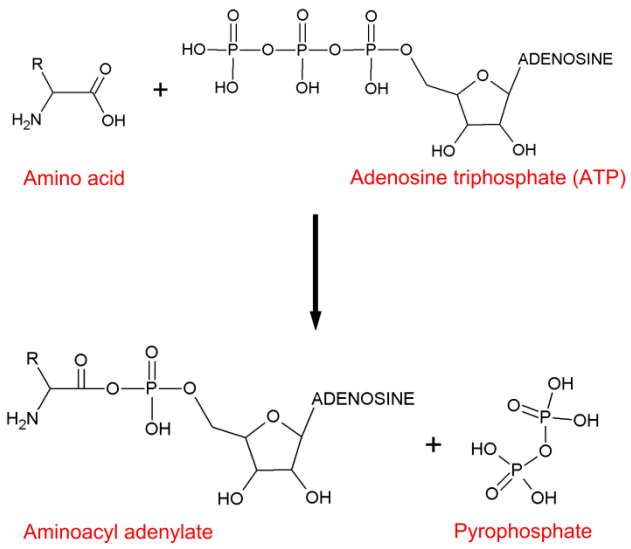

Figure 4. Activation of amino acid by adenylation. The A domain catalyzes the adenylation reaction, which results in formation of aminoacyl adenylate. It involves a formation of phosphodiester bond between the aminoacid and AMP molecule derived from hydrolysis of ATP.

Insights into the substrate binding of A domains was achieved when the first crystal structure was solved for the phenylalanine adenylation subunit of Gramicidin S synthetase 1 (PheA) in the presence of both AMP and L-phenylalanine (Figure 5A) (Conti *et al*., 1997). Although the structure was obtained in the presence of L-isoform, the PheA enzyme is able to activate both stereoisomers of its phenylalanine substrate (Stachelhaus & Marahiel, 1995). Based on the structural information from PheA and sequence alignment of PheA with other adenylation domains, the ATP-binding appeared to be conserved and mediated by shared sequence motifs (Stachelhaus *et al*., 1999). The residues involved in forming the binding pocket for the amino acid substrate were found to locate within a stretch of 100 amino acids (Stachelhaus *et al*., 1999). Combined information from sequence alignment between PheA and 160 other adenylation domains and from the available PheA structure led to the identification of 10 residues that were important for the binding specificity of the A domains (Figure 5B) (Stachelhaus *et al*., 1999). This so-called specificity conferring code was proposed and also shown to be

useful tool in predicting or altering the substrate specificity of adenylation domains (Stachelhaus *et al*., 1999, Eppelmann *et al*., 2002).

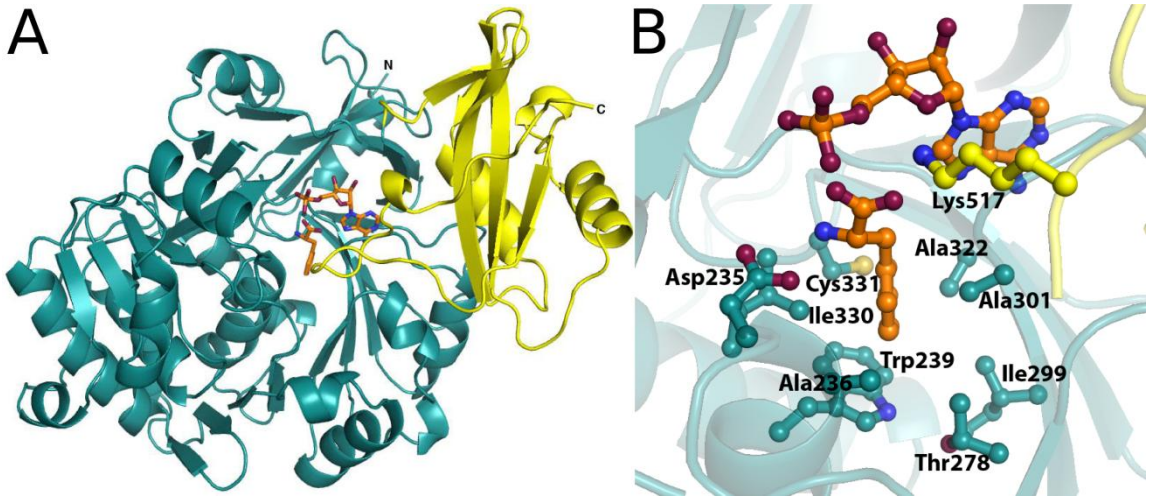

Figure 5. The first solved A domain structure. A) The crystal structure for the phenylalanine adenylation subunit of Gramicidin synthetase 1 (PheA, PDB code: 1AMU) represented the structural organization of A domains (Conti *et al*., 1997). The structure exhibits a two-domain organization consisting of a larger N-terminal domain (dark green) and a smaller C-terminal domain (yellow), connected to eachother with a flexible linker. The N- and C-terminus are marked in the image. B) The binding specificity of phenylalanine is conferred by a set of 10 residues that surround the amino acid substrate. The residues in corresponding positions in other adenylation domains form the specificity conferring code, which can be used to predict and modify the substrate specificity of the adenylating enzyme (Stachelhaus *et al*., 1999).

The central domain in the mediating the nonribosomal peptide synthesis is the small peptidyl carrier protein (PCP), also known as thiolation (T) domain, consisting of roughly 80 – 100 amino acids (Marahiel *et al*., 1997; Lai *et al*., 2006; Hur *et al*., 2012). The PCP domain is located between the adenylation and condensation domain to which it is fused from its N- and C-termini, respectively (Koglin *et al*., 2006). After the carboxylic acid containing residue has been activated as adenylate intermediate, a nucleophilic attack by a reactive thiol group in 4'phosphopantheine cofactor of PCP domain is followed. This results in tethering of the monomeric substrate to the PCP domain as a thioester. The

4’phosphopantheine linker functions like a swinging arm, transferring the activated residue from the adenylation domain to the condesation domain for the peptide bond formation.

The 4’phosphopantheine prostetic group is derived from Coenzyme A (CoA) and is attached to a conserved serine residue in the PCP domain in Mg^{2+}-dependent reaction that is catalyzed by Sfp-type phosphopanthetheinyl transferase (PPTase) (Hur *et al*., 2012). The action of PPTase is needed to convert the inactive *apo*-PCP into its active *holo*-form (Koglin *et al*., 2006; Lai *et al*., 2006; Hur *et al*., 2012). The structural studies using Nuclear Magnetic Resonance (NMR) spectroscopy on the PCP domain from the third module of tyrocidine A synthetase revealed that both the *apo* and the *holo* form of the enzyme were able to maintain two different conformational configurations (Koglin *et al*., 2006). Furthermore it was concluded that two of these four conformations represented the individual apo and holo forms whereas the remaining two were virtually identical and was considered to be common state between the apo and holo form. The three individual conformational states were termed A, A/H and H state (Koglin *et al*., 2006) (Figure 6). The A and H state represent the individual apo-PCP and holo-PCP forms of TycC3-PCP, whereas the A/H conformation is the shared one between the two forms. The canonical four-helix bundle, which is observed in many carrier proteins, is found from the structure of the TycC3-PCP in A/H state whereas the A and H state part of the helical structure unravels (Koglin *et al*., 2006; Hur *et al*., 2006). Further experiments showed that the shift from A/H to H conformation results in movement of the 4’-phosphopantheine cofactor from N-terminal to C-terminal end of the holo-TycC3-PCP, thus supporting the swinging arm mechanism for the transfer of the activated residue from adenylation domain to condesation domain in the peptide synthesis process (Hur *et al*., 2006). It was also shown that the PPTase prefers the A conformation when loading the PCP domain with the 4’-

phosphopantheine cofactor whereas another interaction partner of PCP domains, the type II thioesterase (TEII) only interacts with the H-PCP (Koglin *et al*., 2006). The TEII is an enzyme required for the regeneration of PCP if mispriming of the 4'-phosphopantheine with incorrect adenylate intermediate occurs (Schwarzer *et al*., 2002).

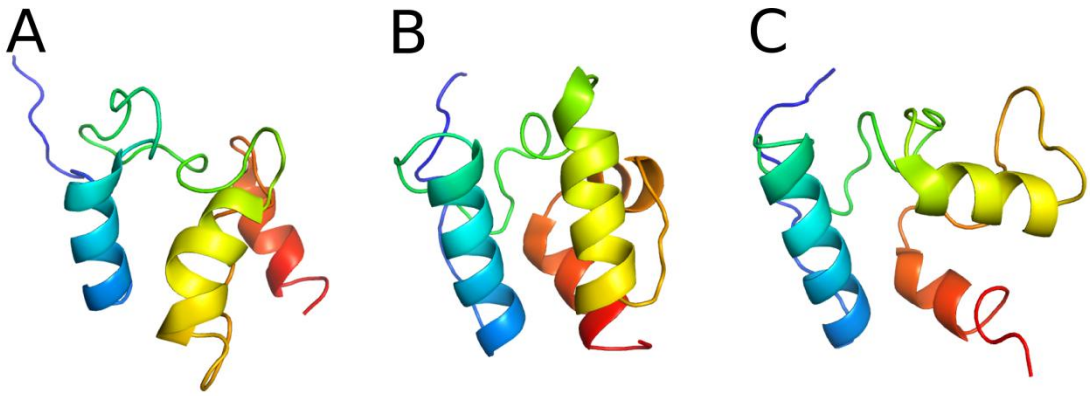

Figure 6. The NMR structures of TycC3-PCP domain in three different conformations. A) The A state (PDB code: 2GDY) is only seen for the inactive *apo*-form of the enzyme. B) The A/H state (PDB code: 2GDW) is proposed to be the intermediate state between the *apo*- and *holo*-forms. C) The H-state (PDB code: 2GDX) corresponds the active *holo*-form. The conformational changes are mediated by the three core helices, αI (blue), αII (yellow) and αIII (red). In the A/H state, an additional α-helix (αIV) appears resulting in the formation of canonical four-helix bundle, observed in many carrier proteins among others.

The step leading to the elongation of the chain is the formation of the peptide bond, which is catalyzed by the condensation (C) domain. Like the A and PCP domains, the C domain is required in each module, with the exception of the initiation module, and is located between two adjacent activating units in the same peptide synthetase polypeptide (Marahiel *et al*., 1997). In some modules the C domain is replaced by Cy-domain, which catalyzes the formation of oxazoline or thiazoline through either intermolecular condensation or intramolecular heterocyclization reaction (Keating *et al*., 2002). The C domain contains a donor site, dedicated for the peptide attached to the PCP

domain of the preceeding module, and an acceptor site for the monomeric residue attached to the PCP domain in the same module (Challis & Naismith, 2004). The reaction in the C domain proceeds through a nucleophilic attack of the α-amino group of the substrate in the acceptor site onto the thioester group of the substrate in donor site, followed by the amide bond formation (Marahiel & Essen, 2009). Since the growth of the peptide chain progresses unidirectionally from N-terminus to C-terminus, the entire peptide is transferred from the upstream donor site to the downstream acceptor site after the peptide bond is formed. The acceptor site has shown to be more substrate specific and even discriminating between different enantiomers whereas the donor site as the binding site for the growing peptide exhibits less strict specificity towards its substrates (Belshaw *et al*., 1999; Linne & Marahiel, 2000).

Insight into the structural characteristics of C domain has been obtained through the crystal structures of VibH, a stand-alone condensation domain from *Vibrio cholerae* vibrioactin synthetase (Figure 7), and of the didomain PCP-C from modules 5 and 6 of tyrocidine synthetase TycC (Keating *et al*., 2002; Samel *et al.*, 2007). In both structures the enzyme consists of N-terminal and C-terminal domains that form a V-shaped fold with the active site at the junction of the two domains. The two subdomains are structurally very similar and based on their sequnce are related to enzymes of the acetyl coenzyme A dependent acetyl transferase superfamily, such as chloroamphenicol acetyl transferase (CAT) (Keating *et al*., 2002; Samel *et al.*, 2007; Challis & Naismith, 2004). The domains exhibit a αβα sandwich conformation where one of the β-strands from each domain contributes to complete a β-sheet in the other (Keating *et al*., 2002). The active site was found to rather be a canyon-like formation than two distinct pockets as was expected (Keating *et al*., 2002; Samel *et al.*, 2007). The opposite faces of the canyon would serve as the acceptor and donor sites so that their substrates point towards a highly

concerved catalytic histidine motif, HHxxxDG, which contains a catalytically important histidine residue (Marahiel *et al*., 1997; Keating *et al*. 2002; Samel *et al*., 2007; Marahiel & Essen, 2009). This histidine is proposed to function as a general base and deprotonate the α-aminogroup for the nucleophilic attack.

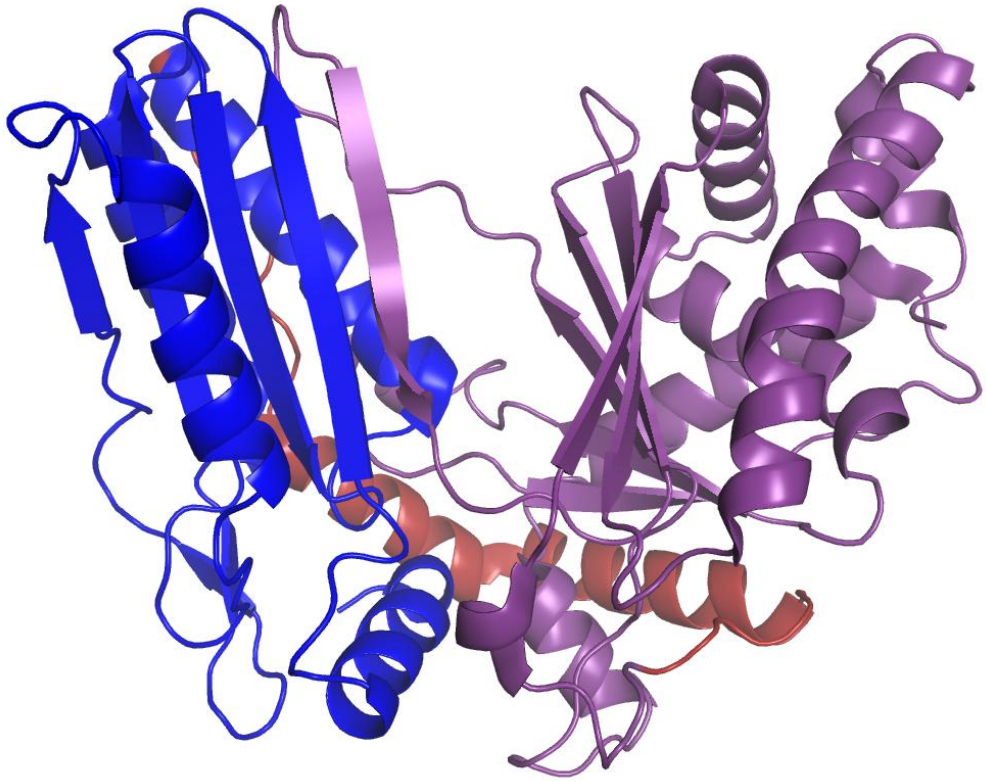

Figure 7. The crystal structure of VibH from *Vibrio cholerae* (PDB code: 1L5A) as a representative of the C domain in NRPS system (Keating *et al*. 2002). The enzyme adopts a V-shaped conformation with the active site located in the canyon-like formation between the N-terminal (blue) and C-terminal (purple) domains. The two domains are connected with 56-residue helical linker, which is depicted here in red.

After the synthesized peptide has reached its final length, it is released from the NRPS machinery. The release of the peptide is usually catalyzed by a dedicated domain at the C-terminal module of the peptide assemply line, known as the TE domain (Sieber & Marahiel, 2005). The TE domain cleaves the peptide product from the 4'phosphopantheneyl cofactor of PCP of the terminal module in a two-step process that leads into linear (penicillin precursor ACV), cyclic (e.g. tyrocidine) or branched-cyclic

(e.g. bacitracin) products. The first step is the transfer of the peptide to the highly conserved serine residues, located within the GXSXG signature motif of TE domain (Konz & Marahiel, 1999; Shen *et al*., 2004). In this step the hydroxyl group of the serine acts as a nucleophile and attacks the thioester bond between the peptide and the 4'phosphopantheneyl cofactor, resulting in the formation of peptidyl-O-TE intermediate (Lautru & Challis, 2004; Marahiel & Essen, 2009). The second step results in the release of the peptide either through hydrolysis to produce a linear peptide or through intramolecular macrocyclization to create a cyclic end product (Marahiel & Essen, 2009; Hur *et al*., 2012). The latter one is more commonly employed by the TE domains in most NRPS systems and results from the attack of internal nucleophile, such as a hydroxyl – or aminogroup in the peptidyl chain, leading to a cyclic lactone or lactam, respectively (Shen *et al*., 2004; Hur *et al*., 2012). The function of the TE domain can be replaced by the activity of NADPH-dependent reductase (R) domain, which catalyzes the release of linear aldehydes or alcohols (Konz & Marahiel, 1999; Marahiel & Essen; 2009). This type of peptide release mechanism has been encountered for example in the biosynthesis of linear Gramicidin A in Bacillus brevis (Sieber & Marahiel, 2005).

1.4.3 Structural features of adenylating enzymes

Adenylation is a chemical reaction where a carboxylic acid group, usually arising from an amino acid moiety, is converted into more reactive phosphodiester group through condensation with ATP (Novelli, 1967). The released pyrophosphate functions as the driving force in this reaction where one mole of water is produced as a byproduct. The resulting phosphodiester linkage between the amino acid and AMP is very reactive and usually the aminoacyl adenylate is used in secondary reactions either by the same enzyme or a different one (Gulick, 2009; Schmelz & Naismith, 2009). There are a number of

enzymes capable of adenylation, arising also from very different functional backgrounds. These comprise the adenylation subunit of Non-Ribosomal peptide synthetases, acyl- and aryl-CoA synthetases, luciferase oxidoreductases, aminoacyl-tRNA synthetases and enzymes involved in NRPS-independent siderophores (NIS) synthesis (Schmelz & Naismith, 2009). The common feature of these enzymes is that they all activate their carboxylic acid containing substrate through adenylation. However, already before structural information was available a sequence-based analysis showed that the adenylation subunit of NRPS proteins, acyl- and aryl-CoA synthetases, and luciferases belong to a same enzymatic superfamily on the basis of the numerous shared concerved sequence motifs (McElroy *et al*., 1967; Babbitt *et al*., 1992; Turgay *et al*., 1992a; Turgay *et al*., 1992b). Together these three enzymes form the superfamily of adenylating enzymes that in addition of shared sequence motifs, possess common structural characteristics.

In both adenylation subunit of NRPS and in acyl-or aryl CoA synthetases the activation via adenylation is followed by thioester-forming reaction where the adenylated molecule is attached to a pantheine thiol group (Gulick, 2009). However, in the NRPS machinery the thioester formation requires the transfer of the adenylate intermediate from the A domain to the PCP domain, which contains the required 4'-phosphopantheine thiol prostetic group. This differs from the acyl-CoA synthetases, which catalyze both adenylation and the thioester formation to produce acetyl-CoA (Gulick, 2009). The follow-up reaction after adenylation catalyzed by luciferases differs from the ones of its enzymatic family members. The lyciferyl-adenylate first goes through an oxidative decarboxylation after which the resulting intermediate decomposes within the active site to produce a photon of light.

The first crystal structure from the family of adenylating enzymes was obtained for the firefly luciferase from *Photinus pyralis* (Figure 8) (Conti *et al*., 1996). The protein consists of two domains, a larger N-terminal domain and a smaller C-terminal one, which are linked to each other by a flexible linker. The N-terminal domain is further divided into three subdomains, which include two β-sheet subdomains and a β-barrel subdomain (Conti *et al*., 1996). The two β-sheet subdomains form together a five-layered sandwich-like αβαβα tertiary fold where two helices lie in the middle of the sandwich surrounded by the two eight-stranded β-sheets with the remaining helices packed against the β-sheets. The C-terminal domain on the other hand consists of two-stranded antiparallel β-sheet and central three-stranded β-sheet that is surrounded by three α-helices. The overall extended conformation of the unliganded enzyme indicated that it represented the enzymatic state prior to substrate binding and that conformational changes would occur upon substrate binding to enclose the putative active site.

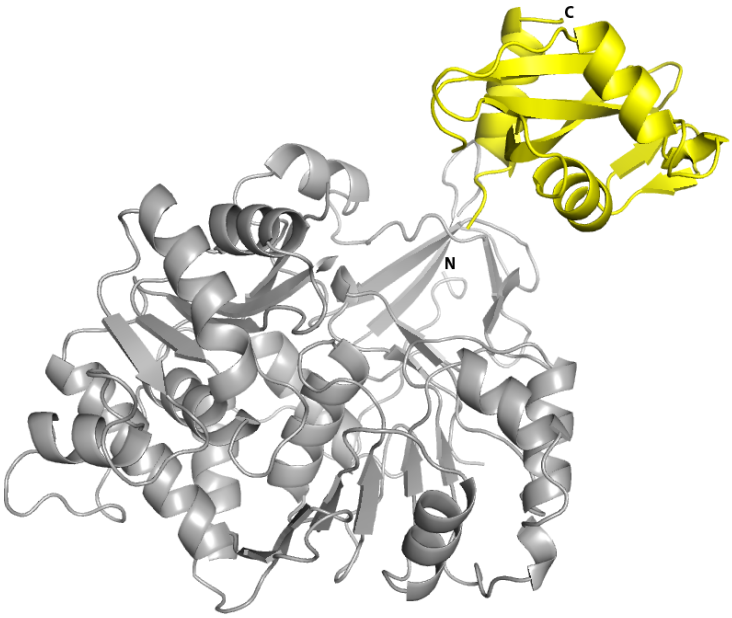

Figure 8. The crystal structure of *P. pyralis* (firefly) luciferase (PDB code: 1LCI, Conti *et al*., 1996). The structure of firefly luciferase was solved in absence of bound substrate and was the first representative for the superfamily of adenylating enzymes. The basic architecture of adenylating enzymes comprises a larger N-terminal domain (in gray) linked to a smaller C-terminal domain (in yellow) with a flexible loop.

The firefly luciferase structure provided information about the general structural organization of the adenylating enzymes. However, in the absence of substrate and a structural comparison from the same superfamily of enzymes, the functional mechanism of these enzymes remained elusive. This was until the structure of the phenylalanine adenylation subunit of Gramicidin synthetase A (PheA) became available (Conti *et al*., 1997). The PheA structure was the first solved A domain structure and more intriguingly represented the enzyme in its substrate-bound state, which enabled characterization of the active site (discussed in detail in the previous chapter). Despite the low sequence identity with firefly luciferase (16%), the tertiary fold of the two proteins was highly similar. As with the luciferase structure, the PheA also consisted of two domains, a larger N-terminal and smaller C-terminal one with the active site in the cleft between the domains. It was noticed however that the C-terminal domain of the PheA had rotated about 140° in order to form the binding site for the substrates. This observation led later to a hypothesis that the different conformational states would represent the steps in the reaction mechanism.

Currently several crystal structures for the members of the adenylating enzyme superfamily have been obtained both in the absence and presence of substrates. These include the crystal structures for the stand-alone 2,3-dihydroxybenzoate (DHB)-activating adenylation domain DhbE from *Bacillus subtilis* (May *et al*., 2002), for the acetyl-CoA synthase (Acs) from *Salmonella enterica* (Gulick *et al*., 2003), and for the D-Alanine carrier protein ligase DltA from both *Bacillus subtilis* (Yonus *et al*., 2008) and from *Bacillus cereus* (Du *et al*., 2008). These enzymes exhibit very similar domain organization with the firefly luciferase and PheA. However, depending on the absence or presence of ligand/substrate or on the reaction state (i. e. if the adenylation reaction has been completed or not), the orientation of the C-terminal domain changes in relation to

the N-terminal one. At least three different conformational states have been observed: an open conformation with the unliganded firefly luciferase structure as an example, an intermediate conformation of which PheA is an example, and finally a so-called closed conformation, represented by *S. enterica* Acs in complex with adenosine-5'-pyrophosphate and CoA (Gulick, 2009). The two conformational states mentioned last have also been termed as the adenylation and thioester-forming conformation, respectively. This nomenclature arises from the reaction that these conformations are proposed to presede. It has been postulated that the three different conformations would mediate the steps in the adenylation reaction mechanism. The open conformation would represent a state prior to substrate binding and the binding of the substrate would cause a shift to the adenylation conformation, which would be the reactive state of the enzyme, and would finally be followed by the closing of the active site to give rise to the thioester-forming conformation (Figure 9) (Yonus *et al*. 2008; Gulick, 2009). This theory is, however, based on structural information from separate enzymes, and does not for example correlate with the information gained from the DhbE structures. The adenylation domain DhbE, which is a stand-alone enzyme responsible for the activation of the aryl acid DHB as the initial step in the synthesis of bacillibactin, was crystallized in unliganded form, together with DHB and AMP, and finally in complex with adenylated DHB (May *et al*., 2002). No conformational change between the three structures was observed, in contrast with the proposed model. Therefore the authors proposed that the binding and activation of the substrate by adenylation would rather involve local changes in the active site.

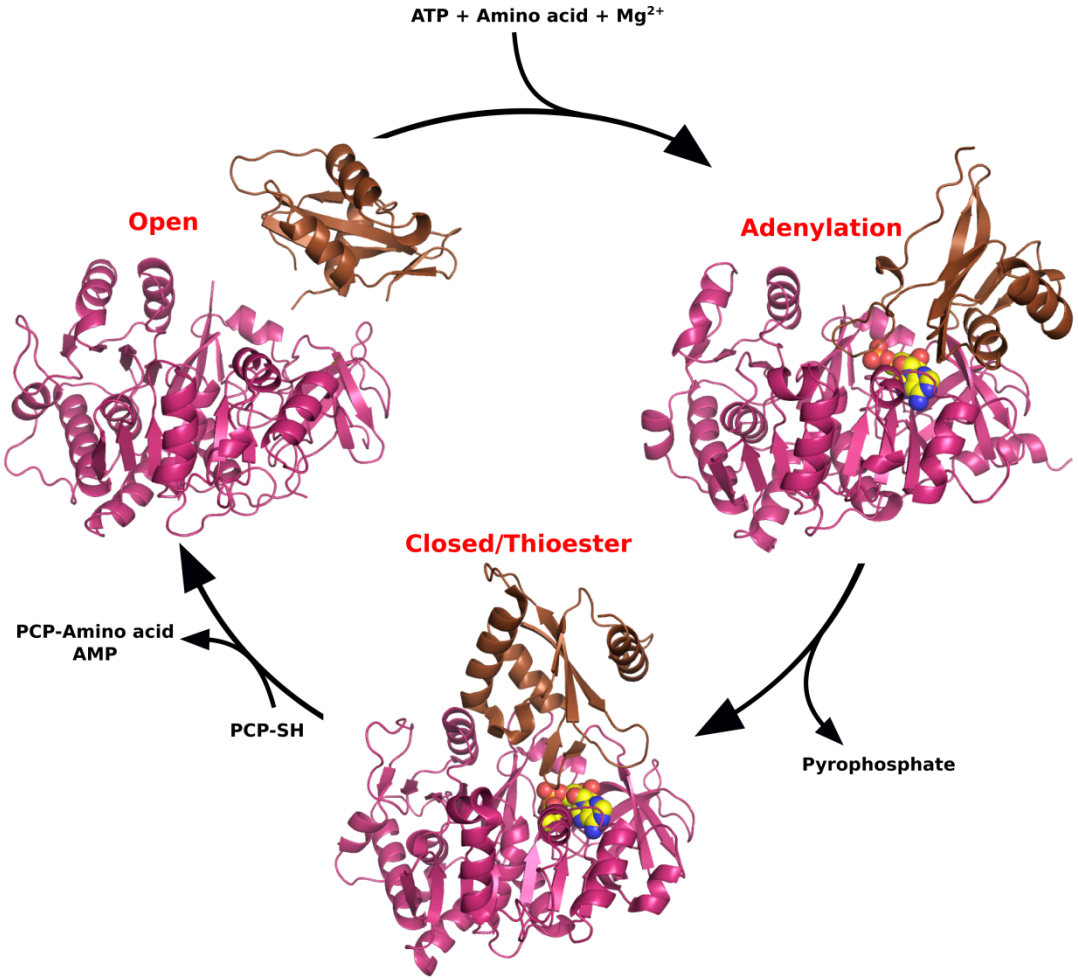

Figure 9. The proposed adenylation domain reaction cycle. The conformational changes in the C-terminal domain (shown in brown) of adenylating enzyme have been proposed to reflect the different steps in the adenylation reaction. The open conformation is reperesented by the structure of unliganded firefly luciferase (PDB code: 1LCI), the adenylation conformation by the structure of PheA and the closed conformation (also termed as thioester conformation) by the structure of DltA (PDB code: 3E7W). The N-terminal domain is depicted in pink and the bound substrate as yellow sphere model. Image adopted from Yonus *et al*. (2008) and Strieker *et al*. (2010).

1.5 Aim of the work

The main aim of the thesis work is the structural characterization of two *Arabidopsis thaliana* proteins, ETR1 and EDR1, both of which are involved in the plant

defense response signaling in ethylene-dependent manner. The most ambitious goal is the expression and purification, and ultimately crystallization of the full-length ETR1, an initiator and regulator of the ethylene-signaling pathway. The structure of the full-length receptor would help to answer questions about the signal relay from the membrane-embedded ligand-binding domain to the cytosolic domains. Since membrane protein expression and purification possess several challenges and a construct without the membrane domain is studied in parallel. A combination of X-ray crystallography and Small Angle X-ray scattering (SAXS) are used for the structural characterization of this ETR1-ΔTM construct. ETR1-ΔTM has already been studied in the group using SAXS but the presence of non-detergent sulfobetaine (NDSB) compound in the buffer was found not to be ideal for these measurements. Since the NDSB is required for the stabilization of the protein in solution, the aim is to improve the purification conditions of the protein in order to make the protein sample more compatible with SAXS. A stable protein is also expected to be more amenable for crystallization.

In addition to the work on the ethylene receptor, structural studies on another *A. thaliana* protein EDR1 are conducted. EDR1 is known to be involved in plant immune responses and also in the regulation of ethylene-induced senescence. It also shares high similarity with CTR1, which makes it interesting target for structural studies in the light of previous work in the group. The structural knowledge of proteins that take part in the plant defense response signaling pathways is important for elucidating the mechanisms by which they regulate their activity or the activity of other proteins in the same or related pathways. Additionally, it could give information needed to genetically modify and thus create plant cultivars more resistant to certain pathogens.

Concerning the proteins that are involved in responses to environmental and biotic stress, another interesting but bacterial enzyme is studied as part of the thesis work. The

target enzyme, ApnA A_1 is the first adenylation domain in the initiation module of the Anabaenopeptin synthetase (a non-ribosomal peptide synthetase) from cyanobacterium *Planktothrix agardhii* strain PCC7821. ApnA A_1 PCC7821 is a bi-specific enzyme that activates both L-tyrosine and L-arginine. As a part of the Anabaenopeptin synthetase, the bi-specificity of ApnA A_1 enables the formation of two different Anabaenopeptins as final peptide products. The possible pharmaceutical applications for the synthesis products of these enzymes make them an interesting topic for structural studies. Knowledge about the residues that define the substrate specificity of the enzyme could help to modify its activity. Enzyme with engineered specificity could be further used in synthesis of new compounds with antimicrobial activity.

2. MATERIALS AND METHODS

2.1 List of online tools

The table 1 summarizes the online tools and their use in this thesis work. The *NetPhos 2.0 Server* was used only for EDR1 kinase domain construct.

Table 1. Online tools used in the thesis work.

Name	Function	Reference
NetPhos 2.0 Server	Prediction of phosphorylation sites	Blom *et al.*, 1999
ProtParam	Calculation of variety of physicochemical properties (molecular mass, isoelectric point and extinction coefficient) for proteins	Gasteiger *et al.*, 2005
PDBePISA	Interfaces, surfaces and assemblies from 3D structures	Krissinel & Henrick, 2007
DALI server	Search for similar 3D structures	Holm & Rosenstrom, 2010
MATTPROB	Calculation of Matthews coefficient	Matthews, 1968; Kantardjieff & Rubb, 2003

2.2 Constructs

The work on plant defense signaling molecules focused on the full-length ETR1 (NCBI accession no. NP_176808) and on the structural characterization of its cytosolic part, referred as ETR1-ΔTM. The latter one comprised GAF domain, histidine kinase domain with dimerization histidine phosphotransfer (DHp) and catalytic ATP-binding (CA) subdomains, and receiver (RR) domain. In addition to ETR1, structural characterization of another plant defense signaling protein, the Enhanced Disease

Resistance 1 (EDR1, NCBI accession no. NP_563824) was conducted. The main construct comprised the catalytic domain of *A. thaliana* serine/threonine kinase EDR1 (EDR1-kd, residues 655-933). In addition to the wild type catalytic domain, three kinase domain variants were studied. These included a catalytically inactive and two G-helix mutants.

Apart from the work on plant signaling molecules, structural studies were conducted with ApnA A_1 PCC7821, which is the first adenylation domain in the initiation module of Anabaenopeptin synthetase from *Planktothrix agardhii* strain PCC7821. In addition to the structural studies on the wild type enzyme, several active site mutants were produced in order to alter the substrate binding specificity of the enzyme. The plasmid containing the wild type ApnA A_1 PCC7821 was obtained from a collaboration partner G. Christiansen (University of Innsbruck).

The main constructs for ETR1, EDR1 and ApnA A_1 PCC7821 are listed in the table 2 and the mutants of these proteins, if applicable, are shown in the table 3. The primers needed for the PCR amplifications of the target genes are listed in appendix 1. The nucleotide sequence for the ApnA A_1 PCC7821 construct is shown in appendix 2.

Table 2. The protein constructs of ETR1, EDR1 and *ApnA* A_1 that were included into the thesis work.

Construct ID	**Protein**	**Domain(s)**	**Molecular weight (kDa)**
FL ETR1	ETR1	Full-length protein	82.6
ETR1-ΔTM	ETR1	GAF++DHp+CA+RR	64.6
EDR1-kd	EDR1	Kinase domain	32.0
ApnA A_1 PCC7821	ApnA A_1 PCC7821	Adenylation domain	61.2

Table 3. The list of EDR1-kd and ApnA A_1 PCC7821 mutants.

Construct	Mutated residue(s)	Mutated into
EDR1-kd	D792	N
EDR1-kd	V876	E
EDR1-kd	F881	E
EDR1-kd	V876/F881*	E/E
ApnA A_1 PCC7821	E204	G
ApnA A_1 PCC7821	E204/S243	G/E
ApnA A_1 PCC7821	S243	C, D, E, G, H, I, K, L, M, N, P, Q, R, T, V, W, Y
ApnA A_1 PCC7821	A307	C, D, E, F, G, H, I, K, L, M, N, P, R, T, V, W, Y
ApnA A_1 PCC7821	K415	A
ApnA A_1 PCC7821	K500	A
ApnA A_1 PCC7821	K415/K500*	A/A

* A double mutant.

2.3 Polymerase Chain Reaction (PCR)

Polymerase chain reaction was used to amplify target genes from suitable DNA source prior to cloning but also to verify the existence of correct clone in bacterial colony (colony-PCR). For this purpose, primers that bind to flanking regions of the target gene were needed. In site-directed mutagenesis PCR was used to introduce a desired mutation into the target gene and consequently to the corresponding protein. Both standard PCR amplifications of target gene and site-directed mutagenesis PCR experiments were conducted using a high-fidelity version of Phusion DNA polymerase (New England Biolabs GmbH, Frankfurt am Main, Germany). In the case of colony-PCR a polymerase with a very low error rate is not necessary and therefore OneTaq polymerase (New England Biolabs GmbH, Frankfurt am Main, Germany) was preferred. The reaction set up for Phusion HF polymerase is shown in table 4 and for OneTaq polymerase in table 5.

Table 4. PCR reaction set up for Phusion HF polymerase.

Reagent	**Amount**
5XPhusion HF or GC reaction buffer	10 µl
2 mM deoxy-nucleotide solution	5 µl
Forward primer (10 pmol/µl)	1.5 µl
Reverse primer (10 pmol/µl)	1.5 µl
DNA template (50 – 150 ng/µl)	1 µl
Phusion HF Polymerase (2 U/µl)	0.7 µl
Milli-Q H_2O	To total volume of 50 µl

Table 5. PCR reaction mixture for OneTaq polymerase as used in the colony-PCR. A sample of the bacterial colony with the target plasmid was added to the solution to provide the DNA template.

Reagent	**Amount**
5XOneTaq standard reaction buffer	5 µl
2 mM deoxy-nucleotide solution	2.5 µl
Forward primer (10 pmol/µl)	1.0 µl
Reverse primer (10 pmol/µl)	1.0 µl
OneTaq DNA polymerase	0.1 µl
Milli-Q H_2O	To total volume of 25 µl

For the colony-PCR the primers were chosen so that one was specific for the vector and the other for the gene insert. Since the *E. coli* expression vectors used in this thesis work contained T7 promoter site, a commercially available T7 forward primer (Eurofins MWG Operon, Ebersberg, Germany) complementary to the nucleotide sequence of the promoter was used in combination with insert specific reverse primer.

The thermocycling conditions for both polymerases were chosen according to manufacturer's protocol unless mentioned otherwise. In a standard PCR run the annealing temperature was set up to be 5°C below the melting temperature (T_m) of the primer with the highest T_m-value. The amplified insert was purified after PCR by either using gel extraction or QIAquick PCR Purification kit (Qiagen GmbH, Düsseldorf, Germany) according to manufactors protocol. For gel extraction the insert DNA was first run on agarose gel, the band corresponding to the insert excised from the gel and finally purified using QIAquick Gel Extraction kit (Qiagen GmbH, Düsseldorf, Germany).

2.4 Site-directed mutagenesis

The EDR1 kinase domain mutants as well as the ApnA A_1 PCC7821 mutants were produced using site-directed mutagenesis. A modified protocol from QuickChange® Site-Directed Mutagenesis Kit (Agilent Technologies, Waldbronn, Germany) was used to introduce the desired mutations. This involved the replication of both strands of the plasmid using oligonucleotide primers with the desired mutation included. The primers were designed so that they were complementary to opposite strands of the plasmid. The primers are extended in a PCR reaction to create a plasmid that contains the mutated target gene. The plasmid from the site-directed mutagenesis PCR contains a nick, which is repaired in *E. coli* after the plasmid is transformed into a suitable *E. coli* strain. Prior to transformation the product from the site-directed mutagenesis PCR is treated with DpnI to digest the parental plasmid. This digestion is based on the ability of the DpnI enzyme to degrade hemimethylated DNA, which is the case with plasmid DNA that has been isolated from *E. coli*.

In the site-directed mutagenesis of EDR1 kinase domain mutants Phusion High-Fidelity DNA polymerase were used in the PCR reaction. The wild type kinase domain of

EDR1 in pMCSG7 vector (Eschenfeldt *et al*., 2009) was used as the template. The thermocycling conditions differed from the standard Phusion HF DNA polymerase protocol in that the number of cycles was reduced to 18. Additionally, instead of the recommended 1kb/min, a doubled extension time of 2kb/min was used to ensure the synthesis of the entire plasmid. For the site-directed mutagenesis of both ApnA A_1 PCC7821 and ETR1-ΔTM Pfu Ultra High Fidelity DNA polymerase (Agilent Technologies, Waldbronn, Germany) was used. The reaction set up for Pfu Ultra HF DNA polymerase is shown in the table 6 and the thermocycling conditions in table 7.

Table 6. PCR reaction set up for ApnA A_1 PCC7821 mutants using Pfu Ultra HF DNA polymerase.

Reagent	**Amount**
10XPfu Ultra Turbo reaction buffer	5 µl
2 mM dNTPs	5 µl
Template DNA	5 µl
10 pmol forward primer	1 µl
10 pmol reverse primer	1 µl
Pfu Ultra HF DNA polymerase	1 µl
Milli-Q H_2O	32 µl

Table 7. PCR thermocycling conditions for the ApnA A_1 PCC7821 mutants.

Step	Temperature	Time	Repeats
Initial Denaturation	95°C	2 min	1
Denaturation	95°C	35 s	18
Annealing	59°C	35 s	18
Extension	72°C	15 min	18
Final Extension	72°C	18 min	1

After the PCR reaction, the formed product was treated with DpnI (New England Biolabs GmbH, Frankfurt am Main, Germany) to degrade the parental plasmid, which does not have the desired mutation. The DpnI digestion was done according to the manufacturer's protocol using CutSmart buffer (New England Biolabs GmbH, Frankfurt am Main, Germany). To ensure the completion of the DpnI digestion the reaction mixture was incubated overnight at 37°C. 5 µl of the overnight incubation product was used to transform high-efficiency and chemically competent *E. coli* DH5α cells according to the standard transformation protocol. From the colonies growing on the plate a subset was chosen to set up pre-cultures for the subsequent plasmid preparation. The correct mutation in the plasmid was verified by DNA sequencing.

2.5 Restriction endonuclease cloning method

The full-length ETR1 was cloned from both cDNA and from pUC57 vector containing the *E. coli* codon-optimized synthetic gene of ETR1 (GenScript USA Inc., Piscataway, NJ, USA) using restriction endonuclease based method. The gene amplification was conducted using standard PCR setup with Phusion DNA polymerase (New England Biolabs GmbH, Frankfurt am Main, Germany) according to

manufacturer's protocol. The ETR1 gene was cloned into pET15b and pETM-11-SUMO3GFP *E. coli* expression vectors using forward primers with NdeI and NcoI and reverse primers with BamHI recognition site, respectively. Both vectors add an N-terminal His_6-tag to the construct and but the latter one also a C-terminal Green Fluorescence Protein (GFP) as a fusion protein. Prior to the ligation step the PCR amplified insert and the target vector were digested with the appropriate enzymes. The inserts as well as the pETM11-SUMO3GFP vector were digested in one step. The pET15B vector was digested in two consequent steps due to the near location of the NdeI and BamHI sites in the vector sequence to ensure the efficient binding of the enzymes to their cutting sites. The basic digestion reaction set up is shown in table 8 using the NcoI/BamHI double digestion as an example. All the enzymes were purchased from New England Biolabs (New England Biolabs GmbH, Frankfurt am Main, Germany) and with the exception of NdeI were the high-fidelity (HF) versions. The pET15b vector was digested for 3 h at 37°C with one enzyme prior to the addition of the second enzyme. After this, the incubation was continued for another 3 h at same temperature.

Table 8. Restriction enzyme digestion reaction for the ETR1 insert for pET15b vector.

Compound	**V(stock solution)**	**Final concentration/amount**
NcoI-HF	1 µl	0.4 U/µl
BamHI-HF	1 µl	0.4 U/µl
10 NEBuffer 4	5 µl	1X
Insert or vector DNA*	x µl	1.0 µg
MilliQ H_2O	38 µl	-

*The amount of vector or insert DNA stock solution dependent on the stock concentration.

After the completion of the reaction both the insert and the vector were purified using Qiaquick Gel Extraction kit (Qiagen GmbH, Düsseldorf, Germany). The purified products were used for the subsequent ligation reaction.

2.6 Ligation in restriction endonuclease cloning

The ligation of the ETR1 gene, both synthetic and cDNA derived, into both pET15b and pETM11-SUMO3GFP vectors was conducted by T4 DNA ligase-dependent method. The T4 DNA ligase is an enzyme that catalyzes the formation of a phosphodiester bond between juxtaposed 5' phosphate and 3' hydroxyl termini in duplex DNA or RNA and can be used to join both blunt end and cohesive end termini.

The ligation reaction was performed with an enzyme from New England Biolabs (New England Biolabs GmbH, Frankfurt am Main, Germany) according to the manufacturer's protocol. The insert and the vector were added to the reaction mixture roughly in 7:1 molar ratio. The reaction set up is shown below (Table 9).

Table 9. The ligation reaction set up in restriction endonuclease cloning.

Reagent	**Amount**
10XT4 ligase buffer	2 µl
Vector DNA*	50 ng
Insert DNA**	110 ng
T4 DNA ligase	1 µl
Milli-Q H_2O	4 µl

*The size of pET15b is 5708 bp and pETM11-SUMO3GFP 6290 bp.

**ETR1 insert size is 2217 bp.

2.7 Ligation Independent Cloning

Ligation Independent cloning or LIC cloning was used to produce two of the main ETR1 constructs. The LIC vectors as well as the cloning protocol were developed at EMBL Hamburg (courtesy of A. Geerlof). The full-length ETR1 was cloned to the pETM-11/LIC, which introduces an N-terminal His_6-tag to the protein. The ETR1-ΔTM had been previously cloned into pET-GB-1a/LIC vector (Mayerhofer, 2011), which adds the β-subunit of protein G as a solubility improving phusion tag to the N-terminus of the target protein.

In LIC cloning the use of restriction enzymes is not required instead it is based on annealing of the complementary overhangs between the insert and the LIC vector. The suitable overhangs are created to the insert by PCR using primers that contain the specific extensions complementary to the LIC-vector. The overhangs needed for insertion of the target gene into EMBL LIC-vectors are as shown here:

Forward primer **CAGGGCGCCA**TG-gene of interest

Reverse primer **GACCCGACGCGGT**TA-gene of interest

Additionally both the vector and the insert are treated with T4 DNA Polymerase to create poly-A (insert) or poly-T (vector) extension to improve the annealing.

The EMBL LIC-vectors contain the gene for enhanced Green Fluorescence Protein (eGFP), which was removed prior to T4 treatment. Two BsaI cutting sites flank the eGFP gene, which allows its removal by an enzymatic digestion reaction. To complete this digestion the reaction mixture shown in the table 10 was incubated at 50°C for an hour. The enzyme and the buffer were obtained from New England Biolabs (New England Biolabs GmbH, Frankfurt am Main, Germany).

Table 10. The reaction components for BsaI digestion of the LIC-vector.

Reagent	**Amount**
10XNew England Biolabs buffer 3	5 µl
LIC vector DNA	5 µg*
BsaI (10 units/µl)	2.5 µl
Milli-Q H_2O	To the total volume of 50 µl

*Usually the amounts of DNA from gel extraction were relatively low and due to this vector DNA was added to fill the reaction volume to 50 µl thus excluding the addition of Milli-Q H_2O.

After the BsaI-digestion the linearized vector was purified using gel extraction. To this end, the sample was run on a 0.8% agarose gel at 100 V for an hour and the band corresponding to the vector was extracted from the gel using the QIAquick Gel Extraction kit according to manufacturer's protocol and subsequently treated with T4 DNA polymerase. The reaction set up for the T4 DNA Polymerase (New England Biolabs GmbH, Frankfurt am Main, Germany) is shown in the table 11.

Table 11. The components required for T4 DNA Polymerase treatment of the LIC-vector.

Reagent	**Amount**
2XNew England Biolabs buffer 2	2 µl
BsaI-digested vector	600 ng
100 mM dTTP	0.5 µl
100 mM DTT	1 µl
100XBSA	0.2 µl
T4 DNA Polymerase (3 units/µl)	0.4 µl
Milli-Q H2O	To total volume of 20 µl

To complete the reaction the reaction mixture was incubated for 30 minutes at room temperature and subsequently the polymerase was inactivated by incubating the reaction mixture at 75°C for 20 minutes.

The insert for the LIC-cloning was amplified using standard PCR reaction and primers with the LIC-cloning extensions and following the protocol for Phusion HF DNA polymerase (New England Biolabs GmbH, Frankfurt am Main, Germany). The PCR amplification was followed by the purification of the insert according to instruction in the used Qiagen PCR Purification kit. The insert was then also treated with T4 DNA polymerase in a similar manner as the vector but using dATP to form complementary overhang with the vector. For the ligation reaction, T4 polymerase treated insert DNA and vector DNA was mixed in 2:1 ratio. The subsequent mixture was first incubated at room temperature for 5 minutes followed by addition of 1 µl of 25 mM EDTA, and the incubation was continued for at least another 5 min but up to 1 hour at room temperature to complete the reaction. The product of the annealing reaction was finally transformed to *E. coli* DH5α cells to seal the nicks and form an intact plasmid. The bacterial colonies growing on the antibiotic containing agar plate were further screened for the correct insertion by colony-PCR. The sequence of the plasmid was verified by DNA sequencing (Eurofins MWG Operon, Ebersberg, Germany).

2.8 DNA gel electrophoresis

Gel electrophoresis is a method for separating macromolecules such as DNA according to their size and charge. The molecules are placed into an electrical field where they migrate towards the positive or negative pole depending on their charge. Due to the phosphate backbone DNA molecule is negatively charged and travels towards the positively charged pole. Therefore the separation of DNA fragments is based on their size

and the velocity of their migration is dependent on the density of the matrix. The most commonly used matrix for DNA applications consists of agarose, a polysaccharide originally extracted from seaweed. Depending on the size of DNA fragments to be separated the concentration of the agarose can vary between 0.7 % and 2.0 %. The sieving range of agarose gel can be from few hundred bases to several kilobases (kb).

DNA electrophoresis was used for both qualitative analysis and purification of DNA molecules from mixtures of differently sized fragments. The agarose concentration used dependent on application of the method. For purification of DNA inserts after PCR and both vectors and inserts after enzymatic digestion was done on a 1.0 % agarose gel. The gel was prepared by mixing 2.5 g of agarose (Serva Electrophoresis GmbH, Heidelberg, Germany) in powder form to 250 ml of 1×Tris-borate-EDTA (TBE) electrophoresis buffer, diluted from a commercially available 10×TBE stock solution (Carl Roth GmbH & Co. KG, Karlsruhe, Germany). The mixture was made homogenous by heating it up in the microwave for up to 3 min at full power (900 W), occasionally gently mixing by swirling. Prior to casting of the gel ethidium bromide (0.025% solution, Carl Roth GmbH & Co. KG, Karlsruhe, Germany) was added for visualization of the DNA fragments by using UV-light. Prior to loading the samples to the gel mixed with 6Xloading dye in sample to loading dye ratio of 5:1. The samples were loaded into the gel together with either a 100 bp DNA ladder (Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany) if the DNA fragment size was smaller than 500 bp or with a 1kb DNA ladder (Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany) if the fragment size exceeded 500 bp. Typically the gel was run at 100 V for 1 hour submerged in 1×TBE buffer.

For analyzing the success of digestion or PCR reaction both after site-directed mutagenesis and colony-PCR, a lower agarose gel concentration of 0.8% in combination

with 1×Tris-acetate-EDTA (TAE) buffer was used. The preparation of the 50×TAE buffer is hown in the table 12. The sample handling and choice of DNA ladder was as previously described in the case of DNA purification from 1% agarose gels. The gel was however run at lower energy of 50 V for 80 min to avoid the melting of the gel.

Table 12. The preparation of 50×TAE buffer. The chemicals were purchased from Roth (Carl Roth GmbH & Co. KG, Karlsruhe, Germany).

Reagent	**Final concentration**
TRIS	2.0 M
Glacial acetic acid	1.0 M
EDTA	50 mM

2.9 Plasmid preparation

The plasmid containing the gene for the target protein was transformed to E. coli DH5α strain according to standard protocol. Colony from the agar-plate was used to inoculate 5 ml of Luria Bertani (LB) medium (see table 13 for recipe) with suitable antibiotics and grown overnight at 37°C. The cells were then harvested the next day by centrifugation at 4000 rpm for 10 min at 4°C. The pellet was used for the plasmid purification, which was conducted using the Qiaprep Spin Miniprep kit (Qiagen GmbH, Düsseldorf, Germany) according to the protocol supplied by the manufacturer. The plasmids were eluted to 32 µl Milli-Q water and their concentration was measured using Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc., Wilmington, USA).

Table 13. The components of LB-medium. The chemicals were purchased from Roth (Carl Roth GmbH & Co. KG, Karlsruhe, Germany).

Compound	**Final concentration**
Tryptone	10 g/l
Yeast Extract	5 g/l
NaCl	5 g/l

2.10 Transformation

The plasmid DNA containing the target gene was introduced into the suitable chemically competent *E. coli* strain by heat-shock transformation method. When transformation was performed after cloning experiments the amount of ligation mixture added to the thawed competent cells was between 4 and 10 µl whereas for standard plasmid transformation 0.5 - 1 µl of plasmid solution was normally sufficient. After this the cells were incubated 30 minutes on ice and then heated for 45 seconds in 42°C water bath in order to the bacterial membrane become porous and take the target plasmid in. The heat-shocked *E. coli* cells were then incubated additional 2 minutes on ice to reduce the damage on the cells followed by the addition of 80 µl SOC (Super optimal broth with catabolite repression) medium (see Table 14 for recipe) and incubation for 1 hour at 37°C. Finally the cells were plated on LB-agar containing the suitable antibiotic and colonies were allowed to grow during the overnight incubation period at 37°C.

Table 14. The components included in the SOC medium.

Component	Final concentration
Yeast extract	0.5 % (w/v)
Tryptone	2.0 % (w/v)
NaCl	10 mM
KCl	2.5 mM
$MgCl_2$	10 mM
$MgSO_4$	10 mM

2.11 Expression in *E. coli*

The expression of ETR1-ΔTM with the deca-His-tag was done as previously described for the His_6-tagged protein (Mayerhofer, 2011). For this purpose the plasmid containing the His_{10}-tagged construct was transformed to the *E. coli* Rosetta 2 (DE3) strain and colony from the kanamycin-containing agar-plate was used to inoculate overnight LB-culture. 1 L of Terrififc Broth (TB, see table 15 for recipe) culture medium was in turn inoculated with 10 ml of the overnight pre-culture and the cells were allowed to grow at 37ºC with 200 rpm shaking (Infors HT Multitron 25, Infors GmbH, Einsbach, Germany) until OD_{600}-value of 1.0 – 1.5 was reached. At this point, the cultures were induced with isopropyl-β-D-thiogalactopyranoside (IPTG), which was added to final concentration of 40 µM. Culturing was then continued at lower temperature of 25ºC for another 18 hours. The cells were harvested by centrifugation at 4500 rpm for 45 minutes at 4ºC and the resulting pellet stored at -20ºC until the day of use.

Table 15. The preparation of the TB-media for *E. coli* cultures. The chemicals were purchased from Roth (Carl Roth GmbH & Co. KG, Karlsruhe, Germany).

Component	**Final concentration**
Casein	12 g/l
Yeast Extract	24 g/l
K_2HPO_4	12.5 g/l
KH_2PO_4	2.3 g/l

The EDR1 kinase domain constructs were transformed and expressed in *E. coli* strain BL21 cells co-expressing chaperones DnaK, DnaJ, GrpE, ClpB, GroEL and GroES (CC4 cell lines, courtesy of A. Geerlof, Helmholtz Zentrum München, Munich, Germany). The pre-cultures were prepared by inoculating 5 ml of LB-medium containing 100 µg/ml ampicillin with a single colony from a plate of freshly transformed cells. The pre-cultures were incubated overnight at 37ºC in shaking incubator (160 rpm, Infors HT Multitron 25) and used the next day to inoculate 1 L of auto-induction medium (Studier, 2005). The components of the auto-induction medium are shown in table 16. The 1 L cultures were first incubated at 37ºC with 200 rpm shaking speed until an OD_{600} between 0.6 and 0.8 was reached. After this point the cultures were cooled down and incubation was continued for another 18 h at a lower temperature of 21ºC. IPTG addition was not needed since the α-lactose in the medium functions as the inducer and starts to act after the bacteria has consumed all the available glucose. Finally, the cells were harvested by centrifugation at 5500 rpm in a Sorvall RC26 centrifuge (Thermo Fisher Scientific Inc./Thermo Electron LED GmbH, Langenselbold, Germany) equipped with JLA-8.1000 rotor (Beckmann Coulter, Inc., Brea, CA, USA) for 25 min at 4ºC.

Table 16. The components of the ZYM-5052 auto-induction medium used for large-scale cultures.

Component	**Concentration**
Trype	1 %
Yeast Extract	0.5 %
Na_2HPO_4	25 mM
KH_2PO_4	25 mM
NH_4Cl	50 mM
Na_2SO_4	5 mM
Glycerol	0.5 %
Glucose	0.05 %
α-lactose	0.2 %
$MgSO_4$	2 mM
1000×Trace elements*	0.2x

*This solution contains 50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM $MnCl_2$, 10 mM $ZnSO_4$ and $CoCl_2$, $CuCl_2$, $NiCl_2$, Na_2MoO_4, Na_2SeO_3 and H_3BO_3 each in concentration 2 mM.

The expression protocol for the ApnA A_1 PCC7821 mutants was obtained from a collaboration partner Prof. Dr. Andrea Rentmeister (University of Münster). For expression the plasmid containing the gene for the ApnA A_1 PCC7821 mutant was first transformed into *E. coli* strain BL21(DE3). The pre-cultures were set up by inoculating 6 ml of LB-medium containing 50 µg/ml kanamycin with a single colony from LB-agar plate. These cultures were then grown overnight in shaking incubator (160 rpm, Infors HT Multitron 2) at 37°C and used next day to inoculate 400 ml of LB-medium containing kanamycin in same concentration than in the pre-cultures. The main cultures were first incubated at 37ºC until suitable cell density was reached, indicated by the OD_{600}-value between 0.2 and 0.5. The expression of the target protein was then induced by the

addition of IPTG to 0.16 mM final concentration. The culturing was then continued at 23ºC for 5 h after which the cells were harvested at 4°C at 6000 rpm (Sorvall RC26 centrifuge with JLA-8.1000 rotor) for 15 min. The cell pellets were stored at -20ºC prior to use.

2.12 Virus preparation for insect cell culturing

The expression in insect cells was conducted using MultiBac expression system developed in Berger lab at EMBL Grenoble (Berger *et al*., 2004; Fitzgerald *et al*., 2006; Bieniossek *et al*., 2008). The target gene was inserted to pFastBac HTa vector by restriction enzyme based cloning method and transformed into *E. coli* DH10EMBacY cells. This cell strain contains the baculoviral genome as bacterial artificial chromosome to which the target gene is integrated by Tn7 transposition. It also contains the helper plasmid for Tn7 transposon enzyme and YFP reporter gene in the virus backbone for quantitative analysis of protein expression. Additionally the DH10EMBacY strain has resistance to kanamycin, tetracycline and chloramphenicol. The transformed cells were plated on an agar plate, which contained in addition to the listed three antibiotics also gentamycin in standard concentrations, 1 mM IPTG and 40 µg/ml 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal, Invitrogen™, Life Technologies GmbH, Darmstadt, Germany). A dilution series of 1:1, 1:10, 1:100 and 1:1000 was made from the transformed cells to optimize the cell density on the agar plates. The plates were then treated according to standard transformation protocol. The positive integration of the target gene was confirmed by blue-white screening since the baculoviral genome also contains a LacZα gene that becomes disrupted upon the insertion of the target gene. The colonies with the correct clone were distinguishable by their white color. Two of these white colonies were picked to inoculate separate 2 ml LB-cultures.

The isolation of the baculoviral DNA with the target gene was conducted according to a modified protocol from Qiagen Spin Miniprep kit using the buffers from the kit. The overnight LB-culture was centrifuged at 4000 rpm for 10 min using a bench-top centrifuge (Eppendorf 5810R, Eppendorf AG, Hamburg, Germany Diagnostics) to harvest the cells. The cells were first re-suspended to 300 µl of P1 buffer followed by addition of 300 µl of P2 buffer and gentle mixing of the suspension by turning the tube 4-6 times upside down. To this 300 µl of N3 buffer was then added and the suspension was mixed gently in a similar manner as in previous step. To remove all *E. coli* debris, the solution was first centrifuged at room temperature for 10 minutes at 13,400 rpm (Eppendorf Minispin, Eppendorf AG, Hamburg, Germany) followed by the transfer of the supernatant to a fresh, sterile tube and an additional centrifugation for 3 minutes. The bacmid was then isolated from the supernatant using isopropanol precipitation. To this end, 700 µl of 100% isopropanol was added and the solution was centrifuged for 10 min at 13,400 rpm at room temperature. The resulting pellet washed with 70% ethanol. After this step the work was continued in sterile conditions to avoid any contamination from *E. coli*. The remaining residual amounts of 70% ethanol were removed from the DNA pellet by drying it in sterile hood. This was followed by addition of filter-sterilized water to dissolve the pellet. The bacmid was prepared for transfection by first adding 200 µl serum free insect cell medium (HyClone SFM4 Insect, Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany) to 20 µl of the re-suspended baculoviral DNA and then mixing the resulting solution with 100 µl of transfection reagent solution. The transfection reagent solution was prior to this prepared by mixing 100 µl of insect cell media with 10 µl of transfection reagent (X-Treme GENE HP, Roche Diagnostics GmbH, Mannheim, Germany).

Prior to transfection a stock culture of *Spodoptera frugiperda* Sf21 cells was grown in the serum free insect cell medium to cell density of 0.5 – 1.0 Million/ml and this stock was used to seed wells of a 6-well tissue culture plate leaving one well empty for medium control. The wells were supplemented with the insect cell medium to total volume of 3 ml. The prepared baculoviral DNA is sufficient to transfect two wells leaving one seeded well un-transfected for cell control. After transfection the cells were grown at 27°C for 60 hours followed by harvesting of the supernatant. The supernatant contains the initial virus or V_0 and was used to infect fresh cells for virus amplification.

2.13 Virus amplification and expression in insect cells

The virus amplification was conducted according to the Multibac protocol (Bieniossek *et al*., 2008). The infected cells were monitored in 24-hour intervals by counting the cells. If the cell count exceeded 1.5 Mio, the culture was divided in two halves and to both equal amounts of culture medium was added. The cells infected with ETR1 gene-containing virus usually doubled only once 24 hours after the infection. After 48 hours the cells stopped doubling and therefore reached the day of proliferation arrest (dpa). To monitor the expression level of the protein 1.0 million cell (Mio) aliquots were withdrawn starting from the dpa and then in 24-hour intervals. 60 hours after dpa the cell suspension was transferred to 50 ml falcon tubes and centrifuged at 150 rcf for 3 minutes to harvest the amplified virus or V_1 virus in the supernatant. The cell pellet was resuspended to media according to the protocol and culturing was continued till 96 hours after dpa. At the 96^{th} hour after infection the cells were harvested by gentle centrifugation at 150 rcf for 3 min and stored in -20°C. In order to evaluate the protein expression, the cell probes that were taken on dpa and on 24-hour intervals after dpa were analyzed on SDS-PAGE gel.

Prior to protein production in large-scale baculovirus infected insect cell (BIIC) stocks were prepared. For this purpose 50 ml of cells in density 1 Mio cells/ml were infected with 6 ml of V_1 and maintained in this density until dpa was reached. Freezing media containing 90% of HyClone insect cell media, BSA in concentration 10 g/l and 10% of DMSO (v/v) was prepared and filter-sterilized. 100 ml of cell culture was centrifuged at 150 rcf for 10 minutes, gently re-suspended to the sterile freezing solution to final density of about 10^7 cells/ml and finally divided to 1 ml aliquots into sterile cryovials. These aliquots were either used directly to infect expression cultures or optionally stored at -20°C.

The protein production was done in 1 L cultures, which were infected using the prepared BIIC stocks. To this end, one BIIC aliquot was thawed rapidly at 37°C water bath, then diluted quickly to 100 ml with the culture media (HyClone) and finally added to 900 ml of uninfected cells at a density of about $0.9*10^6$ cells/ml. The cells were incubated at 27°C and kept at a density of 10^6 cells/ml until the day of proliferation arrest. The protein expression levels were monitored by taking samples every 24 hours as described in the V_1 virus preparation section.

2.14 Membrane protein isolation

The full-length ETR1 was expressed in insect cell strain Sf21 and as a membrane protein was assumed to insert to the cell membranes. Subsequently, the purification of ETR1 required a protocol adapted for membrane proteins. A protocol for membrane protein isolation from *E. coli* was obtained from Dr. Preben Morth (Centre for Molecular Medicine Norway (NCMM), Blindern, Oslo, Norway). It was slightly modified to be better suited for the isolation from insect cells. The initial isolation experiments were carried out using the buffers listed in the table 17. The ultracentrifugations and the

isolation steps in between were performed at Biozentrum Klein Flottbek (Hamburg University, Klein Flottbek, Hamburg) by Ulrike Peters. From each individual isolation step, a small sample was taken for analysis on SDS-PAGE gel and for Western blot.

Table 17. The buffers that were used in the initial membrane isolation experiment for ETR1. The β-mercaptoethanol, protease inhibitor PMSF and DNase were always added fresh to the stock solutions that contained rest of the components.

Lysis buffer	**Membrane wash buffer**	**Membrane resuspension buffer**
50 Mm Tris-HCl pH 7.6	20 mM Tris-HCl pH 7.6	20 Mm Tris-HCl pH 7.6
100 mM NaCl	100 mM NaCl	200 mM NaCl
5% (V/V) Glycerol	5% (V/V) Glycerol	10% (V/V) Glycerol
5 mM β-mercaptoethanol	1 mM PMSF	5 mM β-mercaptoethanol
3 mM PMSF		1 mM $MgCl_2$
1 µg/ml DNase		1 mM PMSF

Prior to cell lysis the insect cells were thawn on ice and washed twice with cold lysis buffer to carefully remove residuals of the insect cell media. For the wash step the pellet wash gently resuspended to 5 ml of the lysis buffer and centrifuged at 1250 rpm for 5 min at 4°C. The washed pellet (m = 1 g) was then resuspended to 8 ml of cold lysis buffer and sonicated for 30 s on ice using 0.5s/0.5s ON/OFF pulsation mode and 45% amplitude. This procedure was repeated once with one minute break in the between. The sonicated suspension was centrifuged for 20 minutes at 18000 rpm at 4°C to remove cell debris, inclusion bodies and DNA. The membrane fraction with the protein should remain in the supernatant, which was transferred on ice to Biozentrum Klein Flottbek for the subsequent isolation steps. To isolate the membranes with the protein attached, the supernatant from the previous centrifugation step was subjected to ultracentrifugation

(45Ti rotor, Beckmann Coulter, Inc., Brea, CA, U.S.A). The ultracentrifugation was performed at 4°C using 45 000 rpm rotational speed for 75 minutes. This resulted in small, soft membrane pellet that was washed by resuspending the pellet to 9 ml of ice-cold membrane wash buffer with Dounce homogenizer and repeating the ultracentrifugation. The wash step was followed by the membrane protein solubilization step. To this end, the washed pellet was resuspended to 3 ml of ice-cold membrane resuspension buffer and a solubilization detergent DDM was added to final concentration of 1% (W/V). For efficient solubilization, the solution was incubated overnight at 4°C with gentle agitation. The isolation process was completed the next day by ultracentrifugation. The success of the isolation and solubilization was evaluated using SDS-PAGE and Western blot.

The effect of pH and salt concentration was also studied to obtain soluble protein. To this end, two different pH-values (6 and 8) in combination with two different NaCl concentrations (20 mM and 500 mM) were tested. The buffers for these experiments were prepared as described in tables 18 and 19. Apart from the modification of the buffers, the isolation was conducted as described above.

Table 18. The low pH buffers for ETR1 isolation experiments.

Lysis buffer	**Membrane wash buffer**	**Membrane resuspension buffer**
50 Mm MES pH 6	20 mM MES pH 6	20 Mm MES pH 6
20 or 500 mM NaCl	20 or 500 mM NaCl	20 or 500 mM NaCl
5% (V/V) Glycerol	5% (V/V) Glycerol	10% (V/V) Glycerol
5 mM β-mercaptoethanol	1 mM PMSF	5 mM β-mercaptoethanol
3 mM PMSF		1 mM $MgCl_2$
1 µg/ml DNase		1 mM PMSF

Table 19. The high pH buffers for ETR1 isolation experiments.

Lysis buffer	**Membrane wash buffer**	**Membrane resuspension buffer**
50 Mm Tris-HCl pH 8	20 mM Tris-HCl pH 8	20 Mm Tris-HCl pH 8
20 or 500 mM NaCl	20 or 500 mM NaCl	20 or 500 mM NaCl
5% (V/V) Glycerol	5% (V/V) Glycerol	10% (V/V) Glycerol
5 mM β-mercaptoethanol	1 mM PMSF	5 mM β-mercaptoethanol
3 mM PMSF		1 mM $MgCl_2$
1 µg/ml DNase		1 mM PMSF

2.15 Protein isolation from *E. coli* cells

The isolation of the soluble protein from the E. coli cells was conducted in similar manner for the ETR1-ΔTM and EDR1 constructs. The pellets from the large-scale *E. coli* cultures were first thawed on room temperature water path then re-suspended to lysis buffer (see table 20) to final volume of 50 ml. The re-suspension of the pellet was conducted by 250 rpm mixing with magnetic stirrer at 4°C. To break the cells the cell suspension was subjected to sonication, which is a method that employs sound energy at ultrasonic frequencies to disrupt the cell membrane. Typically the cell sample was sonicated on ice with ultrasonic homogenizer (Sonopuls HD 3200, Bandelin electronic GmbH&Co. KG, Berlin, Germany) for 4 minutes in 0.4 s/0.9s ON/OFF pulsation mode and this treatment was repeated once. The suspension was then centrifuged at 19 000 rpm for 45 - 60 min at 4°C to remove cell debris. For this purpose SS-34 fixed angle rotor (Thermo Fisher Scientific, Inc.) in combination with Sorvall RC26 centrifuge (Thermo Fisher Scientific, Inc.) was used. The clear lysate was transferred to a fresh 50 ml falcon and filtered using 45 µm membrane prior to Ni-NTA affinity chromatography purification step.

The cell lysis and purification protocol for ApnA A_1 PCC7821 was obtained from Prof. Andrea Rentmeister (University of Münster). As with EDR1, the cell pellet was first thawed on RT water bath after which 10 ml of the lysis buffer described in table 20 was added to resuspend the pellet. The resuspended pellet was then sonicated three times for 3 minutes at 30% amplitude level and having 2 minute break between the sonication rounds. The sonication was followed by centrifugation at 4000 rpm for 20 min at 4°C to remove the cell debris. The lysate was filtered using 0.22 µm membrane for the subsequent NiNTA purification step.

Table 20. The cell lysis buffers for the EDR1 kinase domain and the ApnA A_1 PCC7821 constructs.

Construct	**Lysis buffer composition**
EDR1 kinase domain (WT and mutants)	50 mM Hepes pH 7.5, 400 mM NaCl, 5% glycerol, 5 mM $MgSO_4$, 0.1% CHAPS, 1xEDTA-free protease inhibitor tablet, 10 µg/ml DNase
ApnA A_1 PCC7821	50 mM Tris pH 8.0, 300 mM NaCl, 5 mM Imidazole

2.16 NiNTA purification

Ni-affinity chromatography is a well-known and most widely used version of immobilized metal affinity chromatography (IMAC) and was performed as a first purification step for all the constructs in this thesis work. The isolation of the target protein from the crude cell extract in this purification step was based on the presence of an N-terminal hexa-histidine tag, which was included in the N-terminus of every construct. For the EDR1 kinase domain Ni-NTA agarose beads (2 ml for WT kinase domain and for the D792N mutant, and 1 ml for the G-helix mutants) in gravity flow

column were used for this purpose whereas the corresponding purification step for ETR1-ΔTM and for ApnA A_1 PCC7821 was conducted in commercially available NiNTA superflow cartridge columns. For ETR1-ΔTM 5 ml NiNTA superflow cartridge column (Qiagen GmbH, Düsseldorf, Germany) and for ApnA A_1 PCC7821 1 ml His60 Ni-Superflow cartridge (Invitrogen™, Life Technologies GmbH, Darmstadt, Germany) was used. The basic steps of the Ni-affinity purification were the same for the different proteins consisting of repeated washes and eluting the protein with an imidazole gradient. The main differences arose from the buffer compositions and amounts of wash and/or elution steps, and also in the amounts of buffers used in these steps. The compositions of the two basic buffers used in the wash and elution steps are listed in the table 21. For EDR1 these buffers are later referred as buffer A and B, respectively.

Table 21. Compositions of the basic wash and elution buffers used in the NiNTA purification.

Protein	**Washing buffer***	**Elution buffer****
EDR1	50 mM Hepes pH 7.0, 400 mM NaCl, 5 % Glycerol, 5 mM $MgSO_4$	50 mM Hepes pH 7.0, 400 mM NaCl, 5 % Glycerol, 5 mM $MgSO_4$, 500 mM Imidazole
ETR1-ΔTM	20mM Tris 8.8, 150mM NaCl, 3mM β-Mercaptoethanol, 0.1% (w/v) CHAPS, 5% glycerol	20mM Tris 8.8, 150mM NaCl, 3mM β-Mercaptoethanol, 5% glycerol, 300 mM Imidazole
ApnA A_1 PCC7821	50 mM Tris pH 8.0, 300 mM NaCl, 5 mM Imidazole	50 mM Tris pH 8.0, 300 mM NaCl, 250 mM Imidazole

*Referred as buffer A for EDR1. **Buffer B for EDR1.

Prior to loading the filtered cell lysate to the column or onto the beads, the Ni-NTA matrix was equilibrated with 10 column volumes (CV) of wash buffer or buffer A in

the case of EDR1. After the lysate was loaded the column, the wash steps followed. For EDR1, the beads were first washed with 10 CV of buffer A with 6% buffer B followed by 5 CV of buffer A with 10 % of buffer B. The protein was eluted in four consecutive steps, first with 15 %, 20% and 60% of buffer B in buffer A and finally with buffer B (75 mM to 500 mM imidazole). 2CV of buffer was used in each elution step.

For the ETR1-ΔTM the column was washed after lysate loading by increasing linearly the elution buffer concentration to 17% in a 5 CV range. This was followed by elution gradient from 17 % to 100 % of elution buffer in 9 CV range. To ensure that all of the protein eluted out, elution buffer was run through the column until a plateau in absorbance at 280 nm was reached.

The lysate loading for ApnA A_1 PCC7821 was followed by a two-step column wash, first with 7 CV of washing buffer then with 7CV of 10 % elution buffer in washing buffer. The protein was eluted in a 20 CV linear gradient where the elution buffer concentration from 10 to 100 %. Finally 8 CV of elution buffer was run through the column to elute any remaining protein.

2.17 TEV-cleavage and 2nd NiNTA

The EDR1 constructs were all with an N-terminal His_6-tag whereas the ETR1-ΔTM contained an N-terminal His_{10}-tag and also an N-terminal GB1-tag. The expression vectors used for these constructs contained a TEV-protease cleavage site between the target protein and the tag enabling an easy removal of the tag. After NiNTA purification step the His_6 – or His_{10}-tag was not needed and was therefore removed by TEV-protease treatment. This was conducted during an overnight dialysis step at 4°C. The components of the dialysis buffers for both ETR1-ΔTM and EDR1 kinase domain constructs are listed in the table 22.

Table 22. Dialysis buffers for the EDR1 and ETR1-ΔTM constructs.

Dialysis buffer for EDR1 constructs	**ETR1-ΔTM dialysis buffer**
50 mM Hepes pH 7.0	30 mM Tris pH 8.8
400 mM NaCl	100 mM NaCl
5% Glycerol	3 % Glycerol
5 mM $MgSO_4$	6 mM β-mercaptoethanol
1 mM EGTA	
1 mM DTT	

For the TEV treatement the suitable elution fractions from Ni-affinity purification step were combined and to this TEV-protease in concentration 2 mg/ml were added in 1:10 molar ratio. The resulting protein solution was dialyzed overnight in a dialysis bag (Spectra/Por® 1 Dialysis Membranes, MWCO 6000 to 8000, Spectrum® Laboratories) against 2 liters (ETR1-ΔTM) or 5 liters (EDR1) of dialysis buffer.

The overnight tag-removal step was followed by a second Ni-affinity purification to remove the His-tagged TEV-protease and uncleaved protein material. To this end, the dialyzed sample was loaded onto Ni-beads in gravity flow column and flow through was collected. 1 CV of dialysis buffer was added to the beads to elute the remaining tag-free protein out of the column. After this the material bound to the beads was eluted with 2 CV of elution buffer and collected to a separate falcon tube for analysis of the protease treatment efficiency by SDS-PAGE.

2.18 Size-Exclusion Chromatography

The purification of EDR1 kinase domain constructs was finalized with size-exclusion chromatography (SEC), which is a method that separates proteins according to their molecular weight. The SEC purification was conducted in Hiload 16/60 Superdex G75 column (GE Healthcare Bio-Sciences AB, Uppsala, Sweden). The matrix of this column consists of cross-linked agarose and carbohydrate dextran, and it is suitable for separating proteins in molecular range from 3000 Da to 70 000 Da. The tag-free monomeric EDR1 kinase domain has a molecular weight of 32 kDa and fits into the separating range of this SEC column.

Prior to the SEC run the tag-free protein material was concentrated to volume of about 500 µl using Corning Spin-X UF centrifugal concentrator with a 10 kDa cut-off (Corning Incorporated Life Sciences, Tewksbury, USA) and filtered through a 0.22 µm membrane. The sample was then applied to the column equilibrated with SEC buffer (50 mM HEPES pH 7.5, 400 mM NaCl, 5 % Glycerol and 5 mM $MgSO_4$). Based on the measured absorbance at 280 nm (NanoDrop ND-1000 spectrophotometer, NanoDrop Technologies, Wilmington, DE), protein-containing fractions were analyzed on SDS-PAGE and the fractions with monomeric protein were pooled together. The amount of the obtained protein was determined by measuring the absorbance at 280 nm with NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). An extinction coefficient of 1.38 was used to calculate the correct sample concentration from the measured absorbance reading.

2.19 SDS-PAGE and Western blot

The purity of the protein samples was analyzed using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), which is a very widely used and popular discontinuous buffer system and is also known as the Laemmli method. Bio-Rad Mini-PROTEAN® Tetra Cell was used as the electrophoresis system together with self-made Tris-minigels in 1 mm thickness. Since discontinuous buffer system was utilized, the gel composed of two separate layers of stacking and resolving gel. The recipes for the resolving gels in the three most commonly used densities are listed in the table 23 and the preparation of the stacking gel in table 24.

Table 23. The components of the resolving gel for the self-made Tris-buffer based SDS-PAGE gel. The amounts are optimized for casting two gels in 1 mM thickness for Mini-Protean system.

Reagent	**Amount**		
	7.5 %	**10 %**	**12 %**
4×Tris pH 8.8	2.5 ml	2.5 ml	2.5 ml
Milli-Q H_2O			
10 % SDS	100 µl	100 µl	100 µl
40 % Acrylamide (Roth)	2.0 ml	2.7 ml	3.2 ml
20 % APS	50 µl	50 µl	50 µl
Temed (Roth)	10 µl	10 µl	10 µl

Table 24. The components of the stacking gel for the self-made Tris-buffer based SDS-PAGE gel. The amounts are optimized for casting two gels in 1 mM thickness for Mini-Protean system.

Reagent	**Amount**
4×Tris pH 6.8	1.25 ml
Milli-Q H_2O	3.14 ml
10 % SDS	50 µl
40 % Acrylamide (Roth)	0.53 ml
20 % APS	25 µl
Temed (Roth)	5 µl

Prior to loading the samples onto the gel, 7.5 µl of protein sample was mixed with 2.5 µl of 4×NuPAGE® LDS sample buffer supplied with β-mercaptoethanol, the sample was then boiled for 5 minutes at 75°C and finally centrifuged shortly. The whole 10 µl sample was loaded onto the well of the gel. The EDR1 constructs were always run on 12% gels whereas the ETR1 constructs and cell pellet samples from expression tests were run on gels with density between 7.5 and 10% for more optimal separation. All SDS-PAGE gels were run at constant voltage of 200 V for 48 – 60 minutes depending on the length of the construct. For longer constructs longer run times were usually chosen to better separation at higher molecular weight range.

For the immuno-detection the samples were first loaded onto SDS-PAGE gel and then transferred onto nitrocellulose membrane (Protran - Schleicher & Schuell) using Bio-Rad Mini Trans-Blot cell (Bio-Rad Mini-PROTEAN® Tetra Cell). During the transfer the blotting cell was cooled down on ice and the transfer was performed using constant voltage of 100 V for 1 hour. The membrane was subsequently incubated in PBS-T buffer containing 5% (w/v) milk powder to block unspecific binding of the primary antibody. Prior to this step the membrane had been rinsed with PBS-T buffer for 15 min.

For the blotting mouse antibodies against His6-tag from EMBL Heidelberg were used as primary marker (1:2000 dilution) and Goat anti-mouse-HRP from Pierce as secondary antibody (1:5000 dilution). Detection of ETR1 was carried out by luminescence reaction with Super Signal West Pico Chemiluminescent Substrate (Pierce) as the reagent. The reaction was followed using Bio-Rad GelDoc system.

2.20 SEC-LS/UV/RI

Several methods exist to determine the oligomerization state of the protein sample. The oligomerization state of wild type EDR1 kinase domain and the D792N mutant were evaluated using the Malvern system (Malvern Instruments Ltd, Malvern, UK), which contained light scattering, refractive index and optical UV detectors. For a non-glycosylated protein (as with *E. coli* expressed protein) the molecular mass of the eluted sample can be calculated from the measured light scattering intensity, (LS) and refractive inde signal, (RI) using the following formula:

$$M = K'(LS)/(RI),$$

where $K' = K_{RI}/[K_{LS}(dn/dc)]$ (Wen *et al.*, 1996). K_{RI} and K_{LS} are the instrument calibration constants for the refractive index and light scattering, respectively, and are determined using protein standard. The refractive index increment, dn/dc is for protein solution 0.185.

For the analysis the protein samples were concentrated to 1 – 4 mg/ml concentration using Spin Concentrators with 10 kDa cut-off (Corning® Spin-X® UF 20mL Centrifugal Concentrator, 10,000 MWCO Membrane, Corning Incorporated) and filtered with spin filters (0.22 µm membrane, Millipore). A 100 µl sample was injected to the system equipped with an analytical grade size-exclusion column (Superdex 75 10/300 GL, GE Healthcare Bio-Sciences AB, Uppsala, Sweden).

2.21 Thermofluor

The stability and monodispersity of the protein in soluble form is critical for many applications used to study these macromolecules. Especially in structural studies with x-ray crystallography, obtaining a crystalline sample remains still as one of the main bottlenecks and is usually hampered by the availability of pure and stable sample. Thermofluor is a fluorescence-based thermal-shift assay that can be used to examinate the stability of the protein in various different buffer conditions and in presence of additives in a high-throughput manner. The method is based on measuring the temperature-dependent unfolding of the protein through the binding of the hydrophobic fluorophore to the exposed hydrophobic areas of the unfolding protein.

The thermofluor assays were performed at EMBL Hamburg SPC facility using 96-well thin-wall PCR plates and real-time PCR machine (MyiQ™ Single-Color Real-Time PCR Detection system, Bio-Rad) equipped with a charge-coupled device (CCD) detector for the fluorescence imaging (Boivin *et al*., 2013). For the thermofluor analysis the studied protein was diluted to a concentration between 20 - 100 µM. SYPRO Orange (Invitrogen) was used as a fluorescence dye. This dye is stored as 5000×concentrated solution in DMSO and was diluted to 62.5× concentration with distilled water prior to use. For the buffer screen (see Appendix 3) 5 µl of 5×buffer solution was mixed in a reservoir of the 96-well plate with sufficient amount of water to reach final volume of 25 µl after addition of the protein and the dye. For the additive screen (Appendix 3), 5 µl of suitable 5× additive solution was added. The protein and dye were added last to the wells. The volume of both was 2 µl in order to reach final concentrations of 1.6 – 8.0 µM for the protein and 5× concentration for the dye. After the addition of the protein and the dye the plate was sealed with highly transparent optical-clear quality sealing tape (Greiner Bio-

one) to avoid evaporation and spun down at 4°C at 4000 rpm for 30 s to properly mix the reaction components and to remove airbubbles. The RT-PCR machine was programmed to equilibrate the plate for 5 min at 25°C before the start of the experiment. The plate was subsequently heated from 5°C to 95°C using a temperature ramp of 1°C/min. The fluorescence readings were followed during the temperature ramping by using excitation and emission values of 485/20 nm and 530/30 nm, respectively. These values are optimized for the chosen SYPRO Orange dye.

2.22 Mass Spectrometry

Mass Spectrometry (MS) is a technique that is used to analyze the atomic or molecular components of a sample. The ionized atoms or molecules produce spectra where their abundance and composition can be obtained based on the mass-to-charge ratio (*m/z*). Mass Spectrometry is a widely used technique in proteomics where it can be employed to analyze the primary sequences, post-translational modifications and protein-protein interactions even from more complex protein samples thus gaining knowledge of the function of the protein (Aebersold, R. *et al*. 2003).

The basic elements of MS instrument include an ion source, a mass analyzer for measuring the mass-to-charge ratio of the ionized particles and a detector to register the ions at each m/z-value and the number of these ions. There are several ways to ionize the sample but in proteomics the two most commonly used techniques are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI).

The accurate mass of the EDR1 kinase domain and its inactive version (EDR1-D792N) were analyzed with ESI-MS. Additionally a peptide-fingerprinting analysis with MALDI-TOF was performed for the wild type EDR1 kinase domain to identify the putative phosphorylation sites. For the analysis the protein samples were concentrated to

approximately 5 mg/ml concentration using Spin Concentrators with 10 kDa cut-off (Corning® Spin-X® UF 20mL Centrifugal Concentrator, 10,000 MWCO Membrane, Corning Incorporated). The MS analysis of the EDR1 samples was conducted by Stephan Leicht at the Proteomics Core facility of EMBL Heidelberg outstation. The protocol for the MS analysis including sample handling, data acquisition and data analysis are shown in appendix 4.

2.23 Kinase assay

The kinase activity of EDR1 constructs was evaluated using the coupled kinase assay originally developed by Cook *et al.* (1982). In this assay the consumption of ATP to produce ADP is coupled with the formation of pyruvate from phosphoenolpyruvate (PEP) and subsequent NADH-dependent reduction of pyruvate to lactate by lactic dehydrogenase (Figure 10).

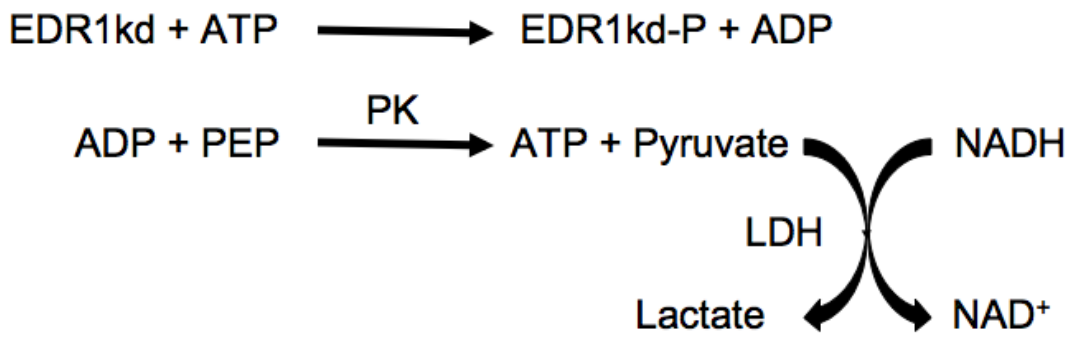

Figure 10. The principle of the coupled kinase assay. The reaction for the studied EDR1 kinase domain is shown as example.

The measurement buffer consisted of 50 mM Hepes at pH7.5, 350 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 1% glycerol, 1 mM NADH, 2 mM PEP and 2.2 U/1.5 U lactic dehydrogenase/puryvate kinase (Sigma). Also the enzyme was added to the buffer in enzymatic concentration of 250 nM. The reaction were started by adding ATP to 2 mM

concentration to one of the wells and then making a serial dilution from this well, having the final ATP concentrations between 0.03125 mM and 2 mM. The measurements were performed in 96well plates using Tecan Infinite M1000 microplate reader (Tecan Group Ltd., Männedorf, Switzerland) set up to measure the absorbance at 340 mm wavelength. The absorbance value was measured on 30 second intervals for a total of one hour. For the linear part of the obtained absorbance curve the standard Michealis-Menten kinetics calculations were performed using GraphPad Prism 5 software (GraphPad Software, Inc., La Jolla, USA).

2.24 Hydroxylamine-trapping assay

The activity of ApnA A_1 PCC7821 was examined using a hydroxylamine-trapping assay, which is based on the protocol published by Kadi&Challis (2009) to study the substrate specificity of NRPS-independent Siderophore (NIS) synthetases. The original assay was created in 1950s as a colorimetric method to detect acyl phosphates through their reaction with hydroxylamine (Lipmann & Tuttle, 1945). The reaction between acyl phosphates and hydroxylamine results in formation of hydroxamic acid, which in the presence of trivalent iron (Fe^{3+}) forms colorful complexes with absorption maxima between 540 nm and 480 nm (Figure 11).

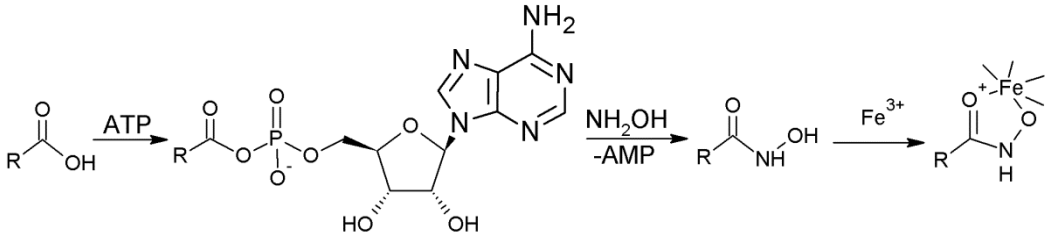

Figure 11. The overall reaction mechanism in the hydroxylamine-trapping assay. In the presence of hydroxylamine the formation of adenylate intermediate is followed by formation of hydroxamic acid. The subsequent reaction between the hydroxamic acid and trivalent iron leads to colorful complex that can be detected by measuring its absorbance at 540 nm.

The hydroxylamine trapping assays were conducted by Stephan Schiefelbein (University of Münster). The activity of wild type ApnA A_1 PCC7821 and the enzyme variants were tested against the 20 canonical aminoacids, excluding the selenomethionine. Additionally the wild type enzyme and the specific mutants were tested against two unnatural amino acids, 4-Azido-phenylalanine (Az) and 4-Fluoro-phenylalanine. All protein samples were diluted to a concentration of 40 µM with the assay buffer that contained 50 mM TRIS at pH 8 and 30 mM $MgCl_2$. The reaction mixture was prepared in a reservoir of a 96-well plate by mixing 2.3 µL of the protein solution with 3.5 µL hydroxylamine (NH_2OH) solution at concentration 2 M (freshly prepared aqueous solution) and 35.2 µL of 3.88 mM amino acid solution. To iniate the reaction, 4.5 µL of 30 mM ATP solution was added to the reaction mixture. The final reaction mixture contained 25 mM TRIS at pH 8, 15 mM $MgCl_2$, 2.0 µM enzyme, 150 mM NH_2OH, 3 mM ATP and 3 mM amino acid. The plate was then incubated for 4 h at 25°C with gentle shaking. Prior to the absorbance measurements, the reaction was stopped by addition of stopping solution, which contained 10% (w/v) $FeCl_3*6H_2O$ and 3.3% trichloroacetic acid (TCA) in 0.7 M HCl. The absorption was measured at wavelength of 540 nm using Tecan Infinite® M1000 PRO microplate reader (Tecan Group Ltd., Männedorf, Switzerland).

2.25 Crystallization

The initial crystallization trials were conducted at the high-throughput crystallization facility at EMBL Hamburg (Müller-Dieckmann *et al.*, 2006) using commercial screens from Qiagen. All initial screens were performed using the sitting-drop vapour-diffusion method at 292K in 96-well Greiner or iQ plates (TTP Labtech Ltd,

Melbourn, UK). In the default experimental set up the drop contained 300 nl protein solution mixed with an equal volume of reservoir solution, and was equilibrated against 50 µl of reservoir solution. Based on the obtained lead conditions, manual optimization was performed at 292 K using 24-well Linbro plates (Molecular Dimensions Limited, UK) and hanging drop vapor diffusion method. For the inactive kinase domain of EDR1 (EDR1-D792N) 5 mM AMP-PNP, a non-hydrolyzable analog of ATP was added to protein sample at 10 mg/ml concentration prior to the crystallization experiments. This protein produced well-diffracting crystals from solution containing 1.2 M $LiSO_4$, 10% Polyethylene glycol (PEG) 3350 and 0.1 M HEPES at pH 7.8 and these crystals were used for the data collection.

The manual crystallization trials for ApnA A_1 PCC7821 were conducted by Sandra Kozak (Sample Preparation and Characterisation Facility, EMBL Hamburg). For ApnA A_1 PCC7821 the protein concentrations of 5, 10 and 13 mg/ml were used for the manual crystallization trials. Well-diffracting crystals were obtained in several conditions and with all tested protein concentrations. However, for data collection and soaking experiments crystals from conditions containing 24-30% Polyethylene glycol 3350 and either 0.1 M MES or 0.1 M BIS-TRIS buffer in pH 5.5 to 6.5 were used.

2.26 Data Collection and Processing

The data collection was conducted at PETRA III beamlines P13 and P14 at EMBL Hamburg, Germany. P13 is a tunable beam line with energy range between 6 and 15 keV and is equipped with MD2 type diffractometer with mini-kappa goniostat. P14 on the other hand is a beam line especially suited for very small crystals due to the availability of micro-beam with a beam-size of 5 x 10 micron^2 and beam-divergence below 0.3 mrad. Additionally the beam line provides a possibility to use an unfocused beam with a

maximum size of 150 micron for larger crystals. The diffractometer at P14 is type MD3. Both beam lines are equipped with PILATUS 6M detectors enabling shutter-less oscillation data collection.

Prior to data collection the protein crystals were soaked in suitable cryo-solution containing glycerol as cryo-protectant and then flash-cooled to 77 K in liquid nitrogen. The cryo-solution for EDR1-D792N crystals contained in addition to the original crystallization solution, 15 % glycerol and 10 mM AMP-PNP. From the best diffracting crystal a data set of 800 images was collected with an oscillation range of 0.1° at beam line P13.

The basic cryo-solution for the ApnA A_1 PCC7821 crystals contained the original crystallization solution with 5% higher PEG3350 concentration, 10% glycerol as additional cryo-protectant and also 4 mM $MgCl_2$. To obtain the structure with the substrate or adenylation reaction product bound, the crystals of the native protein were soaked for 5 to 10 seconds in a cryo-solution containing the suitable substrate or substrates as listed in the table 25. The data for the adenylate complex structures were collected at beam line P13 whereas the data for the AMP-PNP containing structures were collected at beam line P14. From the crystal soaked in AMP-PNP alone a data set of 300 images were collected at oscillation range 0.5° and the same oscillation range but with total rotation range of 180° was used for the crystal soaked in AMP-PNP and L-arginine containing solution. Both the crystal soaked in ATP and L-Tyrosine containing cryo-solution as well as the one incubated in ATP and L-Arginine containing cryo-solution was used to collect a data set of 1800 images but with oscillation ranges of 0.1° and 0.2°, respectively.

Table 25. The additives used in the soaking experiments for ApnA A_1 PCC7821.

Complex structure ID	Additives in soaking solution
Complex 1	2 mM AMP-PNP
Complex 2	2 mM AMP-PNP and 50 mM L-Arginine
Complex 3	2 mM ATP and 50 mM L-Arginine
Complex 4	2 mM ATP and a few milligram of powdery L-Tyrosine

The data for both EDR1-D792N and for *P. agardii* ApnA A_1 PCC7821 were indexed and integrated using *XDS* (Kabsch, 2010) and scaled with *SCALA* (Collaborative Computational Project, Number 4, 1994).

2.27 Structure determination and validation

The structure determination was performed for both EDR1-D792N and for ApnA A_1 PCC7821 using automated molecular replacement and program *MOLREP 11.0* (Vagin&Teplyakov, 1997). The initial model was refined against the native dataset using *REFMAC 5.7* (Murshudov *et al*., 1997) as part of *CCP4 Program Suite 6.3.0* (Potterton *et al*., 2003). This was followed by iterative cycles of manual model building and refinement using programs *COOT 0.7* (Emsley *et al*., 2010) and *REFMAC 5.7*, respectively.

The tyrosyl adenylate and arginyl adenylate in the ApnA A_1 PCC7821 complex structures were created with *JLigand* (Lebedev et al., 2012) using existing restrains from the standard Refmac library to define the link between the monomeric units. The L-arginine (PDB Ligand code: ARG) and AMP-PNP (PDB Ligand code: ANP) were readily available from *COOT* monomer library. The final models were validated through analysis with *MolProbity* (Chen *et al*., 2010). The structure images were created with the

program *PYMOL* (The PyMOL Molecular Graphics System, Version 1.6.x, Schrödinger, LLC).

2.28 Small Angle X-ray Scattering

SAXS measurements were performed at PETRA III beamline P12 (EMBL/DESY, Hamburg, Germany)(Roessle *et al.*, 2007). The beamline is equipped with a Pilatus 2M detector (Dectris). For the ETR1-ΔTM construct, the dialyzed protein sample was concentrated to suitable concentration using Corning Spin Concentrators with 30 kDa cut-off (Corning® Spin-X® UF 20mL Centrifugal Concentrator, 30,000 MWCO Membrane, Corning Incorporated) and filtered with a prewashed spin filter (0.22 µm membrane, Millipore). Due to the high aggregation tendency of the sample, measurements were conducted at lower protein concentrations (0.3 – 1.4mg/ml). The sample concentrations were measured using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and an extinction coefficient of 0.502 for the absorbance value at 280 nm. Each measurement consisted of twenty 50 ms frames, which were averaged using a sample volume of 30 µl at a temperature of 10 °C. A sample-detector distance of 3.1 m was used to cover the momentum transfer range $0.008\ Å^{-1} < s < 0.47\ Å^{-1}$ ($s = 4\pi \sin(\theta)/\lambda$ where 2θ is the scattering angle and λ=1.24 Å is the X-ray wavelength). A solvent blank was measured prior to and following each sample exposure to enable background substraction.

The data were processed using standard procedures, corrected for buffer contribution, scaled for solute concentration and extrapolated to infinite dilution using the program *PRIMUS* (Konarev *et al.*, 2003). The radius of gyration R_g and forward scattering intensity *I(0)* were determined from the scattering profile using Guinier approximation (Guinier, 1938). This relies on the assumption that at very small scattering

angles ($s \leq 1.3/R_g$) the intensity is represented as $I(s)=I(0)\exp((sR_g)^2/3)$. As a mean to validate the obtained R_g and $I(0)$ values, they were also independently calculated using the indirect Fourier transform of the scattering curve with program *GNOM* (Svergun, 1992). The resulting interparticle distance distribution function $P(r)$ was also used to determine the maximum particle dimension D_{max}. The interpolated forward scattering from a reference bovine serum albumin (BSA) sample (calculated molecular mass 66kDa) was used as comparison to calculate the molecular mass (MM_{SAXS}) of the construct. Another molecular mass estimation was derived from the small angle region of the data ($s < 0.25$ Å^{-1}) using the Porod equation (Porod, 1982). This gives the excluded volume of the hydrated particle V_p (Å), which is for globular protein about 1.7 times of the molecular mass in Da (Petoukhov *et al*., 2012). Available high-resolution structures were used to calculate the scattering pattern of the SAXS model and define its fit to the experimental data using the program *CRYSOL* (Svergun *et al*., 1995). The discrepancy χ^2 between the measured and calculated SAXS profiles is described by the following equation:

$$\chi^2 = \frac{1}{N-1}\sum_j \left[\frac{I_{exp}(s_j) - cI_{calc}(s_j)}{\sigma(s_j)}\right]^2$$

where N is the number of experimental points, c is a scaling factor, $I_{calc}(s_j)$ and $I_{exp}(s_j)$ are the calculated and experimental scattering intensity and $\sigma(s_j)$ experimental error at the momentum transfer s_j.

4. RESULTS

4.1 Full-length ETR1

4.1.1 Cloning

The gene for the full-length Ethylene Receptor 1 of *A. thaliana* (ETR1, residues 1-738) was originally amplified by PCR from a cDNA library obtained from the Arabidopsis Biological Research Center at Ohio State University (Kieber *et al*., 1993). This gene was cloned into pETM-11/LIC using LIC-cloning and into pET15b with restriction endonuclease dependent method. To make the expression of the full-length receptor more feasible in *E. coli*, a synthetic gene codon-optimized specifically for this expression system was ordered from Genscript. The synthetic ETR1 gene was obtained both in pUC57 vector and baculovirus vector Fastbac HTa. The ETR1 gene in pUC57 was amplified by PCR and sub-cloned into pETM-11/LIC, pETM11SUMO3GFP and pET15b vectors. All three vectors contain N-terminal His_6-tag. In pETM11SUMO3GFP the cloning sites, namely NcoI and BamHI, were chosen so that a C-terminal GFP fusion was added to the construct. Cloning experiments were conducted in the *E. coli* strain DH5α (Life Technologies GmbH, Darmstadt, Germany). The correct position and sequence of the insert in the target vectors was verified by sequencing (Eurofins MWG Operon, Ebersberg, Germany).

4.1.2 Expression trials in E. coli

The expression of the full-length ETR1 was attempted according to the published protocol from Voet-van-Vormizeele & Groth (2003). To this end, the gene in the pET15b vector was transformed to both C41(DE3) and C43(DE3), which are *E. coli* strain optimized for membrane protein expression (Miroux & Walker, 1996). Additionally, the expression in C43(DE3) cells were tested for the pETM-11/LIC and pETM11SUMO3GFP constructs. The expression tests were conducted in small-scale cultures of both 5 and 50 ml using 2YT media. The cultures were grown at 37°C to OD600-value of 0.4 – 0.6 after which they were induced with 0.5 mM IPTG. The culturing was then continued at 25 or 30°C for 6 hours since according to the paper by Vormizeele & Groth, at the 6^{th} hour the expression of the protein was at its peak. Un-induced cultures were grown in parallel for comparison. To analyze the expression of the target protein, 100 µl samples were taken immediately before induction and then every 2 hours from both induced and un-induced cultures. These samples were centrifuged at 4000 rpm for 2 min at room temperature and remaining pellet re-suspended in 60 µl of standard SDS-PAGE sample buffer (see materials and methods) with β-mercaptoethanol as reducing agent.

To evaluate the expression of the FL ETR1, the cell culture samples were run on SDS-PAGE (10% Bio-Rad Mini gel). They were also analyzed by Western blotting using antibodies against the His6-tag. No expression was detected for the target protein with either of the tested cell strains. Since the ETR1 expression was tested in parallel in baculovirus-insect cell expression system and led to successful production of the protein, the expression in *E. coli* was not continued further. Figure 12 shows the SDS-PAGE and Western blot analysis of ETR1 expression samples from C43(DE3) cells 6 hours after

induction for the three chosen vectors. Culture samples from insect cells were used as a positive control.

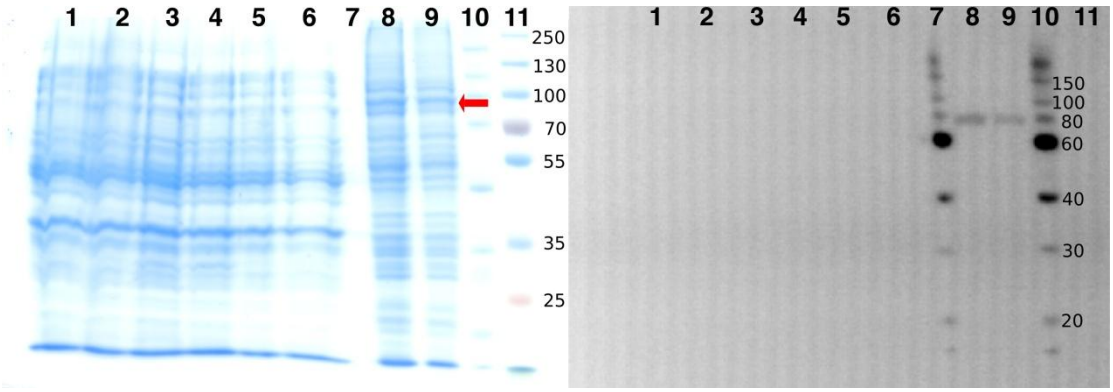

Figure 12. The summary of the *E. coli* expression trials of ETR1. The conditions in the trials were based on the published protocol by Voet-van-Vormizeele & Groth (2003). The samples in the lanes are as following: 1) FL ETR1 in pET15b, un-induced 2) FL ETR1 in pET15b, induced 3) FL ETR1 in pETM-11/LIC, un-induced 4) FL ETR1 in pETM-11/LIC, induced 5) FL ETR1 in pETM11SUMO3GFP, un-induced 6) FL ETR1 in pETM11SUMO3GFP, induced 7) Roti®-Mark 10-150 plus protein marker (Carl Roth GmbH & Co. KG, Karlsruhe, Germany) 8) and 9) Cell culture samples from insect cells expressing FL ETR1 as a positive control 10) As in the lane 7 11) PageRuler Plus Prestained protein ladder (Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany).

4.1.3 Expression in insect cells

The plasmid for the insect cell was ordered from Genscript and contained the *E. coli* codon-optimized synthetic gene for the full-length ETR1 (residues 1 – 738) in pFastBac HTa vector. The preparation of the bacmid and the initial transfection of Sf21 cells on a 6well-plate were performed as stated in the materials and methods. The V_0 virus that was harvested 60 h after the transfection was used to infect fresh batch of Sf21 cells for preparation of the V_1 virus. These V_1 cultures were split to keep optimal cell density of 0.5 – 1.0 Mio/ml until the cells stopped dividing and thus had reached the day

of the proliferation arrest (dpa). The samples taken on dpa and on 24-hour intervals after dpa were analyzed for protein expression. To this end, the cell samples were thawn gently on ice and resuspended to 500 µl of PBS buffer. The resuspended cells were sonicated for 10 s with 25 % amplitude (Bandelin Sonopuls2200, Bandelin electronic GmbH & Co. KG, Berlin, Germany). A small aliquot of the lysate was set aside for analysis on SDS-PAGE gel. The remaining lysate was spun for 3 min at 13,400 rpm (Eppendorf Minispin, Eppendorf AG, Hamburg, Germany) at room temperature. From the supernatant YFP fluorescence was measured using an excitation wavelength of 488 nm and an emission of wavelength of 520 nm with TECAN Infinite M1000 microplate reader (Tecan Group Ltd., Männedorf, Switzerland). YFP was expressed simultaneously with the target protein and its levels should correlate with the amount of protein expressed, i.e. the YFP fluorescence should reach a plateu when the maximum level of expression is reached. However, the expression tests for ETR1 in insect cells did not seem to show the expected pattern for YFP. Instead, a peak of YFP signal was reached 24 hours after dpa and after this point the signal appeared to decrease significantly (Figure 13). Furthermore, 48 hours after dpa the YFP signal had dropped below the signal at the beginning (at dpa). Based on the YFP fluorescence measurements, the sonicated cell samples from dpa, 24 hours after dpa and 48 hours after dpa were analyzed by SDS-PAGE and Western Blot. A band corresponding to the molecular mass of ETR1 monomer was observed, which confirmed the successful expression of the protein (Figure 14). Furthermore, it was found that the ETR1 expression level did not show a similar decrease as was observed for YFP. As a possible explanation for the observed difference, it was postulated that the insertion of the overexpressed receptor to the insect cell membrane had caused breakage of the membrane and the release of the soluble YFP to the culture media. The cell samples taken from the

later stages of expression thus presumably contained more broken cells with the receptor still attached but very little YFP since it had been lost in the media.

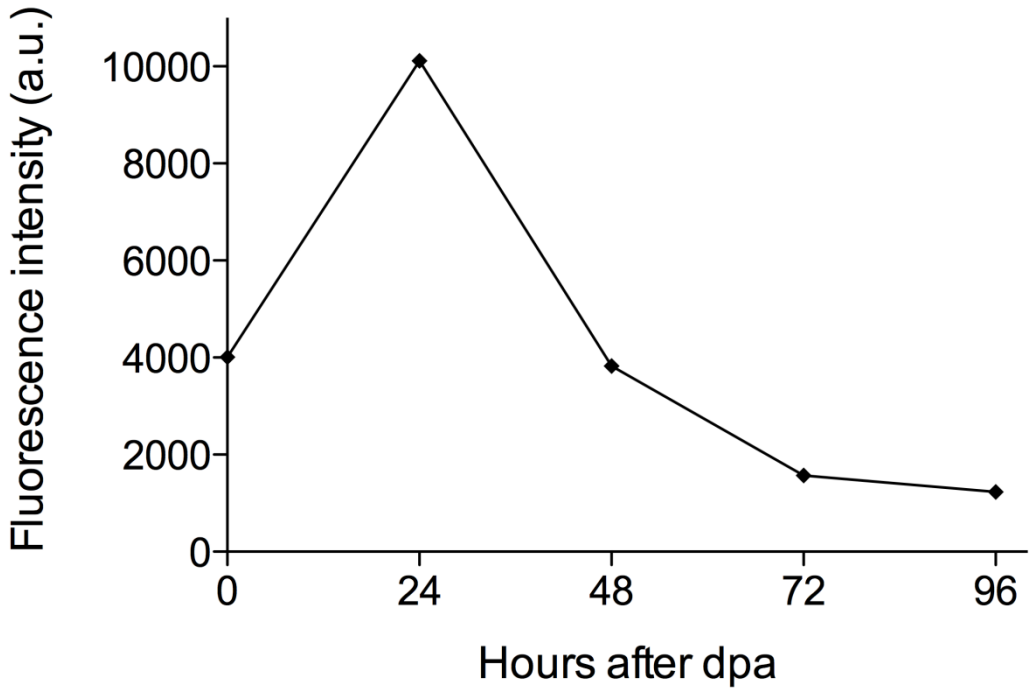

Figure 13. The measured fluorescence signal for YFP (excitation at 488 nm, emission at 520 nm) in the ETR1 expression samples after the day of proliferation arrest (dpa). YFP is expressed simultaneously with the target protein. The decrease in the YFP signal indicated that the insect cells had started to break due to insertion of ETR1 to the membrane thus resulting in release of the YFP to the culture media.

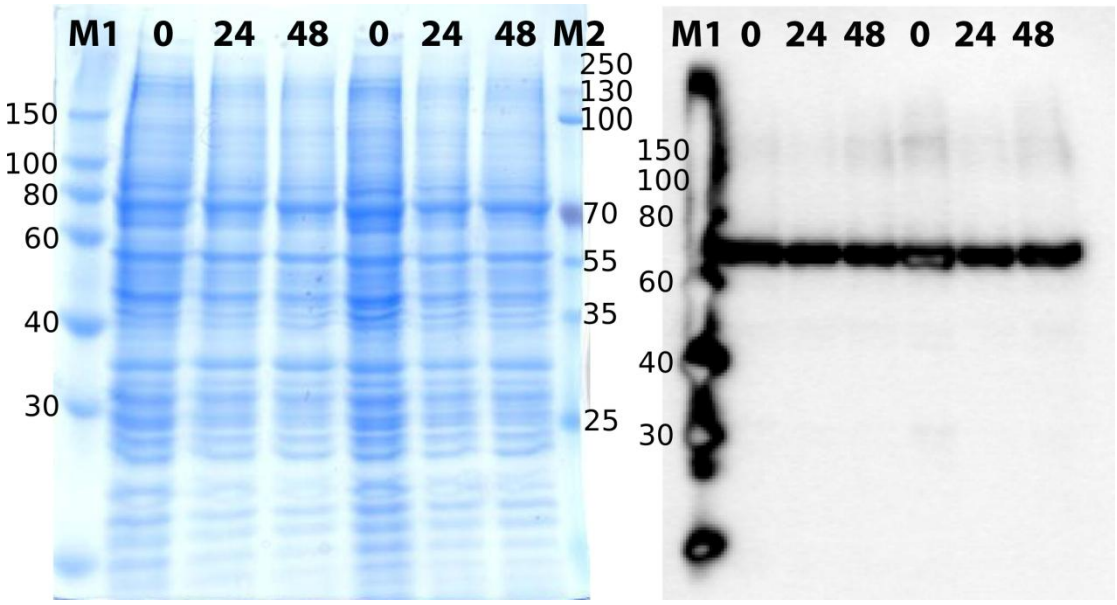

Figure 14. The results of the SDS-PAGE (on the left) and Western blot (on the right) for ETR1 V_0 insect cells cultures. Samples taken at dpa (0), 24 hours and 48 hours after dpa were analyzed from two cultures grown in parallel. In SDS-PAGE gel the protein band for ETR1 (MW 82.6 kDa) appeared above the 70 kDa marker band and in the Westerb blot between the 60 kDa and 70 kDa marker bands. The protein markers used were PageRuler Color Plus Prestained protein ladder (M2, Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany) for SDS-PAGE and Roti®-Mark 10-150 plus protein marker (M1, Carl Roth GmbH & Co. KG, Karlsruhe, Germany) for Western blot.

The V_1 cultures were harvested at 96th hour after dpa by centrifugation at 800 rpm at 4°C. The supernatant contained the V_1 virus, which was further used to produce the BIIC stocks according to the protocol obtained from Berger lab at EMBL Grenoble and as described in the materials and methods. The prepared BIIC stocks were used for protein production. To this end, one aliquot of the BIIC stocks were diluted to 100 ml with HyClone culture media and added to 900 ml of Sf21 cells at cell density of approximately 0.9 Mio/ml. As with the V_1 cultures, the cultures were kept at the recommended density of about 1 Mio/ml until they stopped dividing (dpa reached) and started the protein expression. Due to the cell damage caused by ETR1 expression and assumed membrane localization, the cultures were harvested at 48th hour after dpa. The

cell pellet was stored at -20°C until further use. Samples were taken and handled in a similar manner than for the V_1 cultures. Both the pellet and the supernatant from the sonicated and centrifuged cell suspension were analyzed using SDS-PAGE and immunoblotting with antibodies agaist the N-terminal His_6-tag. A sample from an uninfected culture was used as a control. The results are summarized in figure 15.

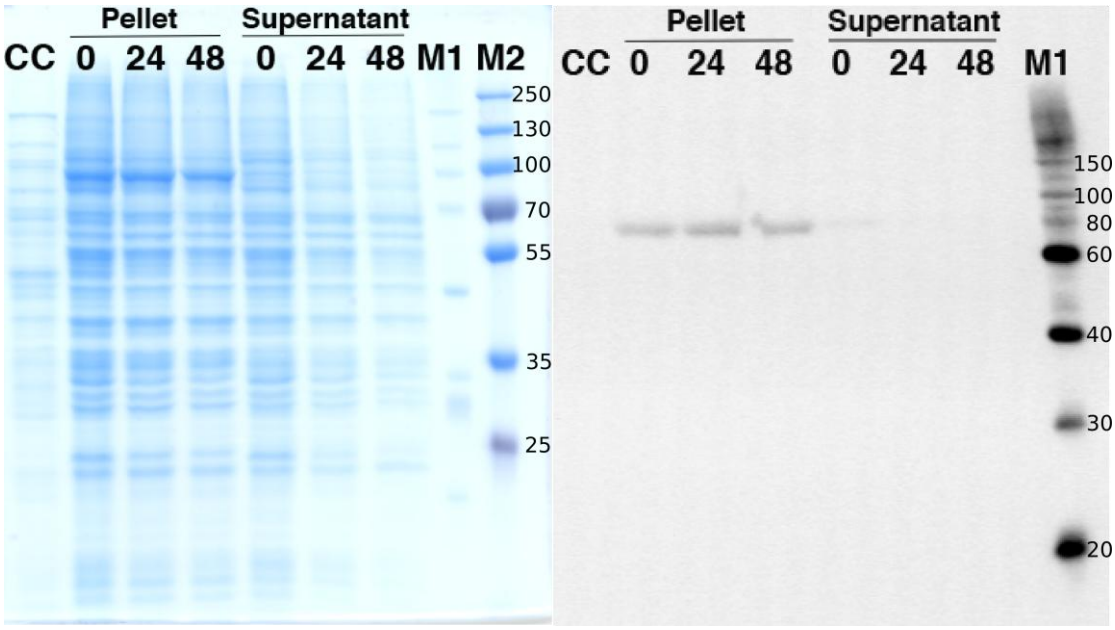

Figure 15. The results of the large-scale expression of full-length ETR1 in insect cells. The samples on the lanes correspond to the pellet and supernatant of the sonicated and centrifuged culture samples, which were taken at dpa (0), 24 hours after dap and 48 hours after dpa. A sample from the uninfected culture provided the cell control (CC). The SDS-PAGE is shown in the left panel and the western blot in the right one. The band below the 100 kDa marker band (M2, PageRuler Color Plus Prestained protein ladder, Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany) in the SDS-PAGE gel for the pellet samples correspond to ETR1. Similarly, a protein band corresponding to a molecular mass of approximately 80 kDa is seen in the pellet fractions on the Western blot.

4.1.4 Isolation from the membrane

The next step after the successful expression of ETR1 in insect cells was its isolation from the membrane. A protocol from Dr. Morth (Centre for Molecular Medicine

Norway (NCMM), Blindern, Oslo, Norway) was used as a base for the isolation and purification tests of ETR1. A schematic representation for the steps in the protocol is shown in the figure 16 (for more in detail describtion, see materials and methods). As in a standard isolation of a soluble protein, also in the membrane protein isolation the target protein should remain in the cleared lysate after the first centrifugation step (18,000 rpm rotational speed). However, membrane proteins are not readily soluble but remain attached to small membrane drafts, which are subsequently pelleted using high-speed centrifugation (i.e. ultracentrifugation). The resulting soft pellet is washed to remove residual impurities from the cells and from the culture medium. After the pellet wash step, the target protein is solubilized with a suitable detergent, which functions by breaking the remaining cell membranes and forming a protecting layer around the hydrophobic, membrane-embedded parts of the protein. As the task of the detergent is to mimic the native lipid environment of the protein, they usually have lipid-like structure and consist of hydrophilic head group and hydrophobic tail. Most commonly non-ionic detergents, such as octyl-β-D-glucopyranoside (OG), decyl-/dodecyl-β-D-maltopyranoside (DM/DDM) are used in initial solubilization step as they have a good solubilization capacity but don't cause denaturation or inactivation of the protein. Other important requirements for the detergent include a suitable critical micelle concentration (CMC) and micelle size. The CMC-value of the detergent defines the concentration above which the detergent forms micelles. As in the standard solubilization experiment a detergent concentration above CMC is used, detergents with lower CMC values are preferred. The concentration of the detergent is in the later purification steps decreased below CMC to avoid formation of micelles, which can interfere with the subsequent characterization methods. For the further use of the protein in applications downstream of

the membrane protein isolation and purification, detergents with small micelle sizes are usually preferred.

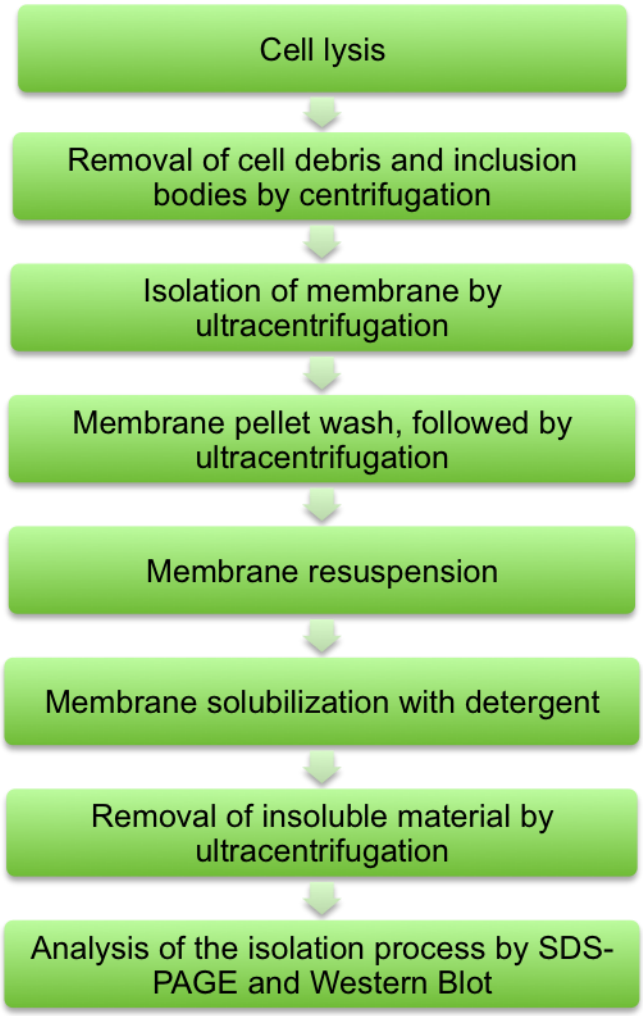

Figure 16. Schematic representation for the ETR1 isolation from insect cell membrane. Prior to cell lysis step the cell pellet wash additionally washed twice with lysis buffer. Sonication was used to break the cells in the cell lysis step.

The samples taken from the different isolation steps of ETR1 were analyzed with SDS-PAGE and Western blot using antibodies against the His_6-tag in order to evaluate the success of the isolation process. In the first isolation trial, 0.5 % DDM (CMC = 0.008%, Sigma-Aldrich) was used in the solubilization step. The chosen solubilization

conditions were insufficient to bring the protein in soluble form since all the protein remained in the pellet after the overnight solubilization and final ultracentrifugation steps. The results of this trial are shown in the figure 17.

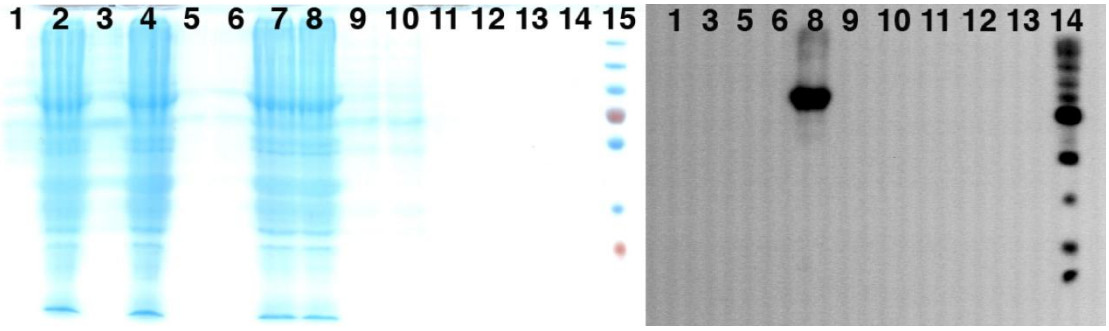

Figure 17. The results of the first isolation trial of ETR1. The samples in the lanes are as following: 1) Pellet washes 2) Crude lysate 3) Cleared lysate 4) Cell debris and inclusion bodies 5) Supernatant after 1st ultracentrifugation (UC) 6) Supernatant after 2nd UC 7) Membrane before solubilization 8) Insoluble material after 3rd UC 9) Soluble fraction 10) As in lane 9 11) - 12) - 13) - 14) Roti®-Mark 10-150 plus protein marker (Carl Roth GmbH & Co. KG, Karlsruhe, Germany) 15) PageRuler Color Plus Prestained protein ladder (Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany).

As an attempt to improve the solubilization of the receptor, the composition of the isolation buffers was modified. To this end, the pH and the NaCl concentration was varied. In the initial isolation trial, TRIS buffer at pH 7.6 was used in combination with 100 Mm NaCl, with the exception of the membrane resuspension buffer where the NaCl concentration was 200 mM. For the buffer optimization trials, TRIS buffer at pH 8 and MES buffer at pH 6 was tested both in combination with lower (20 mM) and higher (500 mM) NaCl concentration. Additionally, the concentration of the detergent was increased from 0.5% to 1.0% in the solubilization step. DDM was used as the detergent as in the initial trial. To evaluate the success of the isolation, selected samples were analyzed in a similar manner than in the initial isolation trial (Figure 18). The change of the pH and NaCl concentration appeared to have an effect only when low salt concentration was

combined with higher pH (pH 8) or vice versa (high salt, pH 6). Selected soluble fractions from these two isolation trials were re-analyzed on SDS-PAGE and Western blot to avoid possible contaminant effect from the pellet fractions (Figure 19). It was found that solubilization was still hampered although more protein was recovered in cell lysis step. Due to the limited access to the equipment required for membrane protein isolation, more solubilization trials were not performed. As the solubilization step appeared to be the biggest bottleneck in the ETR1 isolation from the insect cell membrane, future work would require the testing of other detergents.

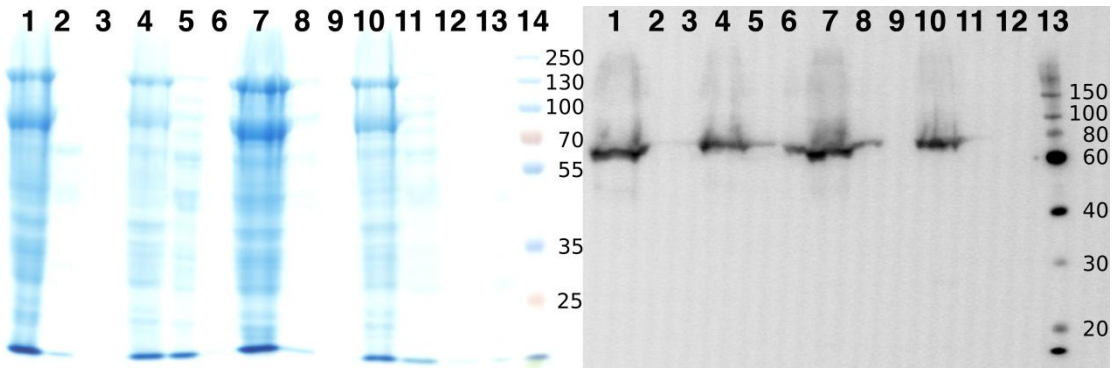

Figure 18. SDS-PAGE gel and Western blot of the selected samples from ETR1 isolation steps with low NaCl concentration at pH 6 and at pH 8 (abbreviations L6 and L8, respectively) or with high NaCl concentration at pH 6 and at pH 8 (H6 and H8). The lanes in the gel and Western blot are as follows: 1) L6 crude lysate 2) L6 cleared lysate (UC) 3) L6 soluble material after overnight solubilization 4) L8 crude lysate 5) L8 cleared lysate 6) L8 soluble material after overnight solubilization 7) H6 crude lysate 8) H6 cleared lysate 9) H6 soluble material after overnight solubilization 10) H8 crude lysate 11) H8 cleared lysate 12) H8 soluble material after overnight solubilization 13) Roti®-Mark 10-150 plus protein marker (Carl Roth GmbH & Co. KG, Karlsruhe, Germany) 14) PageRuler Color Plus Prestained protein ladder (Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany).

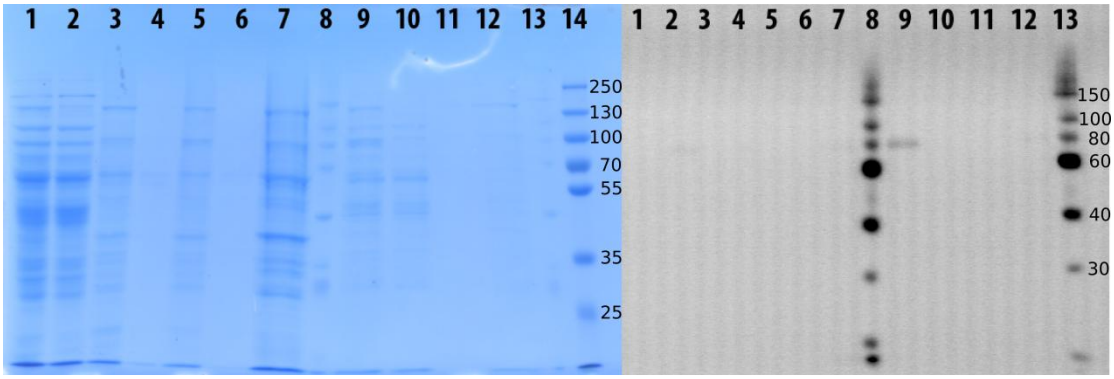

Figure 19. Re-analysis with SDS-PAGE and Western blot for the selected samples from ETR1 isolation steps with low NaCl concentration at pH 8 (L8) and high NaCl concentration at pH 6 (H6). The lanes in the gel and Western blot are as follows: 1) L8 cleared lysate 2) L8 supernatant after 1st ultracentrifugation (UC) 3) L8 resuspended membrane after 2nd UC 4) L8 supernatant after 2nd UC 5) L8 sample after solubilization 6) L8 soluble material (supernatant from 3rd UC) 7) L8 insoluble material after 3rd UC 8) Roti®-Mark 10-150 plus protein marker (Carl Roth GmbH & Co. KG, Karlsruhe, Germany) 9) H6 cleared lysate 10) H6 supernatant after 1st UC 11) H6 soluble material (supernatant from 3rd UC) 12) H6 insoluble material after 3rd UC 13) Same as lane 8 14) PageRuler Color Plus Prestained protein ladder (Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany).

4.2 ETR1-ΔTM

4.2.1 Cloning, expression and purification

The cytosolic part of the main ethylene receptor ETR1 was expressed and purified for structural studies with both X-ray crystallography and Small Angle X-ray scattering method. This ETR1-ΔTM construct consisted of the C-terminal residues 158 – 738 and included the GAF domain, DHp and CA subdomains as part of the HK domain and finally the RR domain (Figure 20).

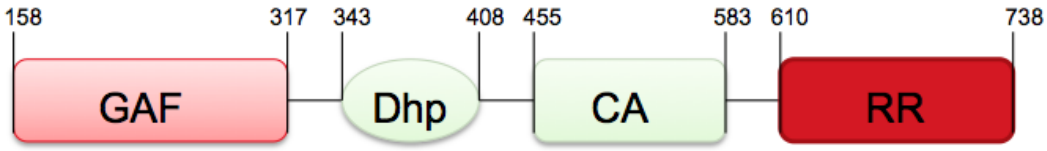

Figure 20. Schematic representation of the ETR1-ΔTM domain organization. The residues that define the domain borders are shown in numbers.

The ETR1-ΔTM had already been previously studied using SAXS and also crystallization trials had been performed, and therefore established expression and purification protocol was available (Mayerhofer, 2011). The plasmid containing ETR1-ΔTM gene from *A. thaliana* in pET-GB-1a/LIC vector was also obtained from H. Mayerhofer. Site-directed mutagenesis was used to introduce four additional histidines to the already existing His_6-affinity tag. This was done in order to improve the binding of the protein to the Ni-affinity matrix, resulting in very pure protein material already after NiNTA purification step and eliminating the need to for SEC step.

The ETR1-ΔTM was expressed in *E. coli* strain Rosetta 2 (DE3) using 1 L culture volume and TB-media. The expression was induced by addition of IPTG to final concentration of 50 µM. The cultures were harvested after 18 h at 20°C by centrifugation. For purification the cell pellet was resuspended to lysis buffer and the cells were lysed by sonication on ice. The resulting lysate was centrifuged at 19000 rpm (SS-34 rotor, Sorvall), filtered through 0.45 µm membrane and loaded to 5 ml NiNTA column (Qiagen NiNTA Superflow cartridge column), which had been pre-equilibrated with washing buffer. The beads were then washed using imidazole gradient to 17% of elution buffer in washing buffer (51 mM Imidazole). The protein was eluted in imidazole gradient from 17% to 100% elution buffer (from 51 mM to 300 mM imidazole). The protein eluted mainly with imidazole concentration higher than 150 mM resulting in very pure protein

material. The different NiNTA purification fractions were analyzed using SDS-PAGE, based on which the fractions containing the target protein were pooled together. The NiNTA purification results for ETR1-ΔTM are summarized in the figure 21.

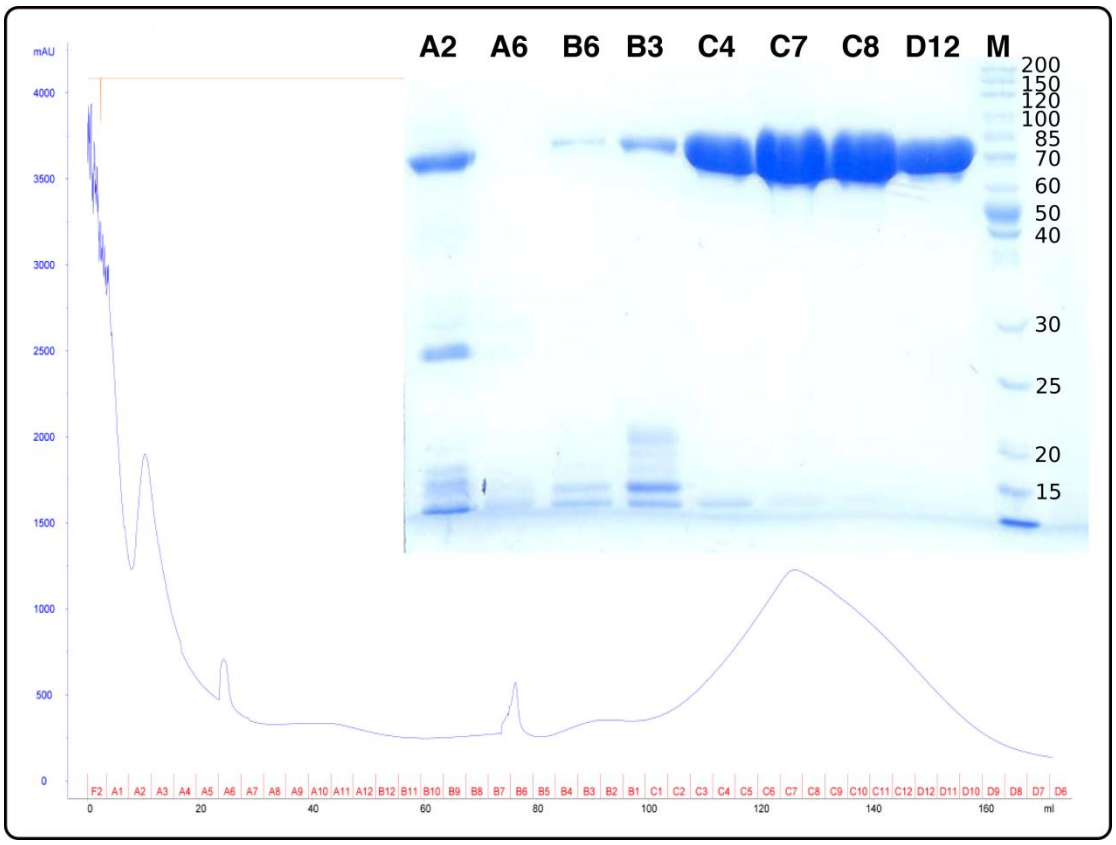

Figure 21. The NiNTA purification graph of ETR1-ΔTM and the corresponding purification fractions shown on SDS-PAGE gel. The fractions A2 - A7 corresponded to the wash gradient (up to 51 mM imidazole) while the fractions A8 – A12 and B12 – B7 comprised the elution gradient (to 300 mM Imidazole). The elution at imidazole concentration 300 mM was continued until a plateau in UV signal at 280 nm was reached. Based on the SDS-PAGE results fractions containing pure protein (C1 – C12 and D12 – D9) were combined. The PageRuler Unstained Protein Ladder (Lane M, Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany) was used as size standard for the SDS-PAGE.

It was observed that upon clevage of the N-terminal GB1-fusion tag together with the His_{10}-tag the protein became highly unstable and began to aggregate. Previously, the stability of the protein had been improved through addition of a non-detergent

sulfobetaine (NDSB) compound 3-(1-Pyridinio)-1-propanesulfonate prior to the SEC run. However, this chemical (referred later simply as NDSB) was found to contribute to the scattering in SAXS measurements, which led to mismatch at the higher angles after solvent substraction. Since thermofluor had been previously successfully employed to find stabilizing compounds for the ETR1-ΔTM construct (Mayerhofer, 2011), a thermofluor experiment was conducted using the Buffer screen of EMBL Hamburg (Boivin *et al*., 2013). The obtained results indicated that increase in NaCl concentration could provide additional stabilization (Figure 22). However, it did not result in more stable protein in the purification trials. Since glycerol has been suggested to prevent protein aggregation (Vagenende *et al*., 2009), addition of this compound in 5% concentration to the dialysis buffer was tested. This appeared to have the desired effect since no visible aggregation formed after TEV treatment. In the test SAXS experiments, it was observed that 5% glycerol also had a minimal effect on the scattering. This problem was solved by reducing the amount of glycerol in the final buffer to 3%. The SDS-PAGE analysis of the ETR1-ΔTM sample before and after TEV-cleavage is shown in the figure 23. The TEV-treated and dialyzed sample was run through Ni-affinity matrix to remove the His-tagged TEV-protease.

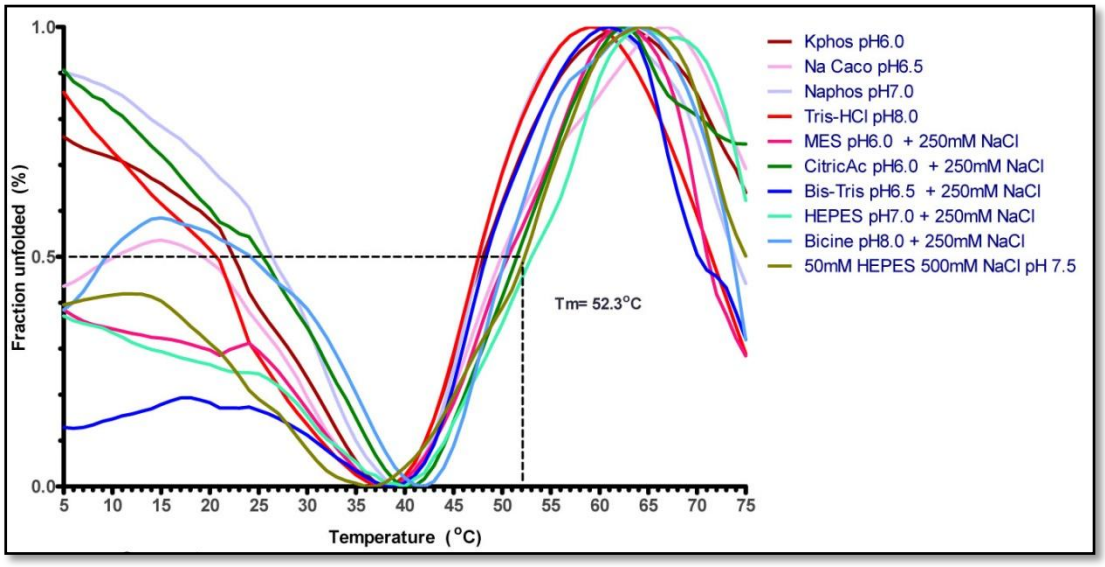

Figure 22. The best buffer conditions for ETR1-ΔTM indentified from the thermofluor experiment. The Buffer screen of EMBL Hamburg (Boivin et al., 2013) was used in this experiment.

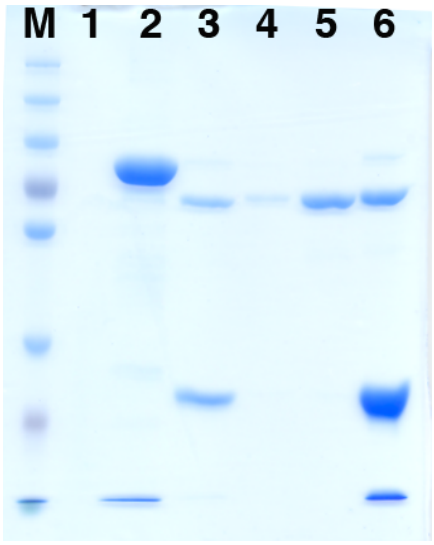

Figure 23. The TEV-treatment results of ETR1-ΔTM. The samples in the lanes are as follows: 1) – 2) Combined NiNTA fraction (before TEV treatment and dialysis) 3) TEV cleaved sample (before 2nd Ni affinity) 4) TEV cleaved diluted sample after 2nd Ni-affinity step 5) As in lane 4 but without dilution 6) Material bound to the Ni-beads after the 2nd Ni-affinity step. Lane M contains the PageRuler Color Plus Prestained protein ladder (Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany).

4.2.2 Crystallization trials

The crystallization trials of ETR1-ΔTM were conducted at the HTx-facility of EMBL Hamburg (Mueller-Dieckmann, 2006). These trials were performed at 19°C for a protein sample with and without the GB1-tag at concentration 4 mg/ml. Higher concentrations could not be tested since the protein became instable and started to precipitate. Four commercial screens (Qiagen Classic I and II, PEGs I and II) were chosen for the initial screening. However, the initial screens didn't result in any promising conditions for further optimization. The main problem appeared to arise from protein aggregation at the concentrations, which are required for crystallization. Since the structural characterization with SAXS does not necessary require high protein concentration, the studies on the ETR1-ΔTM construct were focuzed on the SAXS studies.

4.2.3 Structural characterization by SAXS

Previously the ETR1-ΔTM had been studied using SAXS in the presence of the NDSB compound (Mayerhofer, 2011). The molecular weight estimates calculated from the SAXS scattering curve in these experiments indicated existence of dimeric protein molecule. Furthermore, the *ab initio* and ridig body modeling showed that the two ETR1-ΔTM monomers form together an overall dumbbell shaped dimer. However, due to the presence of the NDSB compound accurate determination of the protein concentration was hampered, which consequently lead to unreliable MM_{SAXS} – value. This molecular weight estimation is calculated from the *I(0)*-value (the scattering intensity at zero angle) using known protein standard (BSA) as comparison and relies on the accurate protein concentration value. Additionally, the scattering of the NDSB caused problems in the background substraction and subsequently resulted in poor fit between the calculated and

experimental scattering curve at higher angles. Although this did not affect the modeling as the information for the overall shape of the molecule is derived from the scattering at lower angles, it reduced the overall quality of fit.

To overcome the negative effect of NDSB on SAXS measurements, other means to stabilize the protein were attempted. As a first attempt, SAXS measurements for the protein with the GB1-fusion tag attached was conducted. Prior to the measurements the sample was concentrated to 4.6 mg/ml and filtered to remove any aggregated protein that would influence the data quality. The concentration was measured using absorbance at 280 nm and the calculated extinction coefficient of 0.502. A dilution series of the protein sample was measured where the lowest concentration was 0.46 mg/ml. Although the SAXS data collected at higher sample concentrations showed signs of protein aggregation, the quality of the data was not affected. Furthermore, there was no buffer mismatch at the higher angles of the SAXS scattering curve as observed in the presence of the NDSB. The obtained data was used to calculate the SAXS parameters, which are shown in table 26. The radius of gyration, R_g and the scattering intensity at zero angle, *I(0)* were calculated using Guinier approximation while the maximum particle dimension, D_{max} was obtained from the interatomic distance distribution function, *P(r)*. The interpolated forward scattering from the BSA standard was used as a comparison to obtain an estimated molecular mass (MM_{SAXS}) for the measured ETR1-ΔTM construct. Another molecular mass estimation was obtained using the Porod equation, which gives the excluded volume of the hydrated particle, V_p (Porod, 1982). The molecular mass of the particle can be derived from V_p-value in $Å^3$, as it is approximately 1.7 times the MM in Da (Petoukov *et al*., 2012).

The *I(0)* and R_g showed a concentration-dependent behaviour by increasing slightly upon the increase in the sample concentration, which indicated a concentration-

dependent oligomerization. Also the MM_{SAXS}-value appeared to depend on the concentration to some extend. At concentrations below 3.0 mg/ml, it varied between 61 and 85 kDa, which is close to the MM of the monomeric ETR1-ΔTM with the GB1-fusion tag (the calculated molecular weight: 67.7 kDa). At the highest used protein concentration (4.2 mg/ml), the MM_{SAXS}-value was slightly higher (~100 kDa) than expected for a monomer. This indicated the existence of monomer-dimer equilibrium. However, the molecular mass estimates from the Porod volumes seemed to vary between 244 kDa (~4 times the MM of the monomer) and 417 kDa (~6 times the MM of the monomer), which would correlate with higher order oligomers. Possible reason for the observed difference between the MM estimations from independent methods could be inaccuracies in the concentration measurement. This affects especially the determination of the MM_{SAXS}-value. Additionally, the flexibility of the system causes problems in the calculation of V_p and thus provides an additional source of error for the estimation of the MM. Attempts to perform structural modeling were hampered mainly due to the increased degrees of freedom, which resulted from the additional rigid body, namely the GB1-fusion tag. Therefore it was decided to find purification conditions that would allow SAXS data collection in the absence of both NDSB and the fusion tag.

Table 26. The SAXS parameters for the ETR1-ΔTM with the GB1-fusion tag.

c (mg/ml)	R_g (Å)	I(0)	D_{max} (Å)	MM_{SAXS}	V_p (Å^3)
0.5	49.0	74.08	171	71	415,000
0.8	53.0	88.94	185	85	426,000
0.9	51.9	82.55	181	79	295,000
1.1	52.1	63.43	182	61	323,000
2.1	58.2	78.47	199	75	523,000
2.1	61.1	84.53	209	81	401,000
3.0	59.6	73.52	204	70	359,000
4.2	70.0	104.90	246	101	710,000
4.2	64.6	97.89	226	94	650,000
4.2	70.2	107.80	234	103	401,000

As described in the previous chapter, addition of 3 % glycerol was found to have a stabilizing effect on the studied ETR1-ΔTM construct. For the SAXS data collection several protein concentrations up to 3.0 mg/ml were tested. However, the final data was obtained using protein samples at concentration between 0.5 mg/ml – 1.4 mg/ml as at higher sample concentrations the protein aggregation created problems regarding data analysis. The data collection statistics together with the SAXS parameters derived from the scattering curve are summarized in the table 27.

For the structural modeling a previously constructed homology model of ETR1-ΔTM (Mayerhofer, 2011) was used. The model was built using the DHp domain structure of ERS1, the structures of the CA and receiver domain of ETR1 (PDB code 1DCF, Müller-Dieckmann *et al*., 1999) and a bacterial GAF domain from *Deinococcus radiodurans* (PDB code 2O9B, Wagner *et al*., 2007). The residues of the DHp domain and the GAF domain were mutated to the corresponding ones in ETR1 sequence. The position of the GAF domain in the model was based on the overlapping region with the N-terminus of the DHp. The relative orientation of the DHp and CA domains to one another was defined using the structure of the histidine kinase domain from *Thermotoga maritima* sensor histidine kinase protein TM0853 as a template. The short but flexible linkers between the GAF and DHp domains, and the DHp and CA domains allowed some movement in the otherwise rigid core structure. The receiver domain was attached to this core with a 15-residue linker.

The ridig body modeling was performed using the program CORAL (Petoukhov & Svergun, 2005). The known dimeric organization of the DHp domains was defined as a core around which the GAF, CA and receiver domains were modeled. The linker regions between the domains were treated as flexible random polypeptide chains. The calculated

scattering curve from the obtained model correlated well with the experimental data based on the χ^2 value of 1.2. In comparison to the results from the SAXS measurements in the presence of NDSB, a better fit also at higher scattering angles were obtained for the new data in the absence of the NDSB compound (Figure 24). This didn't, however have a significant effect on the overall shape of the ETR1-ΔTM SAXS model as the information about the size and the shape of the particle is mainly derived from the lower angle scattering range. The model of ETR1-ΔTM exhibited the previously observed dimeric arrangement where the DHp domains form the central four-helix bundle with the GAF domains attached to top and the CA and the receiver domains to the bottom part of this helical bundle (Figure 25A). However, in the absence of the NDSB compound, ETR1-ΔTM appeared to be more flexible based on the larger D_{max} –value (190 Å) in comparison to the D_{max}-value (158 Å) of the model from the SAXS data in the presence of NDSB. The main source of flexibility appeared to be the 15-residue linker between the CA and the receiver domain, which allowed the movement of the latter domain in relation to the rest of the molecule.

As another modeling approach, the *ab initio* bead-model of the protein was constructed using program DAMMIN (Svergun, 1999). The program uses simulated annealing to create a shape envelope for the target particle that fits to the experimental data. Since the particle is represented as finite volume elements (beads/dummy atoms), contraints for compactness and connectivity are required in order to create a physically and biologically reasonable model. For ETR1-ΔTM additional P2 symmetry constrain was applied based on the observed symmetry-relation between the DHp domains in the crystal structure. The obtained *ab initio* model of ETR1-ΔTM agreed with the dimeric rigid body model from CORAL as shown with the overlay of the two models in figure 25B.

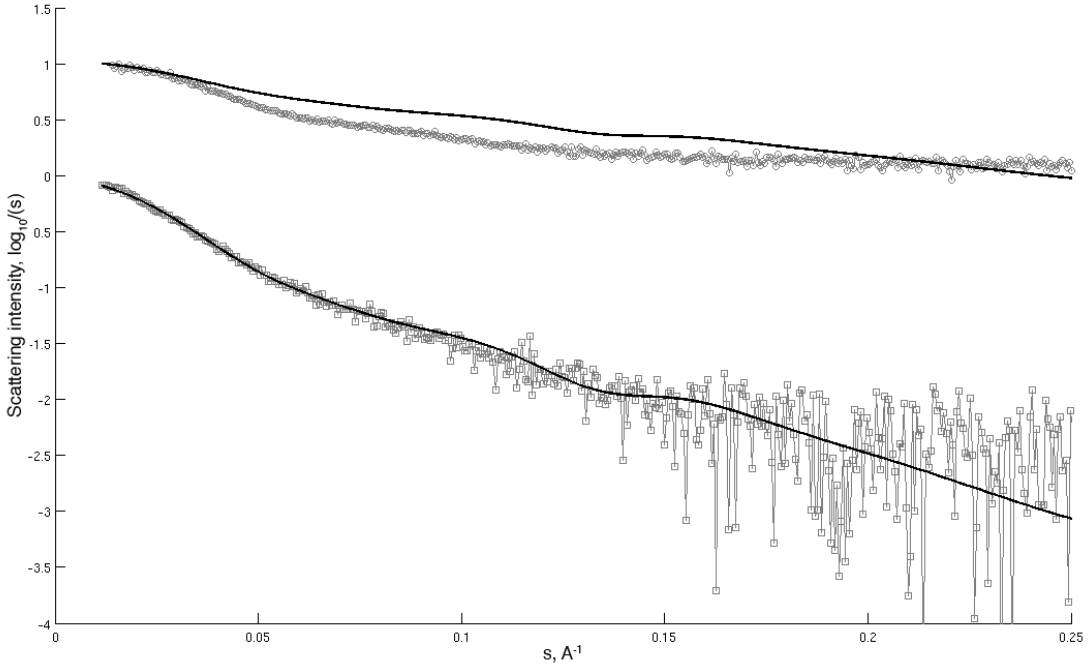

Figure 24. Structural modeling of ETR1-ΔTM against the SAXS data. The fit of the CORAL model with the experimental data is shown for the data both with NDSB (above, circles) and without NDSB (below, squares). The collection and intepretation of the SAXS data in the presence of NDSB has been previously described (Mayerhofer, 2011) and is shown here for comparison. The overall fit of the calculated scattering curve to the experimental one was better in the absence of the NDSB due to the lack of buffer mismatch at the higher angles.

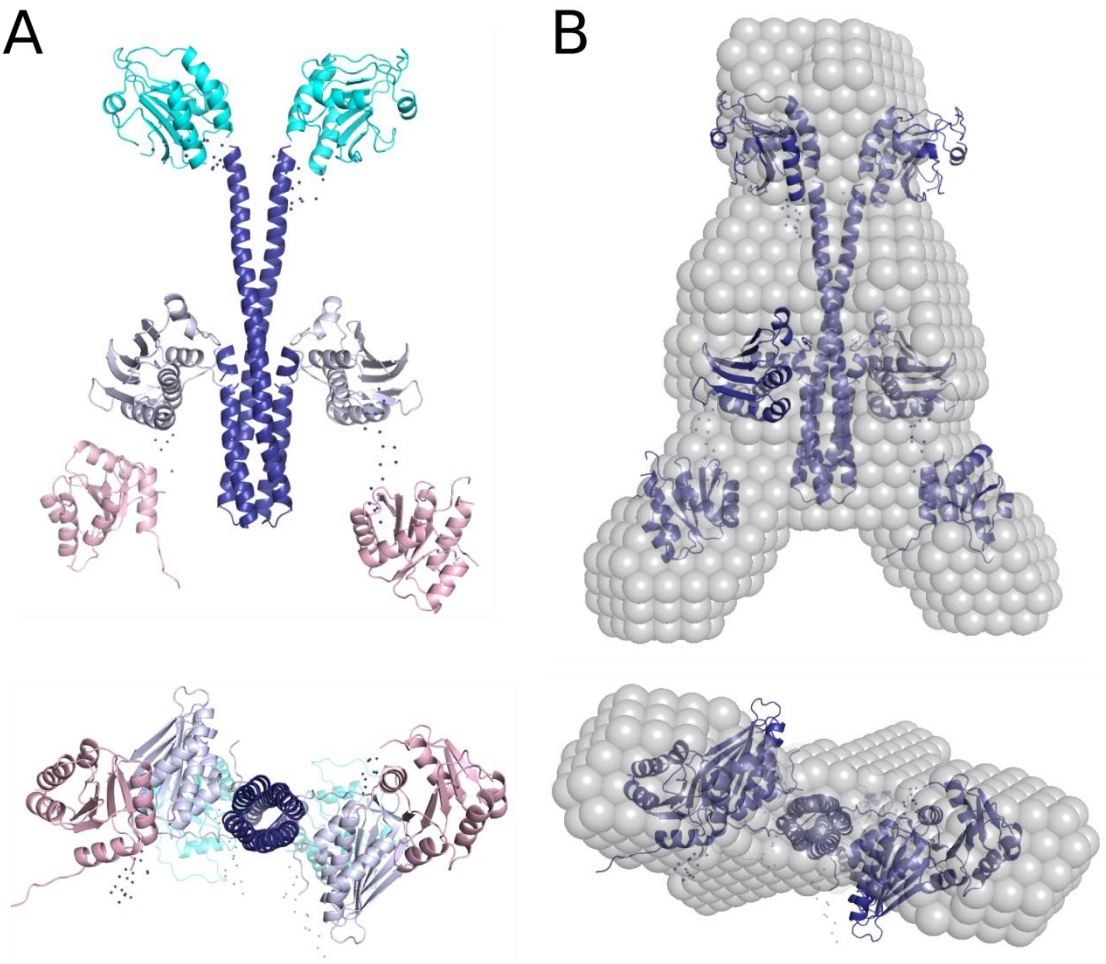

Figure 25. The SAXS models of ETR1-ΔTM in the absence of NDSB. The *ab initio* model from DAMMIN and rigid body model from CORAL for ETR1-ΔTM. A) The ridig body refinement with CORAL resulted in the expected dumbbell shape. B) The shape and size of the *ab initio* bead model (in grey) correlated well with the rigid body model (dark blue cartoon representation) as shown with the alignment.

As the receiver domain in the ETR1-ΔTM model was not very tightly packed against the dimer core, it was expected to exhibit more flexible movement. To analyze the flexibility mediated by the 15-residue linker between the CA and receiver domains, recently developed ensemble optimization method (EOM, Bernado *et al*., 2007) was employed. The program *RANCH* (Bernado *et al*., 2007) was used to produce 10 000 random configurations by allowing the movement of the receiver domains in relation to the rest of the protein. From the created pool of possible conformations, an ensemble of

conformations that provided the best fit to the scattering curve was selected using a generic algorithm. The results were represented as a distribution of R_g –values. The original pool covered a broader R_g range whereas the best-fitted models had an average R_g of about 54 Å, which correlated well with the R_g of about 56 Å calculated for the CORAL model. Additionally, a more compact protein species were observed with an R_g of about 51 Å. This indicated movement between extended and compact conformations mediated by the receiver domain. The results from the EOM analysis are summarized in the figure 26.

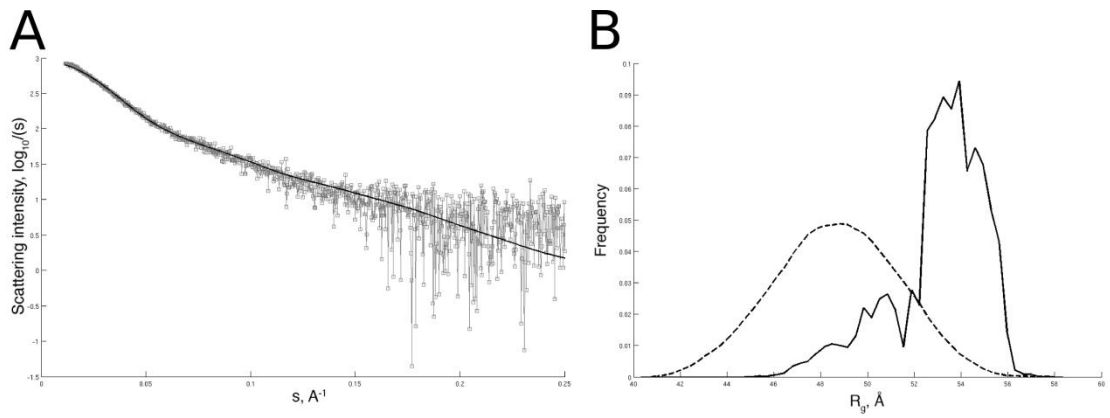

Figure 26. The results from the flexibility analysis of ETR1-ΔTM using EOM approach. A) The fit of the selected ensemble of conformations from EOM analysis to the scattering data. The fitted curve is shown in black line and the scattering curve in grey squares. B) The R_g distribution in the ensemble of conformations fitting the experimental data using EOM. The R_g of the models in the original pool of structures has a broad distribution (shown with dotted line) while the ensemble of the selected structures (shown with solid line) has an average peak at ~ 54 Å and a smaller side peak at ~ 51 Å.

Table 27. SAXS Data Collection and Scattering Derived Parameters.

Data collection parameters	
Instrument	P12, PETRAIII (EMBL, Hamburg, Germany)
Beam geometry (mm)	0.2 x 0.12
Wavelength (Å)	1.24
s-range ($Å^{-1}$)	0.008 – 0.47
Exposure time (ms)	50
Concentration range (mg ml^{-1})	0.66 – 4.16
Temperature (K)	283
Structural parameters*	
I(0) (arbitrary units) (from P(r))	1175 ±10
R_g (Å) (from P(r))	59 ± 6
I(0) (arbitrary units) (from Guinier)	1164 ± 40
R_g (Å) (from Guinier)	55 ± 11
D_{max} (Å)	190 ± 5
Porod volume estimate ($Å^3$)	274000 ± 4200
Dry volume calculated from sequence ($Å^3$)	78367
Molecular mass determination*	
Contrast ($\Delta\rho$ x 1010 cm^{-2})	3.047
Molecular mass M_r (from I(0))	53000 ± 5000
Molecular mass M_r (from Porod volume)	161200 ± 24000
Calculated monomeric M_r from sequence	64771.3
Software employed	
Primary data reduction	Automated radial averaging (Petoukhov *et al*., 2007)
Data processing	PRIMUS
Ab initio analysis	DAMMIN
Validation and averaging	DAMAVER
Rigid body modelling	CORAL, EOM
Computation of model intensities	CRYSOL
3D graphics representations	PYMOL

*Reported for 0.66 mg ml^{-1} measurement

4.3 EDR1

4.3.1 Production of the protein material

The plasmid for the wild type kinase domain of Enhanced Disease Resistance 1 (EDR1) was received from S. Panneerselvam. This plasmid comprised the C-terminal

residues 655-933 of EDR1 cloned into the expression vector pMCSG7 (Eschenfeldt *et al*., 2009). The target gene had been amplified by PCR from a cDNA library obtained from the Arabidopsis Biological Research Center at Ohio State University (Kieber *et al*., 1993). The expression vector introduced an N-terminal His_6-tag followed by a TEV cleavage site to the target protein. Due to cloning, 3 additional amino acids (SNA) remained at the native N-terminus after digestion with TEV protease. The correct insertion of the target gene was verified by DNA sequencing.

The wild type EDR1 kinase domain containing plasmid was used as template to produce four variations of the enzyme using site-directed mutagenesis. These variants comprised the catalytically inactive enzyme and three interface mutants targeting on the conserved G-helix in the enzyme. In order to produce a catalytically inactive kinase domain, a conserved and catalytically important aspartate residue (D792) was substituted with asparagine. The three interface mutants were made based on the solved crystal structure of the inactive EDR1 kinase domain, which implicated that the residues in the conserved G-helix motif are participating in protein-protein interactions at the ES-complex interface. The two substituted residues were valine (V876) and phenylalanine (F881) of the G-helix, both of which were mutated to glutamate. The third interface mutant was the combination of the two, V876E/F881E. The sequence of the four plasmids was verified by DNA sequencing.

The EDR1 kinase domain constructs were expressed in *E. coli* strain BL21 cells co-expressing chaperones DnaK, DnaJ, GrpE, ClpB, GroEL and GroES (CC4 cell lines, courtesy of A. Geerlof, Helmholtz Zentrum München) as described in methods section. The isolation of the protein from the cell lysate was conducted using Ni-NTA affinity chromatography followed by TEV protease cleavage to remove the His-tag and finally size exclusion chromatography (SEC). For the first purification step the cell lysate was

loaded onto Ni-NTA Agarose beads (Invitrogen™, Life Technologies GmbH, Darmstadt, Germany) pre-equilibrated with buffer A (see materials and methods) and washed in two steps with 30 mM and 50 mM imidazole containing buffers, respectively. The protein was eluted in four consecutive steps (elutions 1 to 4) with increasing imidazole concentration, first with 75 mM, then with 100 mM and 300 mM, and finally with 500 mM imidazole using 2CV of buffer in each step. The NiNTA purification fractions were analyzed using SDS-PAGE (Figure 27). Small fraction of the protein material were lost in the 2nd wash step but most were found from the four elutions that were therefore combined for the subsequent purification steps.

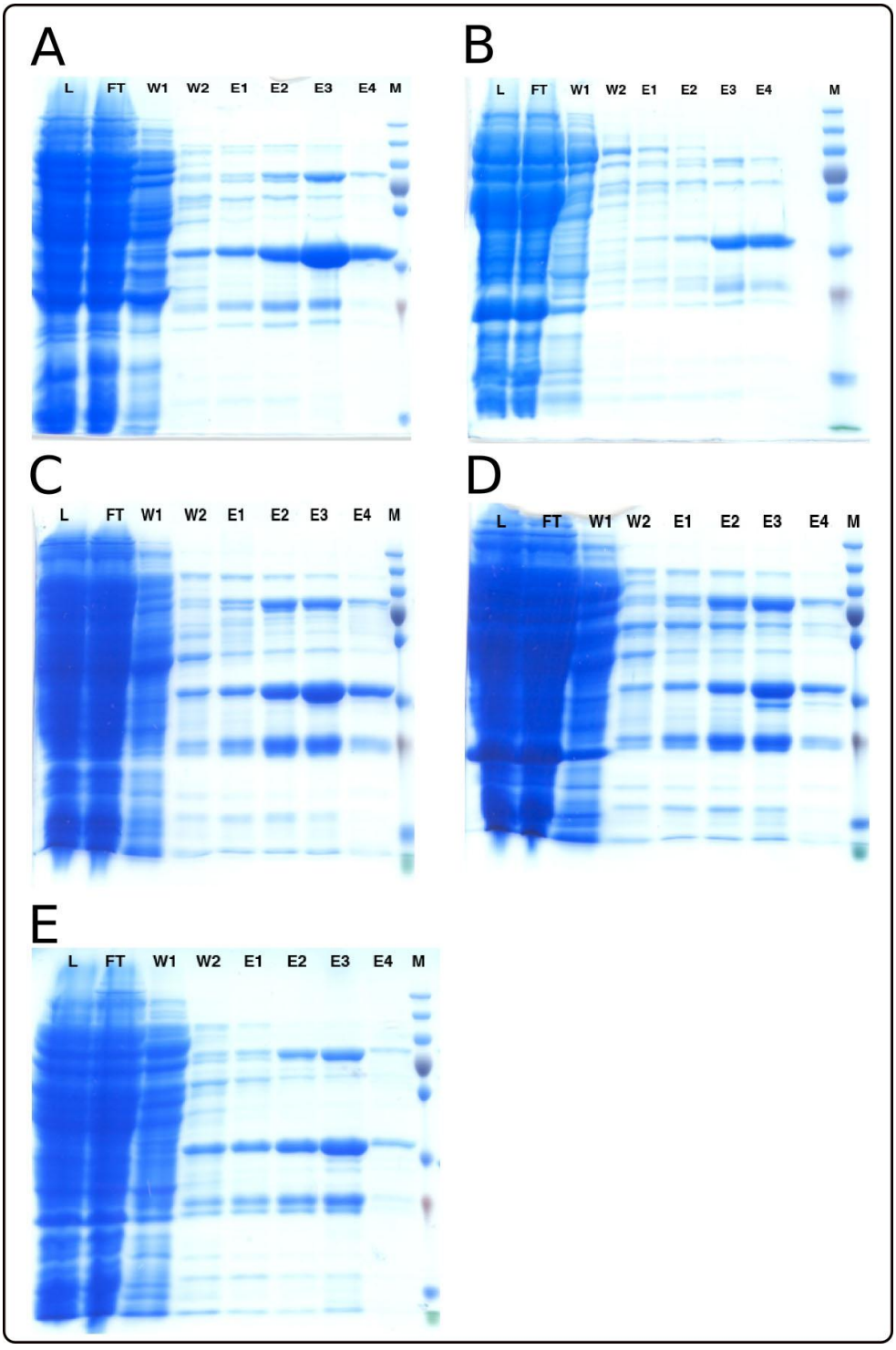

Figure 27. The SDS-PAGE analysis of the NiNTA fractions for the EDR1 kinase domain constructs. The abbreviations used: L = lysate, FT = Flow through, W1 = 1st wash, W2 = 2nd wash, E1 – E4 = The four elution steps with increasing imidazole concentration, M = PageRuler Color Plus Prestained protein ladder (Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany). The results for WT are shown in A, for D792N mutant in B, for F881E mutant in C, for V876E in D and for the V876E/F881E double mutant in E.

The NiNTA purification was followed by the removal of the His_6-tag using TEV protease treatment and subsequent removal of the protease using an additional NiNTA step. Size exclusion chromatography (SEC) was performed as the last purification step for the cleaved protein to yield a pure protein.

The SEC elution profile and the corresponding SDS-PAGE analysis of the wild type EDR1 kinase domain is shown in figure 28. Two distinct peaks were observed in the SEC graph. Based on the SDS-PAGE analysis, the material that eluted nearly at the void volume of the column (V_0 = 42 ml, HiLoad 16/60 Superdex 75 prep grade, GE Healthcare Bio-Sciences AB, Uppsala, Sweden) contained mainly aggregated target protein but also protein contaminants with molecular masses larger than 70 kDa. The presumably monomeric protein eluted separately with an elution volume of about 70 ml. The SDS-PAGE analysis of the monomeric fraction revealed that the obtained protein material was not homogenous. Instead, two protein bands near the 35 kDa molecular weight standard band were observed. The molecular mass of both observed protein species corresponded roughly to the calculated molecular mass of EDR1 kinase domain (32 kDa). Phosphorylation was considered as possible explanation for the observed heterogeneity as the mass spectrometry analysis of the protein indicated existence of multiple phosphorylation sites and multiple protein species (results shown in detail in chapter 4.3.2.2). To obtain more homogeneous sample, a kinase inactive mutant was produced.

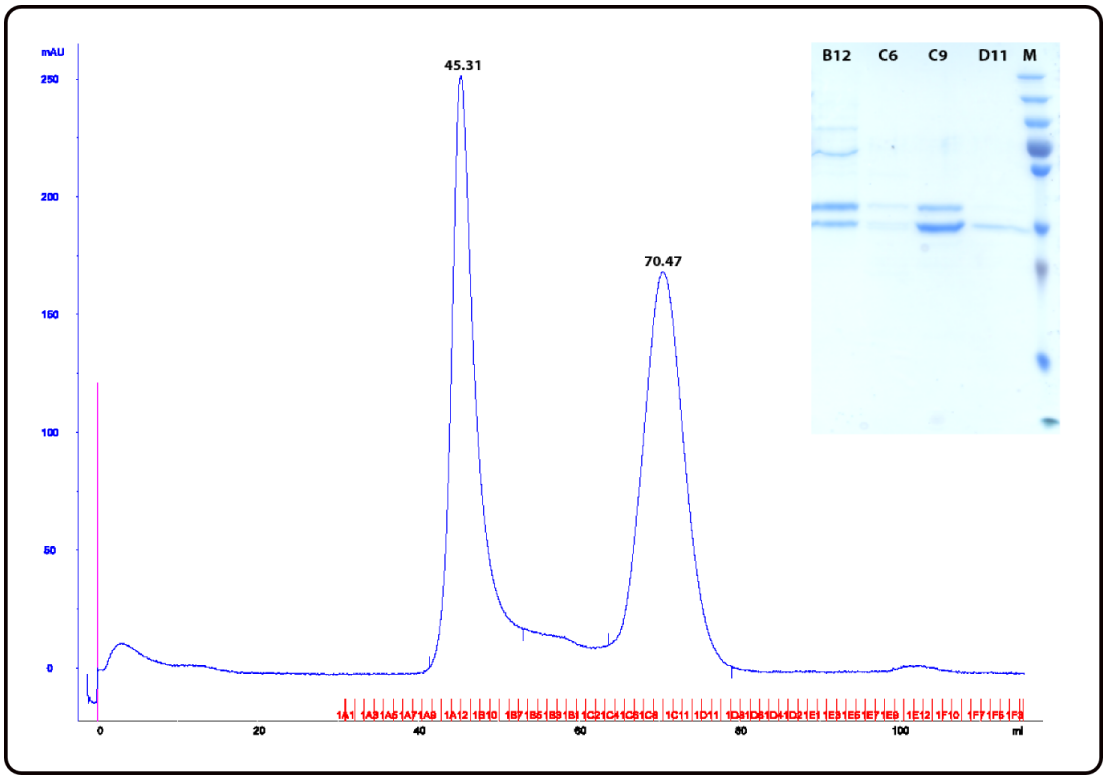

Figure 28. The SEC curve and the corresponding SDS-PAGE analysis of WT EDR1 kinase domain. The lanes in the gel are designated by the corresponding elution fraction from the SEC purification. The first peak in the SEC elution profile corresponded to aggregated target protein and higher molecular weight impurities according to SDS-PAGE analysis shown here for the elution fraction B12. The elution fraction of the presumed monomeric protein (fractions C6, C9 and D11 on gel) appeared to be heterogeneous, presumably due to phosphorylation. PageRuler Color Plus Prestained protein ladder (Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany) was used as protein marker (lane M).

The catalytically inactive variant of EDR1 kinase domain seemed to have a slightly altered SEC profile when compared to the wild type (Figure 28). Altough the SEC profile of EDR1-D792N contained two peaks, significantly less protein came out in the aggregate fraction. Additionally, the SDS-PAGE analysis confirmed that elution fractions with the target protein (elution volume of about 68 ml) contained only one protein species with molecular weight slightly below 35 kDa. This correlates well with the calculated molecular mass of EDR1-D792N, which is 32 kDa. The fractions that

contained the pure monomeric protein were pooled together and concentrated to about 10 mg/ml prior to use.

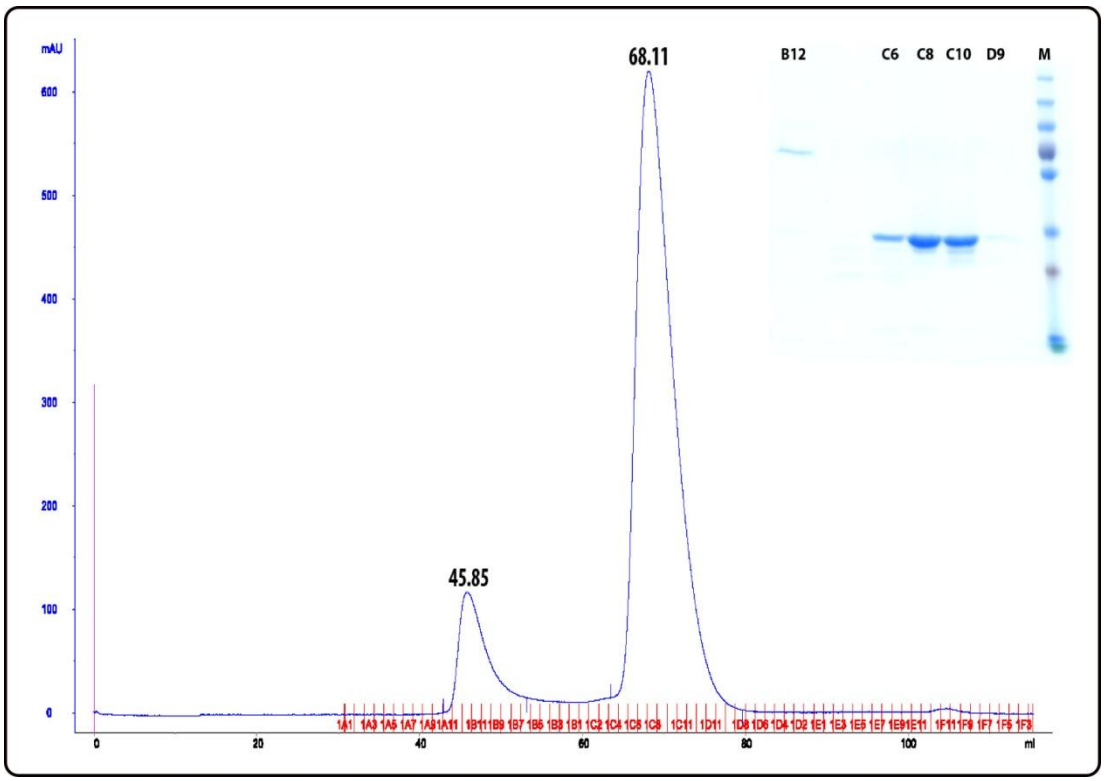

Figure 29. The SEC graph and corresponding SDS-PAGE analysis for EDR1-D792N. The fractions corresponding to the second peak with elution volume of about 68 ml in the SEC profile, containing the pure, monomeric protein (MW = 32 kDa) were pooled together. M = PageRuler Color Plus Prestained protein ladder (Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany).

The SEC results of the EDR1 kinase domain G-helix mutants were very similar to the EDR1-D792N, also based on the SDS-PAGE analysis (Figure 29). The final protein sample consisted of monomeric protein and no additional protein species were observed according to SDS-PAGE. The SEC purification step of the three G-helix mutants resulted in pure protein samples, which were concentrated to roughly 5 mg/ml prior to use in the activity assays.

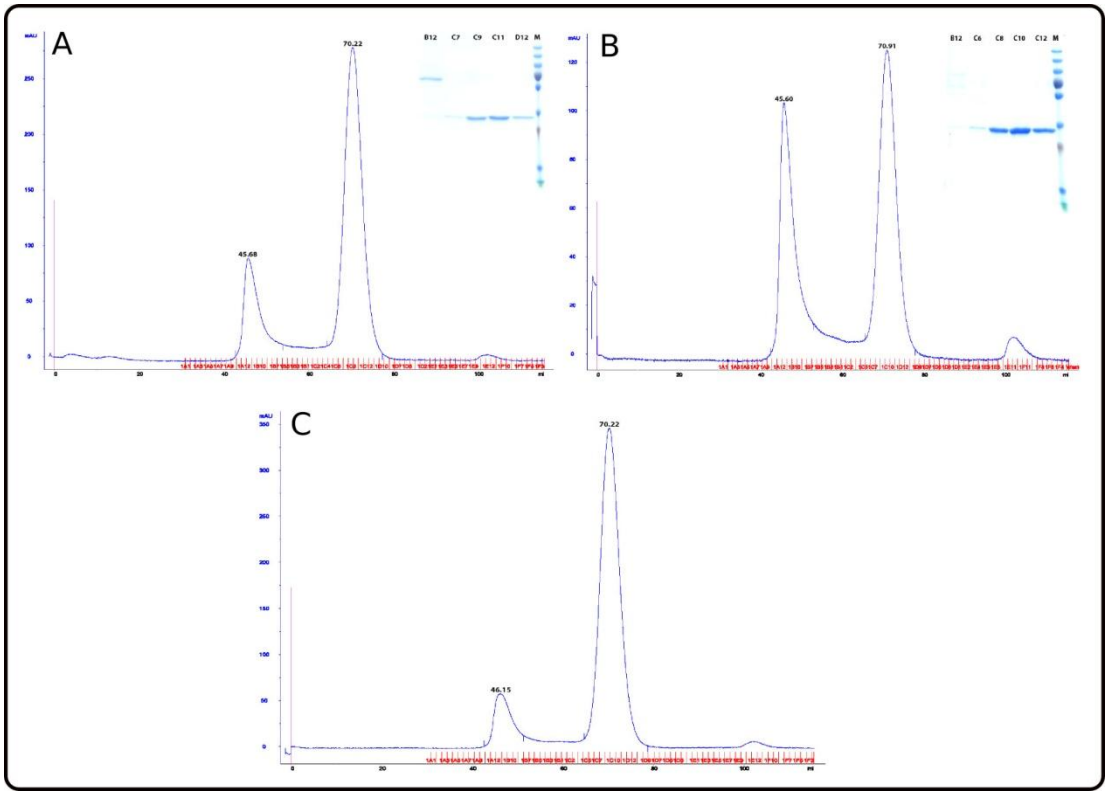

Figure 30. The SEC purification profiles and SDS-PAGE analysis results of the three EDR1 kinase domain G-helix mutants. A) The SEC profile of the EDR1-F881E mutant showed two peaks at elution volumes of about 46 ml and 70 ml, latter one of which contained the pure monomeric protein according to SDS-PAGE analysis. B) Based on the SEC graph of EDR1-V876E, the sample contained more aggregated or larger oligomer of the protein compared to the other two G-helix mutants. Despite of this, highly pure protein was obtained. C) The G-helix double mutant had a similar SEC profile with the D792N and F881E mutants. The fractions of the pure monomeric protein, eluting out at volume of about 70 ml, were pooled together.

4.3.2 Biophysical characterization of EDR1 kinase domain

4.3.2.1 Thermofluor for the EDR1-D792N

The D792N mutant was despite of being homogenous compared to the wild type kinase domain, not at the beginning very amenable for crystallization. To find either more suitable buffer condition or optionally an additive that would stabilize the protein, thermofluor experiment using both buffer and additive screen (Boivin *et al.* 2013) was

performed. The protein concentration used in both screens was 100 µM, which is the upper limit of recommendable protein concentration in thermofluor experiment. From the buffer screen significantly improving conditions were not identified. The results for buffer screen are summarized in figure 31.

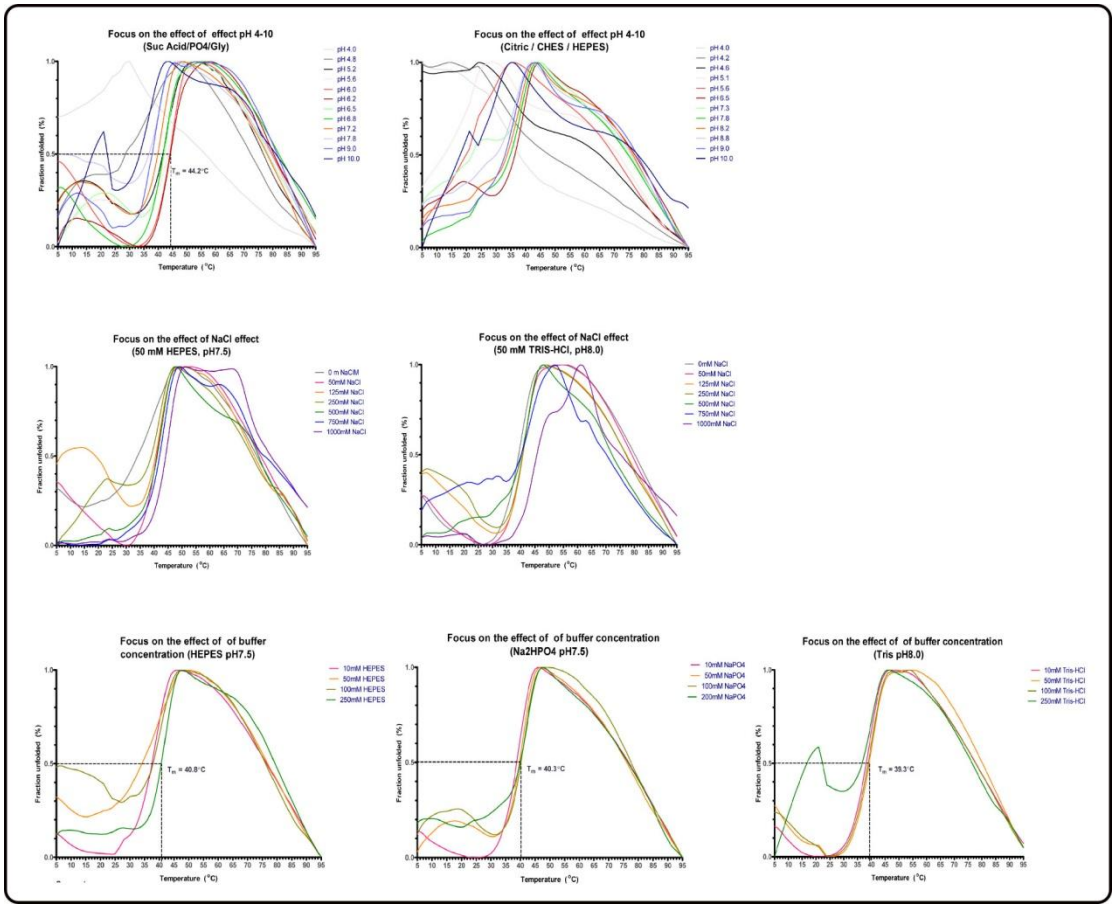

Figure 31. The results of the thermofluor experiment for EDR1-D792N using the buffer screen. No significant stabilizing effect on the protein with the standard buffer components was observed. Not surprisingly, very acidic or basic pH however showed strong de-stabilizing effect.

In search of additives that could have positive effect on the thermostability of the EDR1-D792N, the protein was subjected to the thermofluor analysis with additive screen. This screen included compounds with various different chemical properties, such as chaotropic and dissociating reagents, salts, mono- and multivalent ions, reducing agents,

polyamines, detergents, alcohols, amino acids, sugars and nucleotide cofactors. Additionally three custom chemicals were included, which were the generic kinase inhibitor staurosporin in concentration 0.5 mM and the non-hydrolyzable analog of ATP, AMP-PNP in two different concentrations, 5 mM and 10 mM. Staurosporin was chosen since it was successfully used as additive to crystallize CTR1 (Mayerhofer *et al*., 2012) while AMP-PNP mimics the natural ATP substrate of EDR1.

A modest stabilizing effect was seen with various compounds from different chemical subgroups. Out of the tested salts, both sodium and potassium dihydrogen phosphates had a clear positive effect on the thermostability of the protein. Both increased the melting temperature by 7°C. Similarly a slight increase in the melting temperature was observed with all the tested polyamines (100mM Trimethylamine, 1mM Spermidine and 1 mM Spermidine-HCl) and with glycerol when used in 10 or 20 % concentration. The increase was in this case in 5 to 6°C range. The additives with only a very weak stabilizing effect on the EDR1-D792N included NH_4Cl, NaBr, the detergent Octyl-β-D-Glucoside, taurine and amino acids, such as L-Histidine and in high concentration (500mM) L-arginine when combined with L-Glutamate. These chemicals shifted the T_m-value by 3°C on average.

The most remarkable increase in the melting temperature and was observed when either ATP or its non-hydrolyzable analog, AMP-PNP was added to the protein solution. In the case of ATP an increase from 37°C to 57.3°C was measured whereas with AMP-PNP only slightly lower increase, namely from 37°C to 54.3°C was seen. The effect of AMP-PNP appeared to be independent of the concentration since the increase from 5 mM to 10 mM concentration didn't show any shift in the measured T_m-value. Also the addition of generic kinase inhibitor, staurosporin had a noticeable positive effect on the thermostability of the protein, increasing the melting temperature from 37°C to 50.1°C.

Based on these results, the presence of ATP, AMP-PNP and staurosporin was considered to have desired effect on making EDR1-D792N more amenable for crystallization. The results for the chemical included by default in the thermofluor additive screen are shown in figure 32. The results for the customized screening of thermostability improving compounds are shown in figure 33.

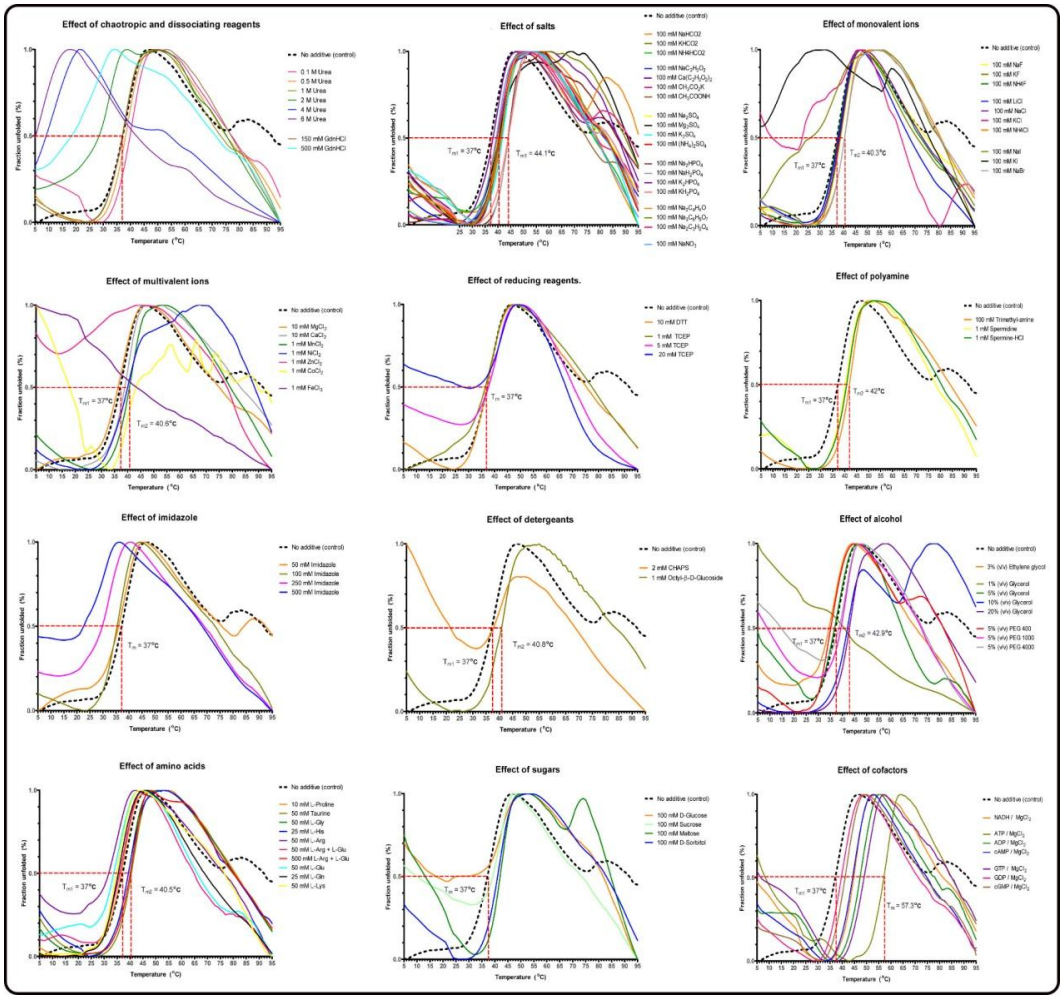

Figure 32. The effect of several different additives with various chemical properties was tested for the EDR1-D792N. From the chemical included in the thermofluor additive screen, there were several that showed a modest stabilizing effect on the protein. These were some salts, such as dihydrogen phosphates, Mn^{2+} of the multivalent ions, all three polyamines included in the screen, the detergent Octyl-β-D-Glucoside, glycerol in concentration 10 or 20 % and some amino acids, such as L-Histidine and combination of L-Arginine and L-Glutamate in concentration 50 mM. The strongest effect on the thermostability of the EDR1-D792N was observed with nucleotide cofactors, especially with the natural substrate molecule ATP, which increased the melting temperature of the protein with 20°C.

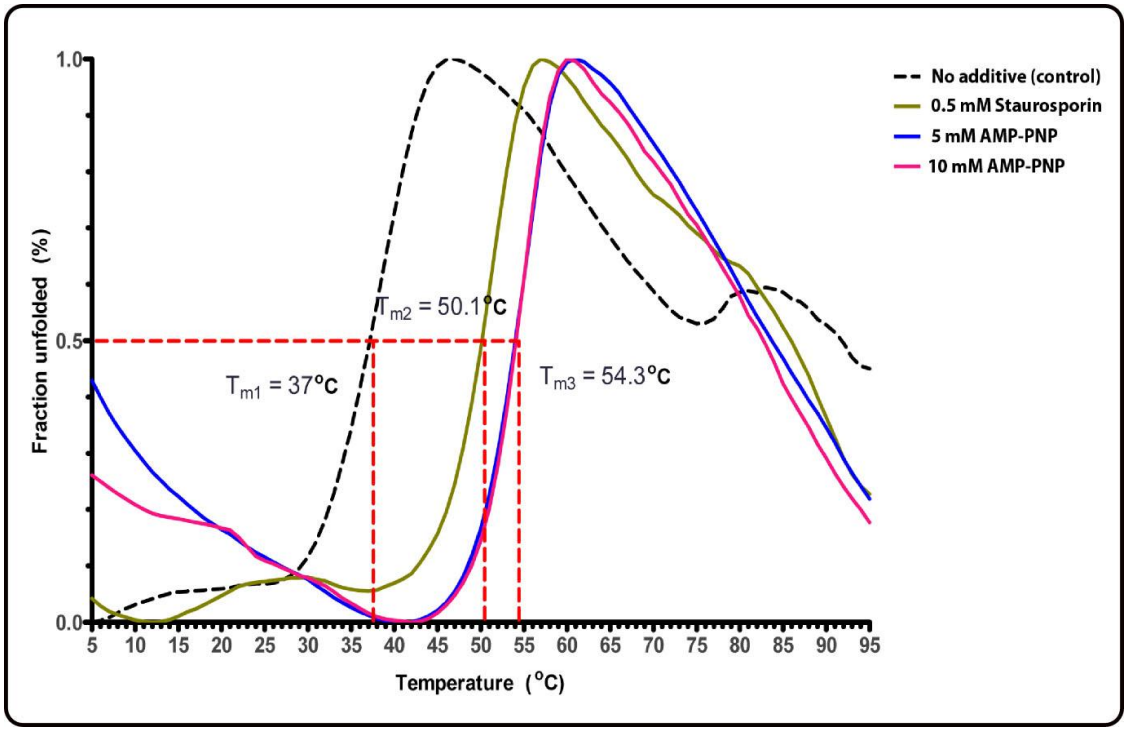

Figure 33. The effect of staurosporin and AMP-PNP on the thermostability of EDR1-D792N. Both stabilized the protein significantly based on the roughly 20°C increase in melting temperature.

4.3.2.2 Phosphorylation sites

The crystallization of the wild type EDR1 kinase domain proved to be a highly challenging task presumably due to the unability to produce material that was homogeneous enough. This was accounted to be the result of multiple phosphorylation sites in the protein. Based on a *NetPhos 2.0 Server* prediction, EDR1 kinase domain was estimated to contain 9 phosphorylation targets, out of which 6 were serine and 3 threonine residues (Figure 34). This program uses artificial neural network method to score the individual putative phosphorylation sites and the residues with highest score are considered to be the most probable sites for phosphorylation (Blom *et al*., 1999). To get a more comprehensive view of the possible phosphorylation targets within EDR1, peptide-fingeprinting analysis with LC-MS/MS was performed to identify phosphorylation

containing peptide fragments from both trypsin and chymotrypsin containing samples. The results from the MS/MS analysis were analyzed using *MASCOT version 2.2.03* (Matrix Science Limited, London, UK). Also the peptide-fingerprinting analysis indicated that the enzyme contained 9 putative phosphorylation sites, although only the phosphorylation of Thr787 could really be confirmed with certainty. This threonine located two residues upstream from the catalytic HRD motif. It was also observed that majority of the pshosphorylation sites concentrated on the activation segment region, which is marked by DFG and APE tripeptide sequences at the N- and C-terminus of this region, respectively. The position of the putative phosphorylation sitesin the EDR1 kinase domain sequence is shown in the figure 34.

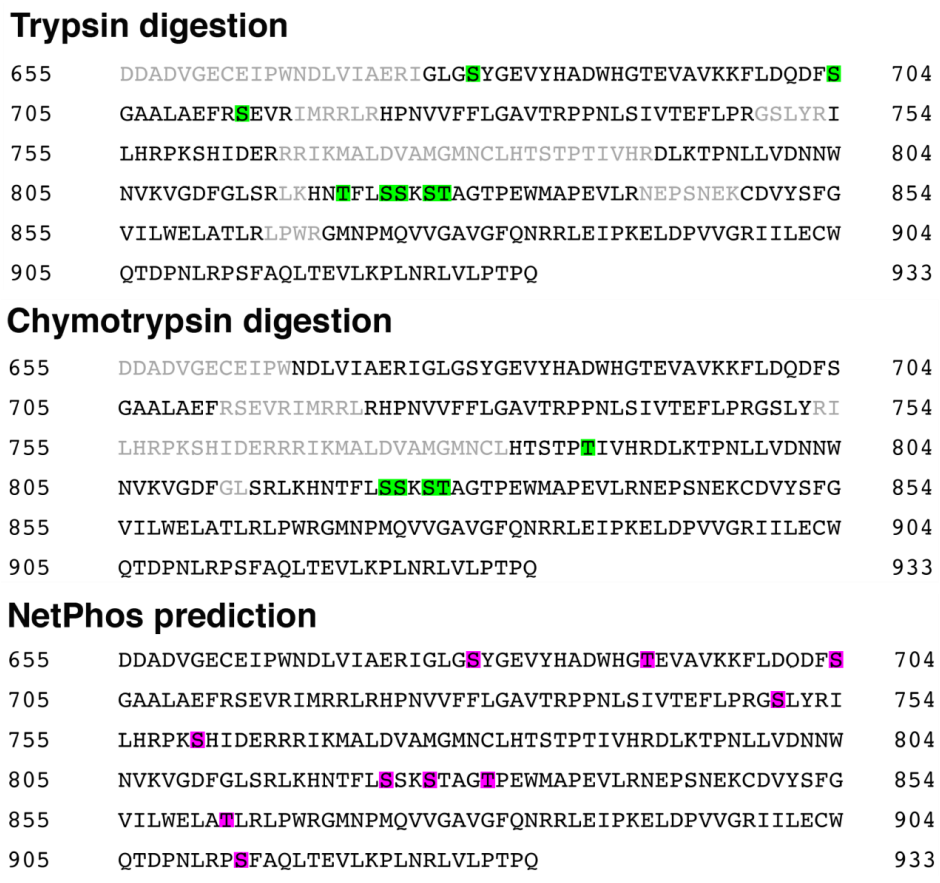

Figure 34. The putative phosphorylation sites of EDR1 kinase domain. The results from the peptide-fingerprinting analysis using trypsin and chymotrypsin digested fragments are shown in the upper and middle panel while the predicted sites from NetPhos are shown in the lower panel.

To examine the number and mass of the different protein species in the final purification fraction, ESI-MS analysis was conducted. As expected the sample for the wild type EDR1 kinase domain appeared to be heterogeneous based on the separate mass species in the ESI-MS spectrum. Three separate protein entities were observed with molecular weights of 32248 Da, 32328 Da and 32408 Da (Figure 35). These correlate very well with the 3-, 4-, and 5-fold phosphorylation of the protein since the calculated molecular weight for the unphosphorylated protein is 32008 Da to which each phosphoryl group adds an additional 80 Da. Furthermore, when the same MS analysis was performed to the kinase inactive mutant of EDR1, the most abundant mass entity had the molecular weight of the unphosphorylated EDR1 kinase domain (Figure 36). These results confirmed that the wild type protein has multiple phosphorylations and furthermore is not homologously phosphorylated. The substitution of the catalytically important aspartate (Asp792) with asparagine renders the enzyme inactive and results in unphosphorylated protein.

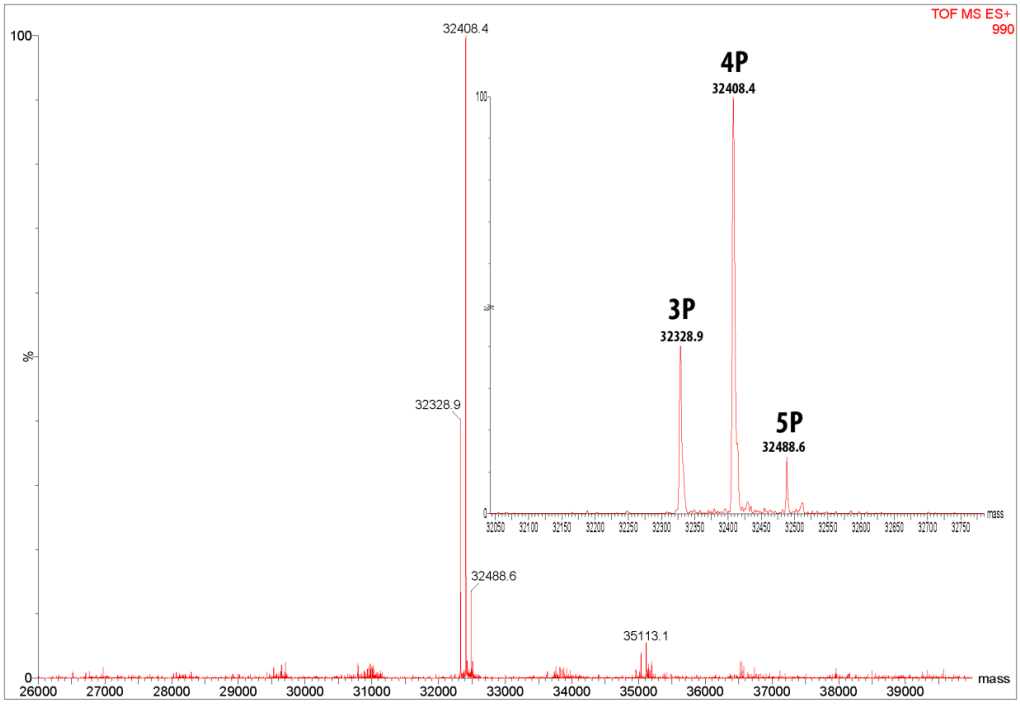

Figure 35. The ESI-MS analysis results for the wild type EDR1 kinase domain (courtesy of S. Leicht, Proteomics Core Facility of EMBL Heidelberg). As expected the sample was heterogeneous and contained multiple protein species in different phosphorylation states. The main forms of the protein appeared to possess 3, 4 or 5 phosphorylated residues based on the obtained mass values.

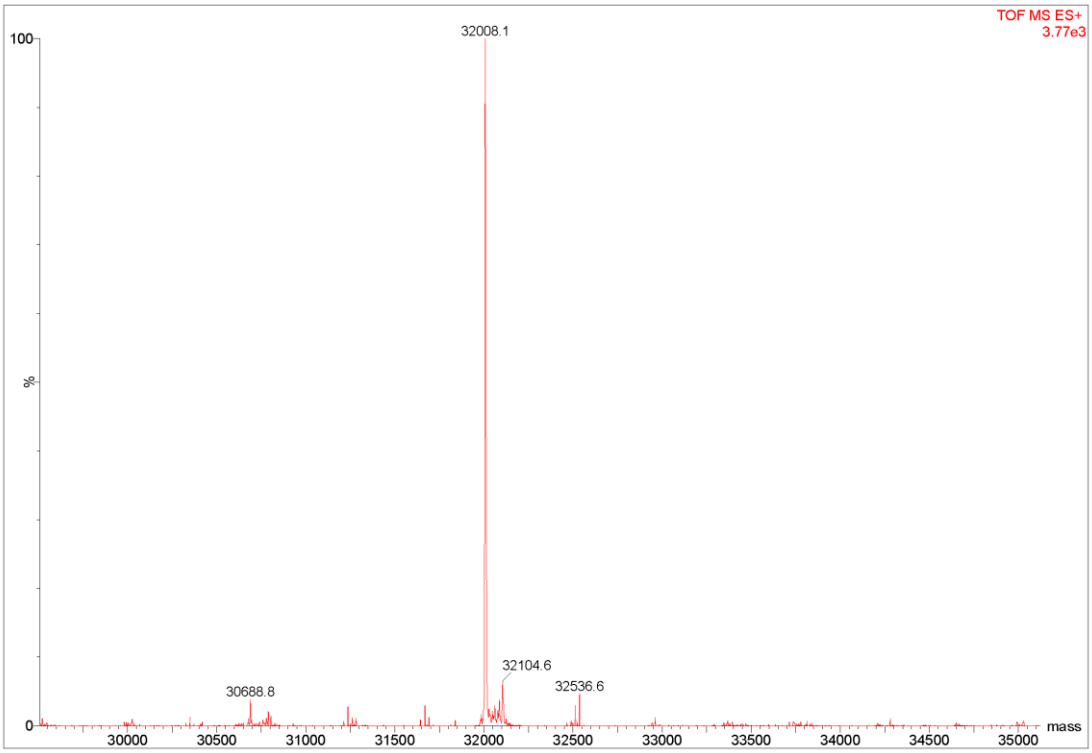

Figure 36. The high accuracy molecular mass analysis was performed for EDR1-D792N mutant using ESI-MS (courtesy of S. Leicht, Proteomics Core Facility of EMBL Heidelberg). The most abundant mass species corresponded to the non-phosphorylated form of the enzyme.

4.3.2.3 The kinase activity of EDR1-kd and EDR1-D792N

The coupled kinase assay was used to study the activity of the wild type EDR1 kinase domain and confirm that the inactive state of EDR1-D792N. In this assay the enzymatic activity of the target protein is the only source for ADP, which is required for the reaction catalyzed by puryvate kinase to produce ATP and Pyruvate from phosphoenolpuryvate (PEP). The formed puryvate in turn is reduced to lactate by LDH in expence of NADH oxidation. NADH has absorption maxima at 540 nm thus providing the means to measure its consumption. A dilution series of ATP (from 1 mM to 0.016 mM) was used and for each protein sample three replicates were measured simultaneously.

The linear part of the raw data curve (A_{540}-values versus time in minutes) for the different ATP concentrations was used to calculate the reaction velocity for each replicate. The reaction velocity (mOD/min) was obtained from the slope of the curve. The mean mOD/min value from the three replicates was then plotted agaist the ATP concentration (µM) and from the obtained curve (Figure 37), the standard enzyme kinetics parameters were calculated according to Michaelis-Menten kinetics. Only the wild type enzyme showed activity with a maximum reaction velocity (V_{max}) value of 6.7 ± 0.3 mOD/min and a Michaelis constant (K_m) of 60 ± 8.9 µM.

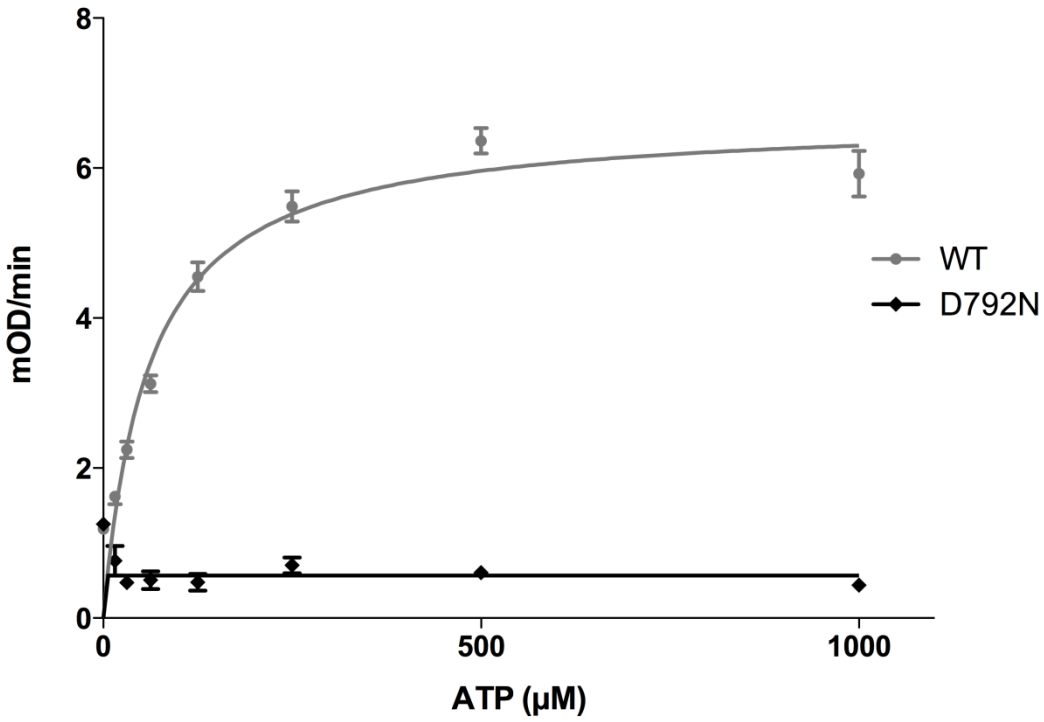

Figure 37. The kinetics curve from the coupled kinase assay for both the wild type kinase domain of EDR1 and for the D792N mutant. The mutation of the catalytically important aspartate residue (D792) renders the enzyme inactive. The error bars represent the standard deviation of the reaction velocity for the three replicates at each data point.

4.3.2.4 Oligomerization state of EDR1 WT and D792N by SEC-LS/UV/RI

The oligomerization state of wild type and D792N mutant of EDR1 kinase domain were studied using SEC-LS/UV/RI to investigate whether EDR1 shows similar dimerization dependent behaviour as CTR1. CTR1 was shown to form dimers when active and inactive mutants of the enzyme appeared to mainly exist as monomers (Mayerhofer *et al*., 2012). Since CTR1 is the closest homolog of EDR1, it was speculated whether the kinase activity of EDR1 would also be dimerization dependent. Both the wild type enzyme and the D792N variant were tested at two concentrations. The Malvern instrument was used at this work to measure the UV absorbance at 280 nm, refractive index (RI) and the right angle light scattering (RALS) of the sample, which eluted out from the SEC column. The measured values for reference BSA sample were used to obtain the calibration constants for the equipment. Based on the calibration constants from the BSA standard and the measured RI and LS signals for the studied protein, the molecular mass of the latter one was calculated.

The UV-absorption based elution profile of the wild type EDR1 kinase domain consisted of one main peak at approximate elution volume of 15.5 ml with both 4.5 and 7.3 mg/ml protein concentration (Figure 38 A and B). The calculated molecular mass for the corresponding protein species was 30.4 kDa at concentration 4.5 mg/ml and 36.9 kDa at concentration 7.3 mg/ml, both of which are roughly the molecular mass of a monomer. Similar results were obtained for the kinase inactive mutant at the tested concentrations of 1.7 mg/ml and 3.5 mg/ml (Figure 39 A and B). Therefore the analysis with SEC-LS/UV/RI indicated that the kinase domain of EDR1 exist mainly as a monomer in solution independent of its activation state.

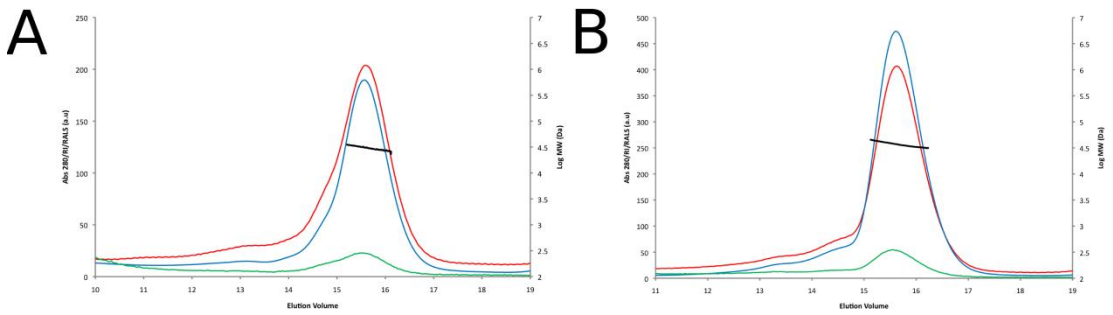

Figure 38. The results for the wild type EDR1 kinase domain from the SEC-LS/UV/RI analysis. The measured UV absorbance at 280 nm is represented with the blue curve, RI with red curve and RALS with green curve. The molecular weight of the main protein species corresponded to the monomeric form of the enzyme, both at concentration 4.5 mg/ml (A) and 7.3 mg/ml (B).

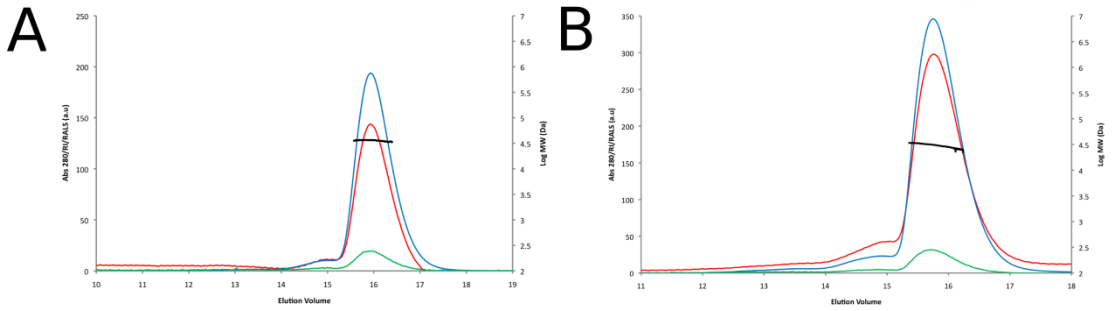

Figure 39. The SEC-LS/UV/RI analysis results for EDR1-D792N. The protein appeared as a single peak in the UV-absoprtion based elution profile (blue curve). Only monomeric form of the enzyme was detected based on the molecular mass estimation. The figure shows the UV absorption at 280 nm (blue curve), the measured RI (red curve) and the measured RALS (green curve) for the samples at concentration 1.7 mg/ml (A) and 3.5 mg/ml (B).

4.3.3 Crystallization trials on the wild type EDR1 kinase domain

The purification of the wild type EDR1 kinase domain resulted always in a heterogeneous material based on the observed double band in the SDS-PAGE gel after the SEC purification step. MS/MS peptide fingerprinting analysis using both trypsin and chymotrypsin additionally showed that the protein contained several phosphorylation sites and was not uniformly phosphorylated based on accurate mass analysis with ESI-

MS. Instead at least three protein species were observed in 4-, 5- and 6-phosphorylated forms, respectively. Despite of the heterogeneity, crystallization was attempted both in the presence of generic kinase inhibitor staurosporin and the non-hydrolyzable ATP analog AMP-PNP, which were added to the protein solution to final concentration of 0.5 mM and 5 mM, respectively. The crystallization trials were carried out using commercial screens (Qiagen Classic I and II, PEGs I and II and AmSO4 Suite) at the EMBL Hamburg high-throughput crystallization facility (Mueller-Dieckmann, 2006). This resulted in one initial hit from a condition that contained 0.1 M Magnesium chloride, 0.1 M MES pH 6.5 and 30% PEG 400 (PEGs II Suite no. 10). In the corresponding drop clusters of very thin needle crystals grew after a couple of days (Figure 40) but due to their dimensions, they were challenging to test at the X-ray source and also eventually did not produce a diffraction. The subsequent manual reproduction and optimization trials unfortunately also failed. The reason for this was presumably in the variations in the phosphorylation state of the protein resulting in heterogeneity not only within the sample but also between different purification batches. Therefore a kinase inactive form of the protein was produced and eventually successfully crystallized.

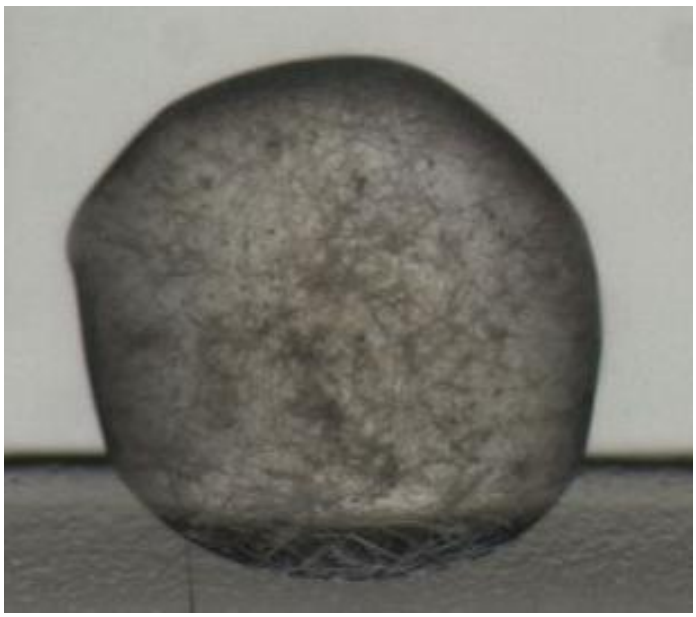

Figure 40. The needle-like initial crystals of EDR1 kinase domain. The crystals grew in a drop containing 0.1 M Magnesium chloride, 0.1 M MES pH 6.5 and 30% PEG 400. These crystals could not be reproduced, most likely due to phosphorylation-dependent heterogeneity between the different protein purification batches.

4.3.4 Obtaining the EDR1-D792N structure

The crystallization of the inactive EDR1 protein kinase domain, EDR1-D792N was performed in the presence of AMP-PNP, a non-hydrolyzable analog of ATP, which was added to the protein solution to a final concentration of 5 mM. Crystallization trials were also conducted to protein material with 0.5 mM staurosporin, a generic kinase inhibitor added, but no crystals were obtained from these trials. Initial crystallization experiments were carried out with 5 commercial crystallization screens (Qiagen Classic I and II, PEGs I and II and AmSO4 Suite) at the EMBL Hamburg high-throughput crystallization facility (Müller-Dieckmann, 2006). All initial screens were performed using the sitting-drop vapour-diffusion method at 292K in 96-well Greiner plates. 300 nl protein solution was mixed with an equal volume of reservoir solution and equilibrated against 50 µl of reservoir solution. Thin, needle-like crystals were growing in several conditions after only few days but the most promising one was with 1.6 M Ammonium sulphate and 0.1 M Bicine pH 9.0 (Qiagen AmSO4 Suite, solution No. 60) (Figure 41A). Rod-like crystals that diffracted to 2.5 Å were obtained from optimization screen with hanging drop vapor diffusion method. These crystals appeared after one week in a drop that contained 1.2 M Lithium sulphate, 10% PEG3350 and 0.1 M Hepes pH 7.8 (Figure 41B).

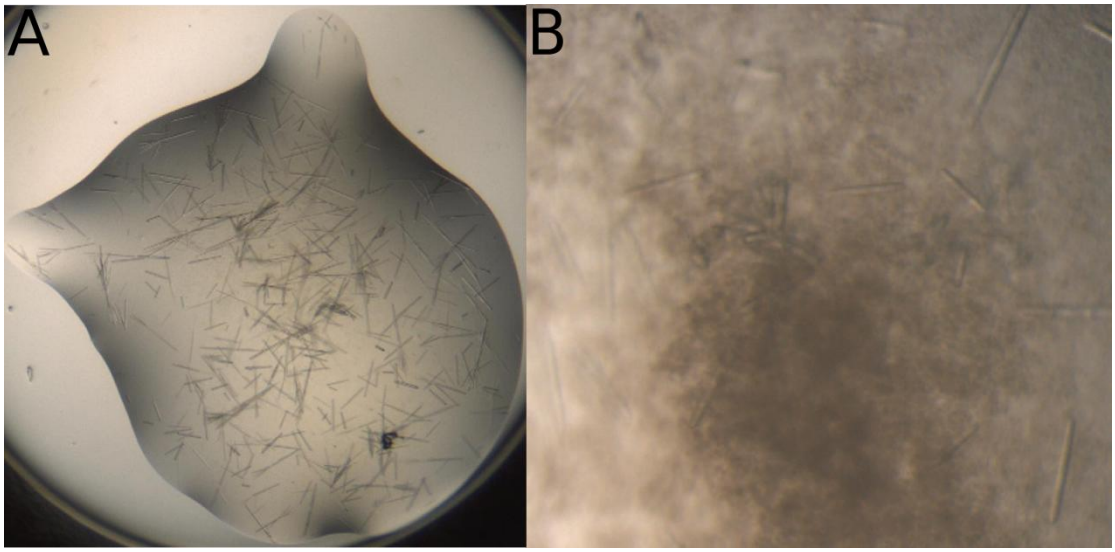

Figure 41. The results from the crystallization trials of EDR1-D792N. A) The initial crystallization experiments with EDR1-D792N produced thin needle-like crystals in condition that contained 1.6 M Ammonium sulphate and 0.1 M Bicine pH 9.0. B) Optimization trials led into slightly thicker rod-like crystals that grew in approximately one week in a drop containing 1.2 M Lithium sulphate, 10% PEG3350 and 0.1 M Hepes pH 7.8.

Prior to data collection, a single crystal of EDR1kd-D792N grown in the presence of AMP-PNP was soaked for about 5s in mother liquor augmented with 15% glycerol as a cryoprotectant and 10 mM AMP-PNP. Out of the best diffracting crystal a total of 800 images were collected with an oscillation range of 0.1º at the PETRA-III beamline P13, EMBL Hamburg, Germany. The data were indexed and integrated using *XDS* (Kabsch, 2010) and scaled with *SCALA* (Collaborative Computational Project, Number 4, 1994).

The crystals of the EDR1-D792N belonged to the trigonal space group $P3_221$ with unit cell parameters $a = b = 90.57$ and $c = 185.90$ ($\alpha = \beta = 90°$, $\gamma = 120°$). The phases were obtained through automated molecular replacement using program *MOLREP 11.0* as part of *CCP4 Program suite 6.3.0* (Collaborative Computational Project, Number 4, 1994) and the crystal structure of the inactive *A. thaliana* CTR1 protein kinase domain (PDB code: 3P86, Mayerhofer *et al*., 2012) as a search model. The asymmetric unit

contained two molecules with a calculated solvent content of 64.24 % and Matthews coefficient V_m = 3.44 $Å^3$/Da (Matthews, 1968). The calculation of Matthews coefficient indicated the possibility of third molecule in the asymmetric unit with corresponding V_m – value of 2.29 and solvent content 46.36 % but attempts to obtain phase solution for three molecules per a.u. failed. Therefore the dimeric packing of the molecules was considered to be correct. The initial model was refined against the native dataset with program *REFMAC 5.7* (included in CCP4 program suite 6.3.0) to obtain interpretable electron density maps for manual model building in *COOT 0.7* (Emsley *et al*., 2010). A large section of unassigned density was identified in the putative active site of each monomer enabling the modeling of one AMP-PNP molecule (PDB Ligand code: ANP) per monomeric protein unit. The final model with R-factor of 0.191 and R_{free} of 0.249 resulted from iterative cycles of manual model building and refinement using the programs *REFMAC 5.7* and *COOT 0.7*. The data quality and refinement statistics are shown in table 28.

Table 28. Data collection and refinement statistics for EDR1-D792N structure. Values in the parenthesis are for the high resolution shell.

	EDR1-D792N
Data collection	
X-ray source	P13, PETRA III
Wavelength (Å)	1.3858
Space group	$P3_221$
Cell dimensions	
a, b, c (Å)	90.57, 90.57, 185.9
α, β, γ (°)	90.0, 90.0, 120.0
Resolution (Å)	92.93–2.55 (2.69-2.55)
R_{merge}	9.5 (73.0)
$I / \sigma I$	8.6 (2.0)
Completeness (%)	97.5 (96.4)
Redundancy	3.3 (3.2)
Refinement	
Resolution (Å)	78.45 – 2.55
No. reflections	29,533
R_{work} / R_{free}	19.1/24.9
No. atoms	
Protein	4,305
Ligand/ion	130
Water	6
B-factors	
Protein	68.1
Ligand/ion	76.9
Water	89.6
r.m.s. deviations	
Bond lengths (Å)	0.02
Bond angles (°)	1.9

4.3.4 Structural characterization

4.3.4.1 EDR1-D792N structure

The crystal structure for EDR1-D792N was obtained in the presence of the ATP analog AMP-PNP with two protein molecules per asymmetric unit. As expected, the inactive kinase domain of EDR1 adopted the characteristic protein kinase domain fold as described first for the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase (PKA-Cα) (Knighton *et al.*, 1991). The structure of EDR1-D792N

consisted of two lobes: N-terminal one (residues 655 – 744) and larger C-terminal one (750 – 933) connected to eachother with a short linker (residues 745 – 749) (Figure 42). The first 6 N-terminal residues appeared to be disordered based on the lack of interpretable electron density. Additionally, the last or the last two C-terminal residue(s) were missing in the protomers B and A, respectively. As in the canonical kinase core structure, the N-terminal lobe in EDR1-D792N consists mainly of β-strands that assembly to form the central β-sheet required for the binding of the nucleotide. However, in contrast to the typical five-stranded β-sheet seen in other protein kinase domain structures the corresponding β-sheet in EDR1-D792N is formed by six β-strands. Due to disruption in the hydrogen-bonding pattern through the altered orientation of the Arg674, the first β-strand is split into two separate β-strands, referred here as β1A and β1B. The β1A strand (residues 670 – 673) interacts with the C-terminal residues 685 - 688 of the β2-strand whereas the β1B forms an antiparallel interaction with four N-terminal residues of β2 (residues 681-684). In addition to the β-sheet, the N-terminal lobe of EDR1-D792N contains the αC-helix (residues 707 – 720), which is one of the characteristic structural elements in protein kinases and is involved in mediating the changes between the inactive and active state of the protein. The loop preceding the αC-helix is more disordered in the protomer B of EDR1-D792N than in the protomer A based on the lack of density for the residues Ala702 and Phe703. In both the residues from Asp200 to Ala706 show higher B-factors compared to the average for the whole protein molecule. This phenomenon is not unusual in flexible linker regions that mediate movement, as is the case here.

The C-terminal lobe of EDR1-D792N is mainly α-helical and consists of seven α-helices (αD – αI) and four 3/10 helices. Additionally this subdomain includes an antiparallel two-stranded β-sheet made out of strands β7 (residues 798-800) and β8 (residues 806-808). As expected for an inactive kinase, two short β-strands (β6 and β9

according to the nomeclature in PKA-Cα structure) are missing in EDR1-D792N structure. The residues of the β6-strand (residues 787 – 789) precede the conserved HRD motif (residues 790 – 792), which contains the highly conserved and catalytically important aspartate residue (Asp792). This motif is part of the catalytic loop, which is located between the αE-helix and β-strand β7 and in EDR1 spans the residue range from 784 to 797. The β-strand β9 is part of the structurally flexible activation loop, which is flanked by the conserved DFG and APE motifs and usually contains the primary phosphorylation site of the enzyme. In protomer B this activation loop is fully structured whereas in the protomer A a four-residue segment (residues 825-829) could not be build into the model. Both protomers also showed weak or complete lack of density for the residues Lys759 and Ser760, which are part of a surface loop between αD- and αE-helices in the C-terminal lobe.

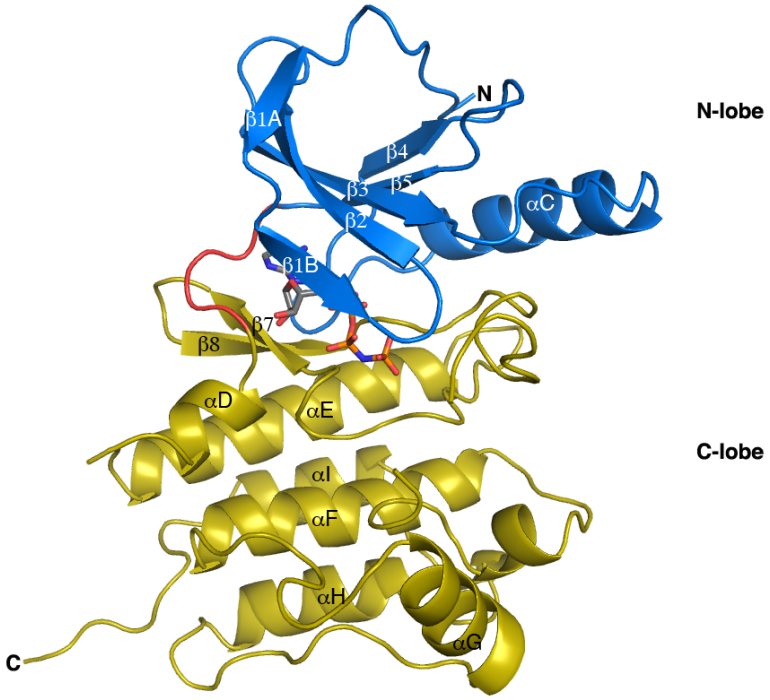

Figure 42. The EDR1-D792N structure. As expected, EDR1-D792N exhibited the typical bilobal structure of protein kinases as shown here for the protomer B. The N-terminal lobe (in blue) typically consists of five-stranded β-sheet (β1-β5), however in the case of EDR1-D792N the β1 is split into two β-strands, β1A and β1B. Additionally the N-lobe includes the important αC-helix, which changes its orientation when the kinase shuffles between inactive and active state. The larger C-terminal lobe (in gold) has mainly α-helical structure and is formed by the α-helices αD – αI. The protein is shown as cartoon representation and the bound AMP-PNP molecule is depicted as grey sticks.

4.3.4.2 Dimerization

The involvement of dimerization in protein kinase domain activation is well-established phenomenon both among prokaryotic and eukaryotic members of Ser/Thr protein kinases (STPKs) and protein tyrosine kinases (PTKs). An example is the activation of well-known oncogenic protein kinase BRAF, which is dependent on both homodimerization and also heterodimerization with pseudokinase KSR (kinase suppressor of Ras) (Rajakulendran *et al.*, 2009) A recent example for the dimerization-

dependent activation is the kinase domain of CTR1 (CTR1-kd), which is in its active state forms similar back-to-back dimers observed for its eukaryotic homolog B-RAF (Mayerhofer *et al*., 2012). In CTR1-kd three residues, Arg604, Phe611 and Met612 mediate the central interactions in the back-to-back interface in a similar manner as observed for BRAF and KSR (Mayerhofer *et al*., 2012). Mutations that block the back-to-back dimer formation significantly reduce the kinase activity of the protein thus indicating the biological importance of this interface (Rajakulendran *et al*., 2009; Mayerhofer *et al*., 2012).

Despite its high similarity with CTR1-kd, EDR1 kinase domain didn't appear to form dimers based on the static light scattering experiments on the Malvern system. However, in the crystal structure EDR1-D792N had surprisingly adopted a dimeric arrangement where the active sites of the protomers faced away from eachother as expected for an authentic back-to-back interface (Figure 43). To examinate the possibility to a similar back-to-back interface as observed for CTR1-kd, the dimer interface of EDR1-D792N was analyzed. It was noticed that the position of the conserved back-to-back interface residues (Arg722, Phe729 and Leu730 in EDR1) differed significantly from the expected one. In the CTR1-kd structure the corresponding residues are correctly positioned at the center of the dimer interface, where they are able to interact with one another. Similar interactions are not observed in EDR1-D792N structure due to an altered orientation of the two protomers in comparison to CTR1. Instead, the three residues are approximately 14 Å apart from their counterparts and locate near the periphery of the dimer interface (Figure 44A).

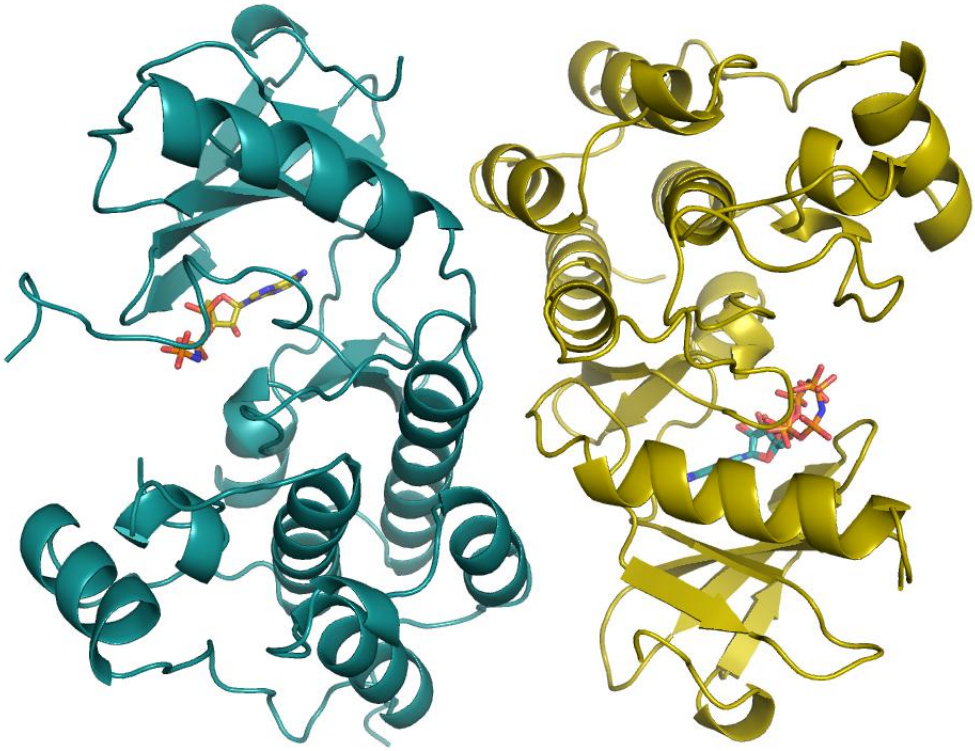

Figure 43. The EDR1-D792N dimer as observed in the asymmetric unit. The active sites of the two protomers are located at the opposite faces in the observed dimer. This resembles the arrangement in the classical back-to-back interface.

The lack of interaction between the central back-to-back interface residues in EDR1-D792N indicated that this enzyme had adopted a different type of dimeric arrangement in the asymmetric unit. The biological relevance of the observed dimer interface was analyzed using PISA (Krissinel&Hendrick, 2007). A set of 9 residues from each chain was attributed to the dimer interface and covered 740 $Å^3$ and 755 $Å^3$ for the protomers A and B, respectively (Figure 44B). Firstly, the small number of strong interactions (hydrogen bonds, salt bridges) at the interface indicated a weak protein-protein contact. Secondly, the calculated surface area at the interface was significantly smaller than the one in the back-to-back dimer interface of CTR1, which covers 1249 $Å^3$.

Based on these two observations, it was considered to be unlikely that the EDR1-D792N would have similar function as the canonical back-to-back interface. Additionally, the interface was classified as crystallographic contact by PISA.

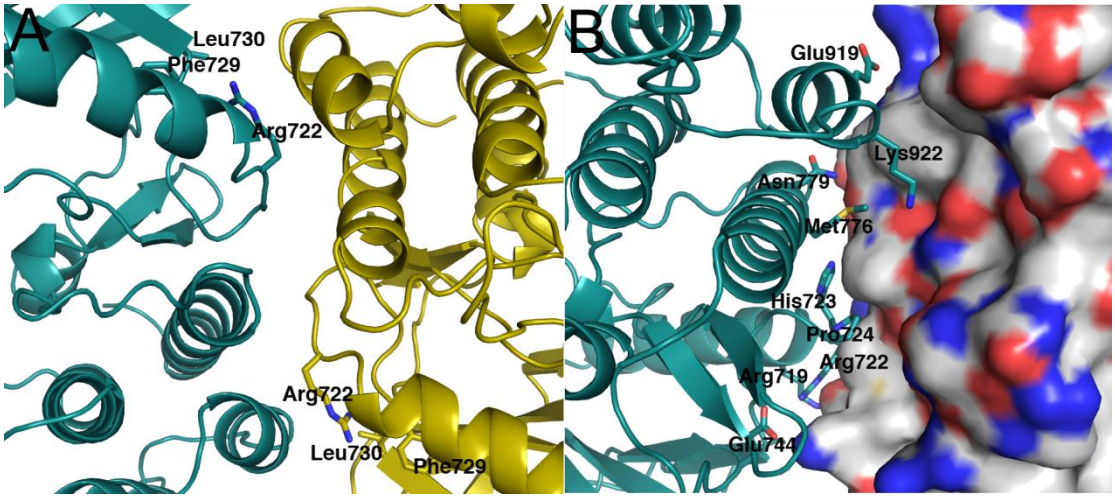

Figure 44. The EDR1-D792N dimer interface as observed in the asymmetric unit. A) The position of the back-to-back dimer interface residues in the EDR1-D792N dimer interface. B) The dimer formation in EDR1-D792N is mediated by a set of 9 residues, which cover a total of 740 $Å^3$ and 755 $Å^3$ for the protomers A (shown as green cartoon representation) and B (as surface representation with carbon atoms in grey, oxygens in red and nitrogens in blue), respectively.

4.3.4.3 EDR1-D792N protomers in different activational state

The accurate mass analysis with ESI-MS indicated that the EDR1-D792N would exist in a non-phosphorylated state and therefore inactive state. This was further supported by results from the kinase activity assays. Surprisingly, for one of the protomers (protomer B) a block of unassigned density was observed at the proximity of two threonine residues (Thr820 and Thr827). Furthermore, in both cases the observed density was continuous with the side chain of the threonine residue, which allowed the modeling of phosphorylated residue at the two positions. Although this didn't correlate with the results from the activity assays and from the accurate mass determination experiments for EDR1-D792N, the location of the phosphorylations supported their

authenticity. Firstly, both of the putative phosphothreonine residues are located in the activation loop, which is known to be a primary site for phosphorylation in the active protein kinases. Furthermore, both Thr820 and Thr827 were listed as putative phosphorylation sites according to the peptide fingerprinting analysis of the WT EDR1 kinase domain. Additionally, in the sequence alignment of EDR1 with CTR1 (Figure 45), Thr820 correspond to Thr704, which is a known phosphorylation site in CTR1 (Mayerhofer *et al.*, 2012). The same phenomenon was seen in the structural alignment of EDR1-D792N with active form of CTR1 kinase domain (PDB code: 3PPZ), where the phosphothreonine 820 (TPO820) adopted a nearly identical position with the phosphothreonine 704 (TPO704) (Figure 46A). Finally, the second phosphothreonine (TPO827) in EDR1-D792N was found to lie within a so-called basic pocket, where it interacts with two conserved arginine residues, Arg791 and Arg815 (Figure 46B). These interactions are characteristic to the primary phosphorylation site in the active form of RD-kinases (Nolen *et al*., 2004).

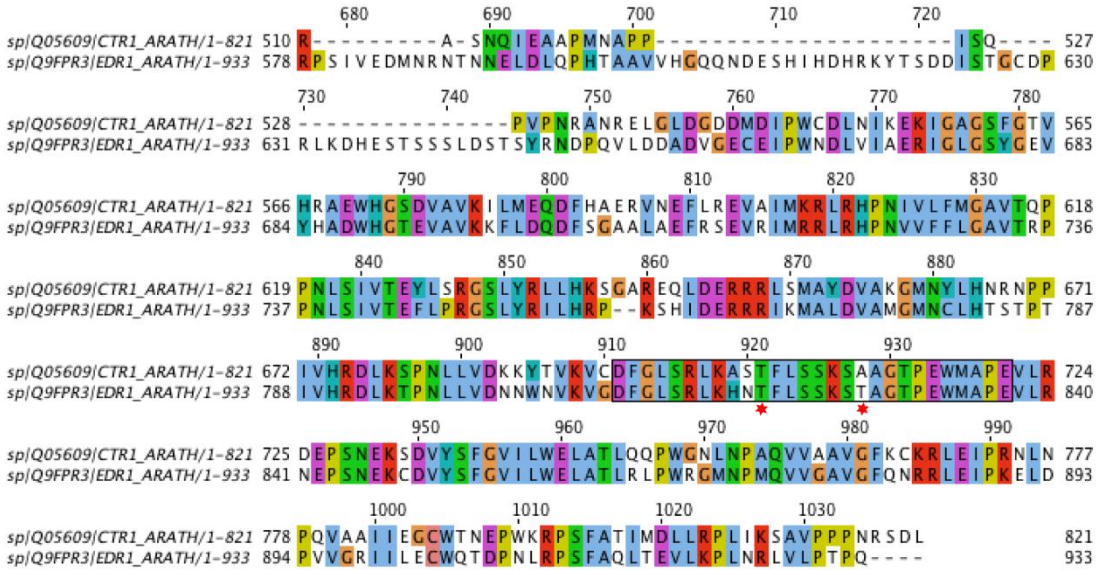

Figure 45. The sequence alignment of EDR1 with CTR1. The activation segment is marked with square and the phoshoporylation sites of EDR1 as seen in the EDR1-D792N structure with red stars.

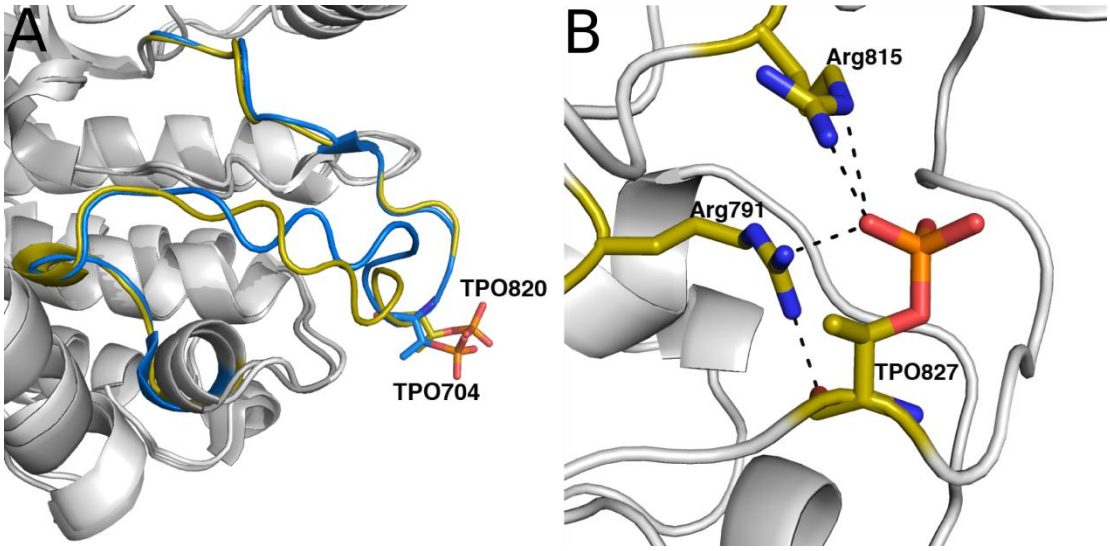

Figure 46. The location of the two phosphothreonines in EDR1-D792N protomer B. A) The phosphothreonine 820 (TPO820) of EDR1 overlays with TPO704 of CTR1 (PDB code: 3PPZ) in the structural alignment of the two enzymes. The structures of EDR1-D792N and CTR1 kinase domain are shown as light grey and grey cartoon models with the activation loops in gold and blue, respectively. The phosphoresidues are shown as sticks. B) The second phosphoresidue, TPO827, interacts with two arginine residues within a so-called basic pocket as seen with other RD-kinases for the primary phosphorylation site.

Interestingly, neither of the above mentioned threonine residues showed any sign of phosphorylation in the protomer A of EDR1-D792N. Furthermore, Thr827 is part of the activation loop segment, which is unstructured in this protomer. Since the stabilization of activation loop through phosphorylation is a known phenomenon in the transition from inactive to active state in protein kinases, the observed difference between the two protomers indicated that they might be in different activational states. In order to see if there would be additional factors to support this hypothesis, the structures of the two protomers were investigated in detail. The structural alignment of the EDR1-D792N protomers showed a high overall similarity in the protein backbone (r.m.s.d.-value of 0.495 Å for 203 C_{α}-atoms) exept for the catalytically important αC-helix and the

activation loop (Figure 47A). Both are known to be the key regulatory elements that mediate the transition between inactive and active state (Huse and Kuriyan, 2002). In the active-like protomer B, the N-terminal end of the αC-helix is approximately 5Å closer to the core of the protein than in the other protomer. The difference arises from the combined translational and rotational motion perpendicular to the pivotal axis of the helix. The observed rotation is approximately 10° around the helical axis. The αC-helix contains a conserved glutamate residue (Glu714), which in an active kinase forms a salt bridge with a conserved lysine residue (Lys696) from the β3-strand. The altered orientation of the αC-helix in the protomer B enables the formation of this salt bridge (Figure 47B). In the protomer A the glutamate is pulled away from its counterpart as would be expected for an inactive kinase (Figure 47C).

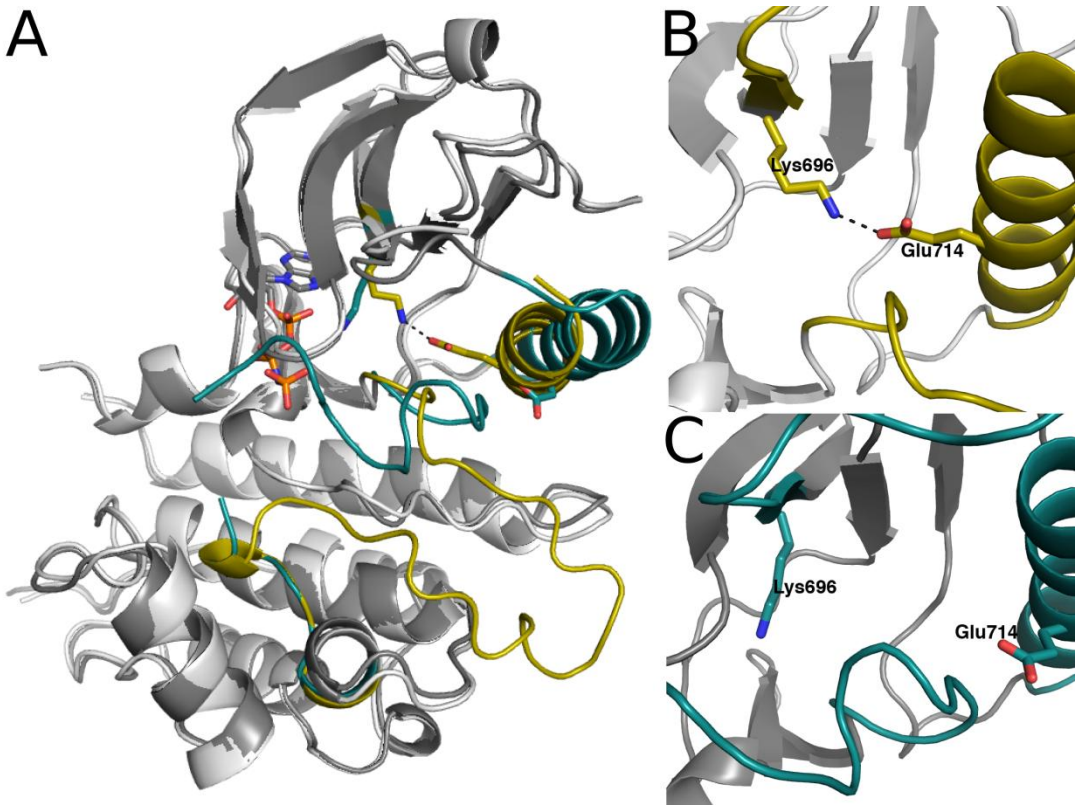

Figure 47. The overlay of the two EDR1-D792N protomers reveals distinct differences in their folding pattern. A) The most profound differences are seen in the orientation of the catalytical αC-helix (residues 707 – 720) and of the activation loop (residues 810 – 837) highlighted here in dark green and gold for the protomers A and B respectively. B) Additionally in the protomer B a hydrogen bond between the conserved lysine (Lys696) at the N-terminal part of the protein and conserved glutamate (Glu714) in the αC helix is observed. This hydrogen bond is one of the hallmark features for an active kinase and does not exist in the EDR1-D792N protomer A. C) In the protomer A the corresponding lysine and glutamate residue locate far apart due to the altered position of the αC-helix.

The structural rearrangements that stabilize the activation loop are known to occur in the transition from inactive to active kinase. The phosphorylation of this flexible protein segment plays an important role in the stabilization process. In the protomer B of EDR1-D792N the interaction of TPO827 with Arg791 and Arg815 facilitate the correct folding of the activation loop. The former arginine residue is part of the conserved HRD motif in the catalytic loop and forms a hydrogen bond with the phosphate group oxygen

of TPO827 through its guanidinium group. The latter arginine is part of the activation loop and localizes two residues apart from the conserved DFG motif that marks the beginning of this loop. The hydrogen bonds between the TPO827 phosphate group oxygen and the hydrogen donors NH2 and NE of Arg815 help to orient the N-terminal end of the activation loop. As a result, the residues of the β6 and β9 are brought into close contact and appear to form nearly a β-sheet like hydrogen bonding network (Figure 48A). This resembles very closely to the structural arrangement observed in active kinases. In the protomer A the lack of phosphorylation and subsequent lack of structural organization in the activation loop results in increased distance between the putative β6 and β9 strand residues as would be expected for an inactive kinase (Figure 48B).

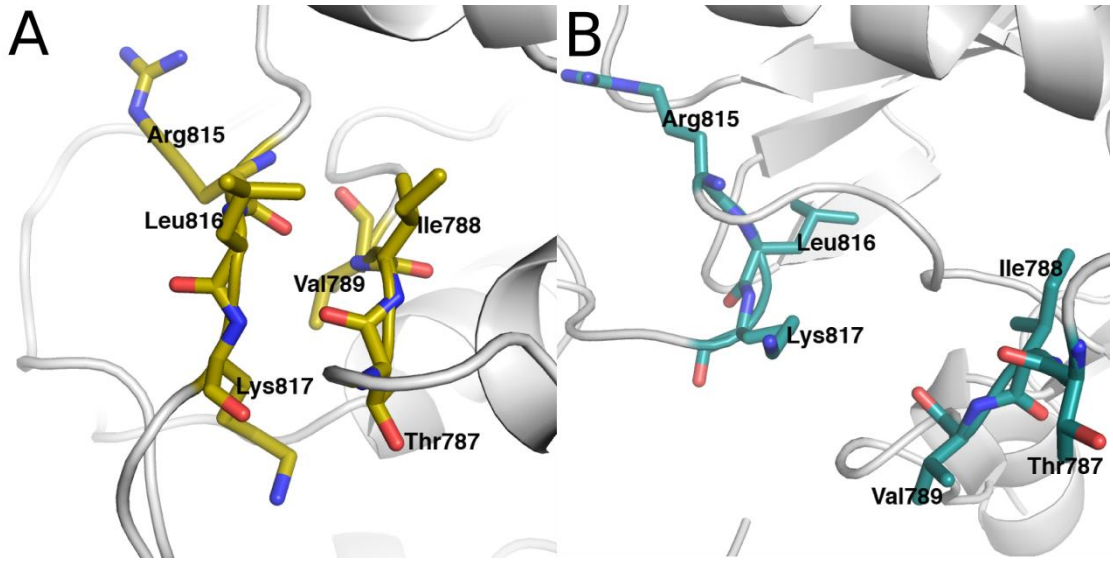

Figure 48. The organization of the putative β6- and β9-strand residues. One of the characteristic features of an active kinase is a small β-sheet formed by the β6- and β9-strand. A) In the active-like protomer, the rearrangement of the activation loop brings the residues of the putative β9-strand (815 – 817) in close proximity with the residues of its pairing β-strand (β6, residues 787 – 789). This observation further supports the active-like state of the protomer B. B) The corresponding residues in the protomer A are not interacting with eachother as would be expected for an inactive kinase.

4.3.4.4 Enzyme-substrate complex

The structural analysis of the EDR1-D792N protomers indicated that one possessed the characteristics of an inactive kinase whereas the other resembles closely an active kinase. Recently a similar phenomenon was seen in the crystal structure of p21-activated kinase PAK1, where an inactive protomer was paired with a protomer in an active-like conformation (Wang *et al*., 2011). The crystallized protein in this case was also a catalytically inactive variant of the enzyme. In the structure of this PAK1-$KD^{K299R/D389N}$ the protomers formed an enzyme-substrate complex, where the activation loop of the inactive protomer served as a substrate for the protomer in the active-like state. Since the orientation of the EDR1-D792N protomers within the asymmetric unit differed from the observed ES-complex of PAK1-$KD^{K299R/D389N}$, the crystallographic contacts across the different asymmetric units were examined. Interestingly, an interface that mimicked the trans-autophosphorylation complex was indeed found between the symmetry-related molecules from the neighbouring asymmetric units (Figure 49A). As in the PAK1-$KD^{K299R/D389N}$ structure, the activation loop of the inactive protomer ("substrate") protrudes to the active site of the active-like protomer ("enzyme"). Furthermore, a serine residue (Ser824) from the activation loop of the substrate molecule is optimally positioned to interact both with the mutated catalytically important aspartate residue (Asp792) and with the γ-phosphate group of the bound AMP-PNP in the active site of the neighbouring protomer (Figure 49B). This appears to mimic the state prior to the phosphoryl group transfer. The side chain hydroxyl group of Ser824 is within 3.8Å and 3.0Å range from the γ-phosphate of AMP-PNP (conformer B) and from the ND2 atom of Asn792, respectively.

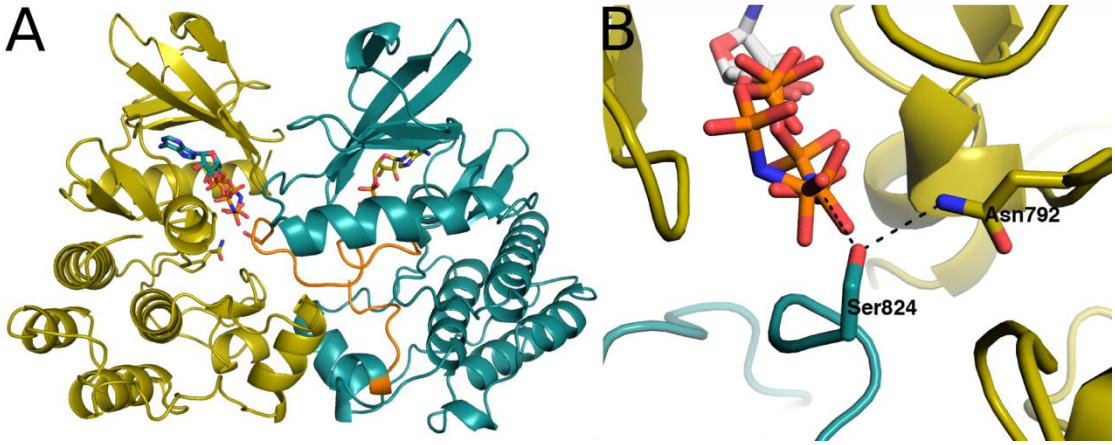

Figure 49. The ES-complex of EDR1-D792N. A) The ES-complex interface is observed between the adjacent protein molecules across the different asymmetric units. The partially disordered activation segment (in orange) of the "inactive" monomer A (in dark green) protrudes to the active site of the monomer B (in gold). B) A serine residue (Ser824) of this activation segment interacts with the γ-phosphate of the AMP-PNP molecule and with the mutated catalytically important aspartate (Asp792) as indicated with the black dashed lines.

In addition to its role as a substrate for the adjacent active-like protein molecule, the activation loop of the inactive protomer functions as an anchor to dock the substrate molecule into correct position. A set of 5 residues from this activation loop interacts via hydrogen bonds with the residues of the adjacent molecule (Figure 50A). These include two serine residues Ser823 and Ser824, which form hydrogen bonds through their OG1 and main chain oxygens, respectively. Their interaction partners are Lys794 from the catalytic loop and Thr830 from the activation loop of the neighbouring molecule. Leu822 and Phe821 participate in the coordination network via their main chain oxygen atoms, which find their hydrogen bond parthers from the side chain hydrogen donors of Ser679 and Lys794. Additional hydrogen bonds are observed between the side chain of Asn819 and the main chain oxygens of Trp867 and Met870. Asn819 forms also a hydrogen bond to the side chain of Glu832.

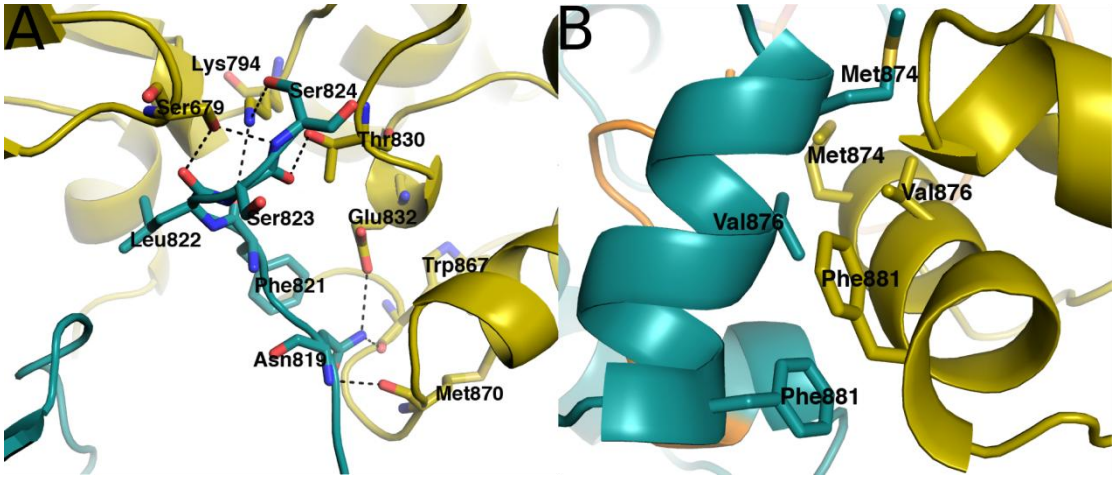

Figure 50. The anchor residues in the ES-complex interface. A) The residues from the activation loop of the inactive protomer (in dark green) form hydrogen bonds (black dashed lines) with the residues from the adjacent molecule (in gold). B) The specific αG-helix residues provide the important hydrophobic interactions between the protein molecules at the ES-complex interface.

The residues of the αG-helix also play an important role in the ES-complex interface of EDR1-D792N as previously observed in the crystal structure of *Mycobacterium tuberculosis* PknB receptor serine/threonine kinase variant (Mieczkowski *et al*., 2008). The specific residues of the αG-helix stack together through hydrophobic interactions as shown in the figure 50B. The most important interactions appear to be mediated by the side chains of Val876 and Phe881. As the mutation of the important αG-helix residues in *M. tuberculosis* PknB led to a significant loss of enzymatic activity, it was investigated whether the substitution of Val876 and Phe881 with glutamate would have the same effect on the kinase activity of EDR1. In addition to the single mutants, a double variant where both αG-helix residues were substituted with glutamate was made. The kinase activity of the enzyme variants was tested using the coupled kinase assay as described in materials and methods. The wild type EDR1 kinase domain was tested in parallel as comparison. The results of the coupled kinase assay results for the mutants and

for the wild type enzyme are shown in figure 51. As expected, the wild type enzyme was active with V_{max} and K_m values similar to the ones calculated before (see chapter 4.3.2.3). The three αG-helix mutants appeared to have lost their activity thus confirming the importance of the Val876 and Phe881 mediated interaction to the activity of the enzyme.

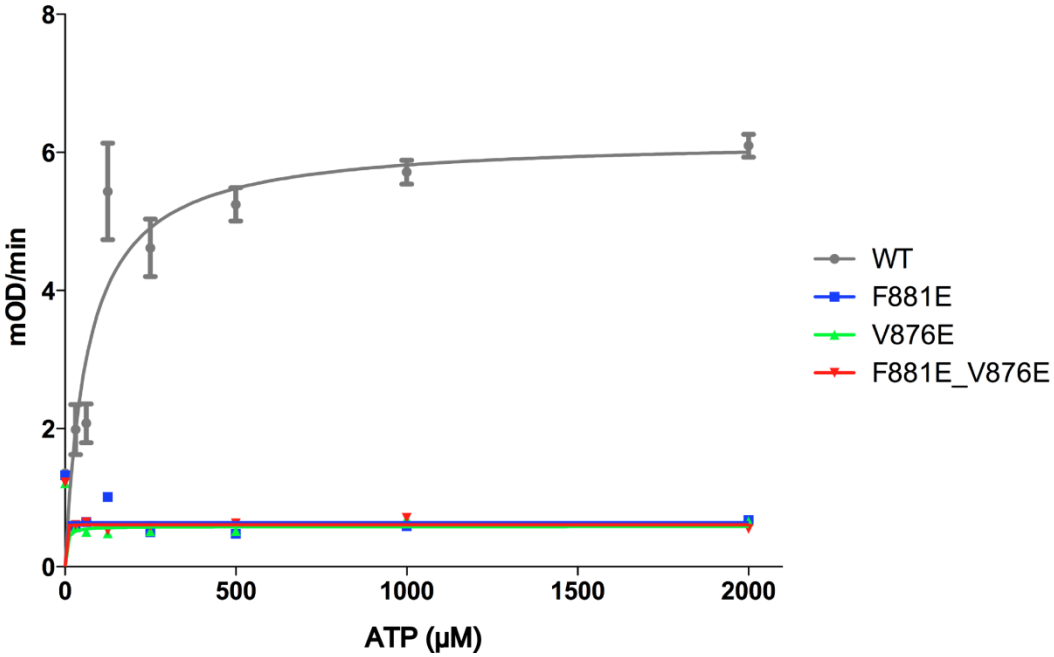

Figure 51. The results from the coupled kinase assay for the αG-helix mutants of EDR1 kinase domain. The mutation of the two αG-helix residues (Val876 and Phe881) renders the enzyme inactive, which supports the importance of the αG-helix mediated interaction at the ES-interface and furthermore confirms the requirement of the ES-complex interface for the autophosphorylation activity if the enzyme.

4.3.4.4 Nucleotide-binding in EDR1-D792N

The combined co-crystallization and soaking with AMP-PNP resulted in successful incorporation of the nucleotide into the active site of the inactive EDR1 kinase domain. In both protomers a bulk of unassigned density was observed in the cleft between the two lobes that allowed the placing of the nucleotide analog to this site (Figure 52).

The observed density at the active site of the protomer B appeared to be large enough to fit two conformations for the AMP-PNP molecule. In the two conformers the β-phosphate groups point to opposite directions allowing however the γ-phosphates to be positioned in similar environment.

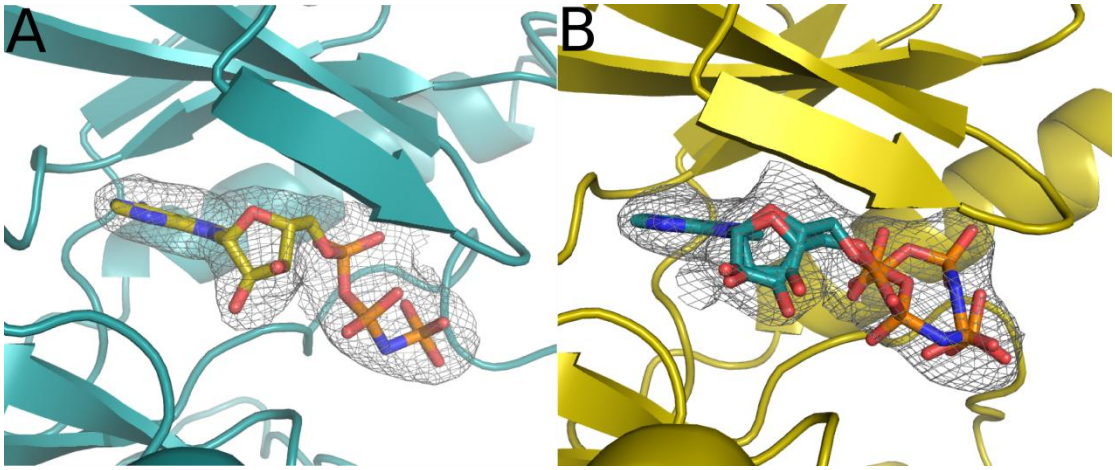

Figure 52. The AMP-PNP bound to the active site of EDR1-D792N. A clear electron density for the AMP-PNP was observed in the active site of both protomer A (A) and B (B). The difference electron density for the images was calculated by omitting the nucleotide from the EDR1-D792N model. The map contour level was set up for 0.0502 electrons/$Å^3$ ($\sigma = 2.0$). The program FFT (Read & Schierbeek, 1988) as part of CCP4 Suite were used to create the omit maps.

The location of the nucleotide-binding pocket is conserved among the protein kinases. The lid of the binding pocket is formed by the highly conserved P-loop, which is found from all kinases and is important for the ATP-binding. The tip of the P-loop interacts with the γ-phosphate group of the nucleotide. Other structural elements that define the binding pocket include the linker region defines the bottom of the pocket and contributes to the nucleotide binding both through hydrophobic interactions and hydrogen

bonds. The frame of the AMP-PNP lies on top of β-strand β7 and the preceding 3/10 helix, both if which are part of the C-terminal lobe.

In both EDR1-D792N protomers the residues that form the binding pocket for the adenosine unit appeared to adopt a specific rotamer and show conserved interaction with the ligand (Figure 53). The adenosine ring is flanked by the side chains of Val683 and Leu799, which form hydrophibic interactions with the bound ligand. The bottom of the nucleotide-binding pocket is capped by the phenylalanine side chain (Phe745). The same phenylalanine also takes part in the hydrogen bond network through its main chain aminogroup, which interacts with the N1 nitrogen of the nucleotide. More variability is observed in the surroundings of the phosphate-tail. This is partly due to the conformational flexibility in the triphosphate unit of the AMP-PNP in protomer B but also changes in the rotameric form of certain residues, such as Arg753 and Lys696. In protomer A these two residues are involved in the hydrogen bond network with the ligand. However, in the active-like protomer the side chains of these two residues are facing away from the active site. Furthermore, the Lys696 is involved in the conserved hydrogen bond that is characteristic for an active kinase.

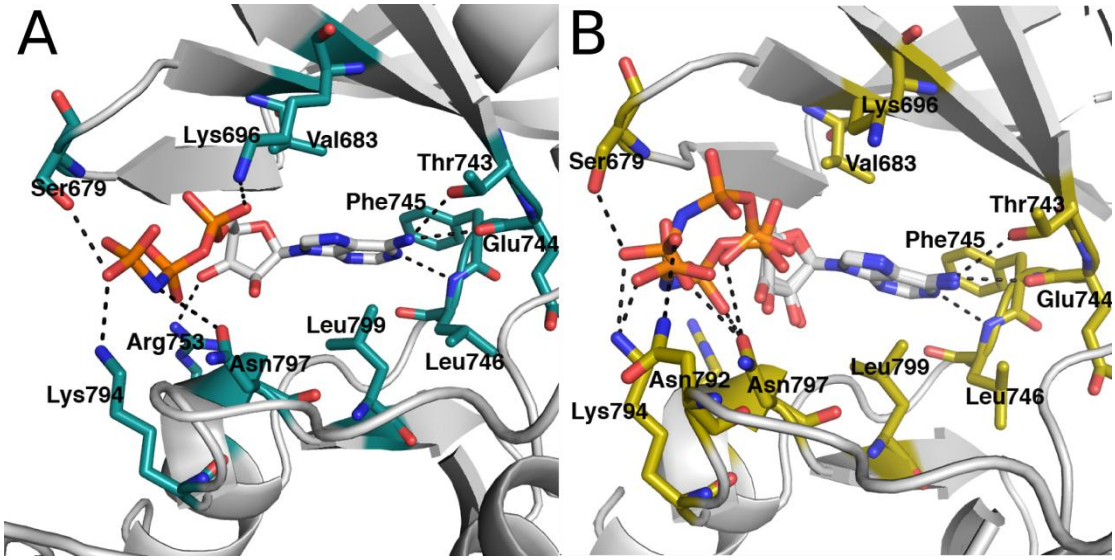

Figure 53. The AMP-PNP binding in the EDR1-D792N. A) The residues involved in the binding of the AMP-PNP in protomer A are shown as green stick representation. B) The AMP-PNP binding pocket in protomer B similar to its counterpart in protomer A. A few differences in the hydrogen bond network exist due to the altered orientation of Arg753 and Lys696 and also because of the additional AMP-PNP conformer in protomer B.

4.4 Adenylation domain ApnA A_1 of Non-Ribosomal Peptide Synthetase from *Planktothrix agardhii* strain PCC7821

4.4.1 The ApnA A_1 PCC7821 constructs

The ApnA A_1 is the first adenylation domain in the Anabaenopeptin synthetase, which is a Non-ribosomal peptide synthetase from cyanobacterium *Planktothrix agardhii* responsible for the synthesis of anabaenopeptin. The studied ApnA A_1 enzyme from *P. agardhii* strain PCC7821 consisted of 546 amino acids and possessed the unusual ability to adenylate two very different amino acids, namely L-arginine and L-tyrosine

(Christiansen *et al.*, 2011). The wild type enzyme was studied using X-ray crystallography to gain insights into its bi-specificity. From the structural studies, three active site residues, Glu204, Ser243 and Ala307 were found to play important role in the substrate binding and recognition. The importance of the residues in positions 243 and 307 has already been indicated in the phylogenetic and biochemical studies on ApnA A_1 enzymes from different *Planktothrix* strains (Christiansen *et al.*, 2011). Furthermore, according to sequence aligment of ApnA A_1 PCC7821 with PheA, the three mentioned residues are part of the so-called specificity-conferring code (Table 29). The residues included in the specificity conferring code have been indicated to define the substrate specificity of the adenylating enzyme (Stachelhaus et al., 1999). Substitutions to the positions 204, 243 and 307 of ApnA A_1 PCC7821 were made to investigate the effect on the binding specificity of the enzyme. The created variants are listed in the table 2 (in materials and methods section).

Table 29. The ApnA A_1 PCC7821 residues that correspond to the specificity-conferring code (Stachelhaus *et al*., 1999) from PheA.

PheA	**ApnA A_1**
Asp235	Asp200
Ala236	Val201
Trp239	Glu204
Thr278	Ser243
Ile299	Ile271
Ala301	Gly273
Ala322	Val298
Ile330	Ile306
Cys331	Ala307
Lys517	Lys500

4.4.2 Protein production

The wild type ApnA A_1 PCC7821 and majority of the mutants were expressed and purified in the laboratory of Prof. Andrea Rentmeister (University of Münster). The mutants of ApnA A_1 PCC7821 purified at EMBL Hamburg are listed in the table 30 together with the final yield from total 800 ml of bacterial culture. The detailed expression and purification protocol has been described in materials and methods section. The NiNTA purification graphs and corresponding SDS-PAGE results for the Ala307 variants (A307C, A307D, A307K and A307N) are shown as an example in the figure 54. The protein species in the elution fractions appear slightly above the 55 kDa molecular marker band in the SDS-PAGE gel, thus corresponding roughly to the calculated 63.4 kDa molecular weigth of the ApnA A_1 with the His_6-tag attached. Similar results were obtained for the serine mutants, S243E, S243G, S243H and S243Q, and for the lysine mutants K415A, K500A and K415A/K500A. Based on the SDS-PAGE analysis, the elution fractions that contained the target protein were pooled together and dialyzed overnight at 4°C. The dialyzed sample for each mutant was analyzed on SDS-PAGE gel to confirm the final purity level of the sample. The results for this analysis are shown in the figure 55. High level of purity was observed already after dialysis step and therefore no additional purification steps were performed.

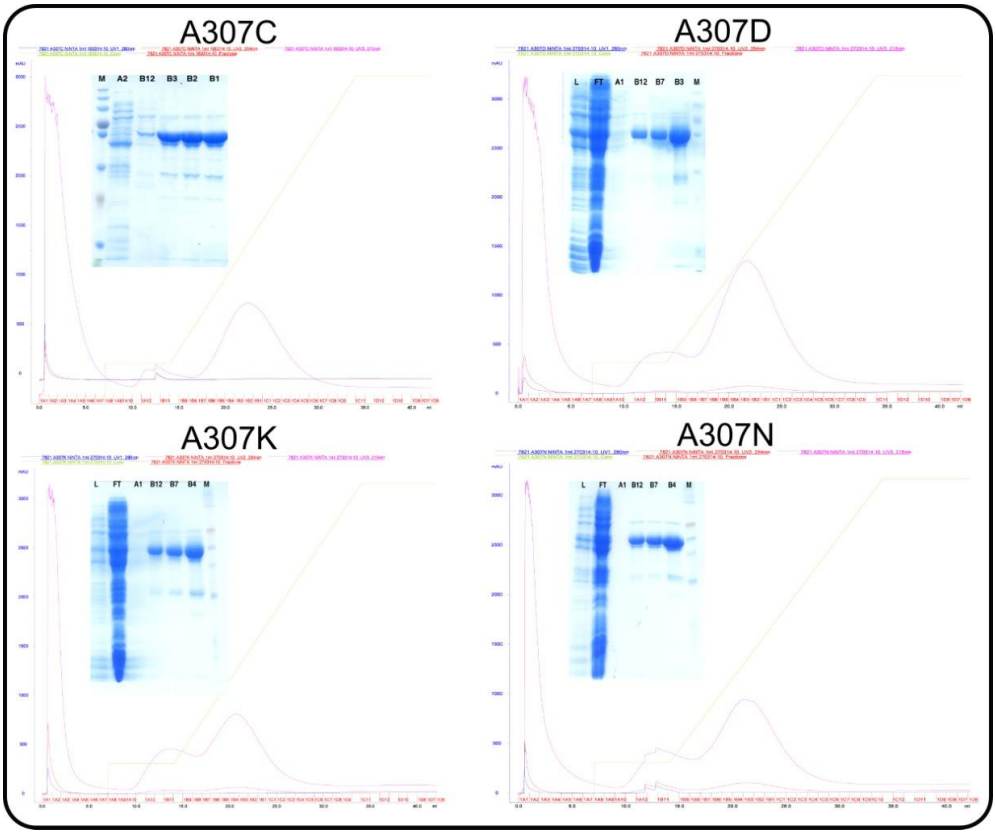

Figure 54. An example of the NiNTA purification results for the ApnA A1 PCC7821 variants. The NiNTA purification results for four of the A307 variants are shown. The abbreviations used: L = lysate, FT = Flow through, M = PageRuler Color Plus Prestained protein ladder (Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany). The remaining lanes correspond to a specific purification fraction in the elution profile.

Table 30. The yield of pure protein after final purification step for the ApnA A_1 PCC7821 mutants purified at EMBL Hamburg.

ApnA A_1 PCC7821 mutant	**Final yield (mg)**
A307C	2.9 mg
A307D	5.8 mg
A307N	4.5 mg
A307K	4.1 mg
S243E	1.2 mg
S243G	1.6 mg
S243H	3.6 mg
S243Q	6.1 mg
E204G	1.8 mg
K415A	6.1 mg
K500A	4.1 mg
K415A/K500A	5.2 mg

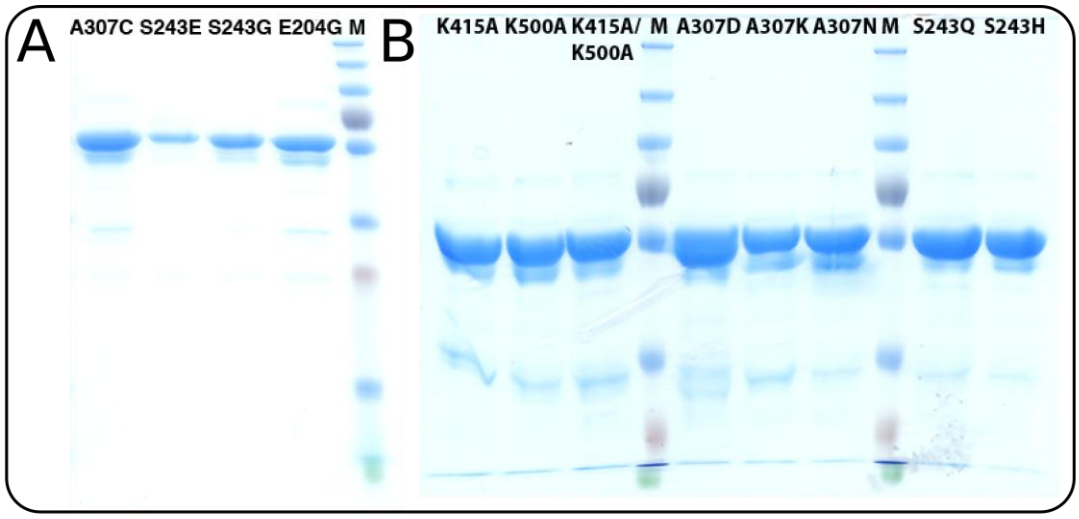

Figure 55. The dialyzed samples for the ApnA A_1 PCC7821 mutants were analyzed on SDS-PAGE to evaluate their purity. M = PageRuler Color Plus Prestained protein ladder (Fermentas/Thermo Fisher Scientific – Germany GmbH, Schwerte, Germany). A) The SDS-PAGE analysis for A307C, S243E, S243G and E204G confirmed their purity. B) Same applied for the lysine (K415A, K500A and K415A/K500A), alanine (A307D, A307K, A307N) and serine mutants (S243Q and S243H).

4.4.3 Crystallization

The WT ApnA A_1 PCC7821 for the crystallization experiments was purified by Julia Sandberg-Meinhardt (University of Hamburg). The initial crystallization trials were performed using commercial screens and sitting-drop vapor diffusion methods at the EMBL Hamburg high-throughput crystallization facility (Müller-Dieckmann, 2006). The selected commercial screens included PACT Suite, pHClear and pHClear II Suite, Classics I and II Suite, PEGs and PEGs II Suite from Qiagen. Prior to crystallization the protein was concentrated to 10 mg/ml.

Two lead conditions were obtained from the Qiagen PACT suite, one with 25% Polyethylene glycol 1500 and 0.1 M Malonic acid-Imidazole-Boric acid (MIB) buffer at pH 5.0 (Solution no. 14) and the other with 25% Polyethylene glycol 1500 and 0.1 M

Malic acid-MES-TRIS (MMT) buffer at pH 6.0 (Solution no. 39) (Figure 56A and B). Based on these two conditions, manual optimization was performed using hanging-drop vapor diffusion method at 19°C and protein concentrations of 5, 10 and 13 mg/ml. In the manual screens the concentration of the main precipitant PEG1500 was varied between 10 - 30 % in combination with BIS-TRIS, MES and HEPES buffers at pH 5.5, 6.5 and 7.5, respectively. Similar trial was additionally conducted with another PEG compound of larger molecular weight, PEG 3350. For both PEGs a selection of salts, namely $MgCl_2$, $CaCl_2$, $(NH_4)_2SO_4$ and $LiSO_4$, were tested as additives at concentration 0.2 M. The crystals used for data collection and soaking experiments grew in one week in drops containing 24-30% Polyethylene glycol 3350 and 0.1 M MES or BIS-TRIS buffer in pH 5.5 to 6.5 (Figure 56C). These crystals diffracted up to 1.90 Å and belonged to the space group $P2_12_12_1$.

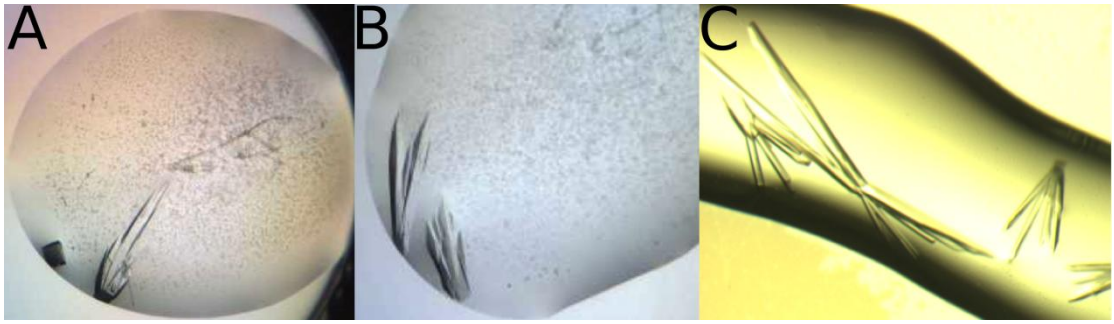

Figure 56. The ApnA A_1 PCC7821 crystals obtained from initial and manual optimization screens. A) Crystals from Qiagen PACT condition no. 14 (25% PEG 1500, 0.1 M MIB Buffer at pH 5.0). B) Crystals from Qiagen PACT condition no. 39 (25% PEG1500, 0.1 M MMT Buffer at pH 6.0). C) The optimized crystals from manual screens using PEG3350 at concentration 24 – 30 % and either 0.1 M MES or BIS-TRIS buffer at pH range 5.5 to 6.5.

4.4.4 Structure determination and validation

The raw data was indexed and integrated using program *XDS* (Kabsch, 2010) and scaled with *SCALA* (Collaborative Computational Project, Number 4, 1994). The data quality statistics are in the appendix. The phases were obtained from automated molecular replacement with program *MOLREP 11.0* (Vagin&Teplyakov, 1997) using the crystal structure of phenylalanine activating subunit of Gramicidin synthetase 1 (PheA, PDB code: 1AMU) as a search model. In each case the asymmetric unit contained one molecule with Matthews coefficient of 2.0 $Å^3$/Dalton and approximate solvent content of 40 %.

The final ApnA A_1 PCC7821 models were obtained through repeated cycles of manual model building in *COOT 0.7* (Emsley *et al.,* 2010) and refinement in *REFMAC5.7* (Murshudov *et al*., 1997). For the complex structures part of the protein was build into the electron density using *ARP/wARP* (Langer *et al*., 2008) and manually modified in *COOT*. The tyrosyl adenylate exists in the PDB with ligand ID YAP but is not readily available in *COOT* monomer library or in the *REFMAC* library. *JLigand* was used to create the adenylated L-tyrosine and the corresponding CIF component definition file using similar compound Phosphoric acid 2-amino-3-(4-hydroxy-phenyl)-propylester adenosin-5'yl ester (PDB Ligand code: TYA) as a template. The arginyl adenylate has not been observed in existing protein structures as ligand according to current knowledge. The program *JLigand* was employed to create a link between L-arginine and AMP monomers using existing restrains from the REFMAC library. The structure composition file from JLigand was then used in both *COOT* and *REFMAC* to fit and refine the created molecule to the electron density in the model. The data collection and refinement statistics are shown in the table 31.

Table 31. Data collection and refinement statistics for ApnAA1 PCC7821 structures. These were solved in absence of substrates (referred as Apo enzyme) and in complex with AMP-PNP (Complex 1), with both AMP-PNP and L-arginine (Complex 2), with arginyl adenylate (Complex 3) and finally with tyrosyl adenylate (Complex 4). Values in the parenthesis are for the high-resolution shell.

	Apo	**Complex 1**	**Complex 2**	**Complex 3**	**Complex 4**
Data collection					
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions					
a, b, c (Å)	68.81, 81.19, 89.59	68.31, 81.81, 89.73	68.43, 82.69, 90.15	68.14, 81.85, 89.65	68.17, 82.24, 90.06
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	19.97–2.02 (2.13–2.02)	89.73–2.250 (2.37–2.25)	90.15–2.00 (2.11–2.00)	89.65–1.90 (2.00–1.90)	90.06–2.00 (2.11–2.00)
R_{merge}	6.9 (65.1)	8.8 (42.6)	6.0 (51.5)	11.1 (114.0)	12.9 (85.3)
$I / \sigma I$	14.1 (2.6)	15.5 (5.7)	21.7 (3.9)	16.6 (2.4)	12.2 (2.2)
Completeness (%)	98.5 (91.4)	99.1 (98.4)	100.0 (100.0)	99.1 (99.0)	100.0 (100.0)
Redundancy	6.5 (5.9)	5.5 (5.7)	6.6 (6.7)	13.2 (13.0)	6.5 (6.6)
Refinement					
Resolution (Å)	19.97-2.02	60.48-2.25	60.94-2.00	60.45-1.90	60.74-2.00
No. reflections	33,052	24,495	35,245	34,891	40,188
R_{work} / R_{free}	20.4/23.6	18.2/23.0	19.0/24.5	18.3/22.7	18.7/23.1
No. atoms					
Protein	3,240	3,825	3,838	3,904	3,845
Ligand/ion	6	75	81	51	51
Water	109	105	145	164	122
B-factors					
Protein	39.4	30.4	35.7	31.7	30.3
Ligand/ion	59.7	38.7	42.9	44.1	41.1
Water	42.0	27.9	33.6	31.4	30.0
r.m.s. deviations					
Bond lengths (Å)	0.01	0.01	0.01	0.02	0.02
Bond angles (°)	1.3	1.5	1.5	2.0	2.0

4.4.5 Structural characterization

4.4.3.1 The crystal structure of the apo ApnA A_1 PCC7821

The first crystal structure for the ApnA A_1 PCC7821 was obtained in the absence of the amino acid and nucleotide substrates, and was therefore referred as the apo form of the enzyme. The apo-ApnA A_1 PCC7821 structure consisted of residues 3 – 410 and

covered mainly the expected N-terminal domain (Figure 57). For the first two N-terminal residues and for the C-terminal residues 411 – 547 no interpretable electron density was observed, which indicated a lack of stabile three-dimensional organization or intristic flexibility in this region. The structure shared the closest resemblence with the N-terminal domain of PheA (residues 17-428) with root mean square deviation (r.m.s.d.) of 0.777 Å for C_{α}-atoms of 283 residues in the structural alignment of the two enzymes. The apo-ApnA A_1 PCC7821 structure exhibited the typical fold of the N-terminal domain of adenylating enzymes, and accordingly comprised three subdomains. Two of these domains consisted of a large β-sheet where most of the β-strands were parallel to each other and connected via α-helix. These two subdomains, referred to as subdomain A and B analogous to PheA, formed together a five-layered αβαβα fold where the two β-sheets enclosed two of the α-helices between them with the remaining helices surrounding this core structure. In the apo-ApnA A_1 PCC7821 structure the β-sheet of the subdomain A was shorter than in the PheA structure due to a missing β-strand corresponding to β-strand A7 in PheA. Therefore the subdomain A of ApnA A_1 PCC7821 consisted only of six β-strands (A1-A6 in PheA). Additionally, the α-helix in the linker region of β-strands A3 and A4 found in PheA does not exist in ApnA A_1 PCC7821. The subdomain B of ApnA A_1 PCC7821 was very similar with PheA, with the exception that N-terminus of this domain contained an additional α-helix in PheA. The third subdomain of ApnA A_1 PCC7821 (residues 328 – 409) was mainly formed by β-strands and had a tertiary fold that resembled a distorted β-barrel. This subdomain was referred to as subdomain C, based on the nomenclature in PheA structure.

The highly conserved phosphate-binding loop (P-loop), which is a common motif found from most ATP- and GTP-binding proteins, was located in the subdomain A. The P-loop has in the adenylation domains of Non-ribosomal peptide synthetases consensus

sequence LAYxxYTSG(ST)TGxPKG (Marahiel *et al*., 1997), which corresponds to the residues 149 – 164 in ApnA A_1 PCC7821. In crystal structures of ATP-binding proteins the whole P-loop or parts of it are occasionally missing, examples of which are the firefly luciferase and PheA structures. In both structures, the central residues of the P-loop have apparent conformational flexibility based on the absence of electron density (Conti *et al*., 1997). In the apo-ApnA A_1 PCC7821 the P-loop also appeared to have flexibility based on the weaker electron density for the two central residues Thr159 and Gly160. The entire P-loop could still be build into the model.

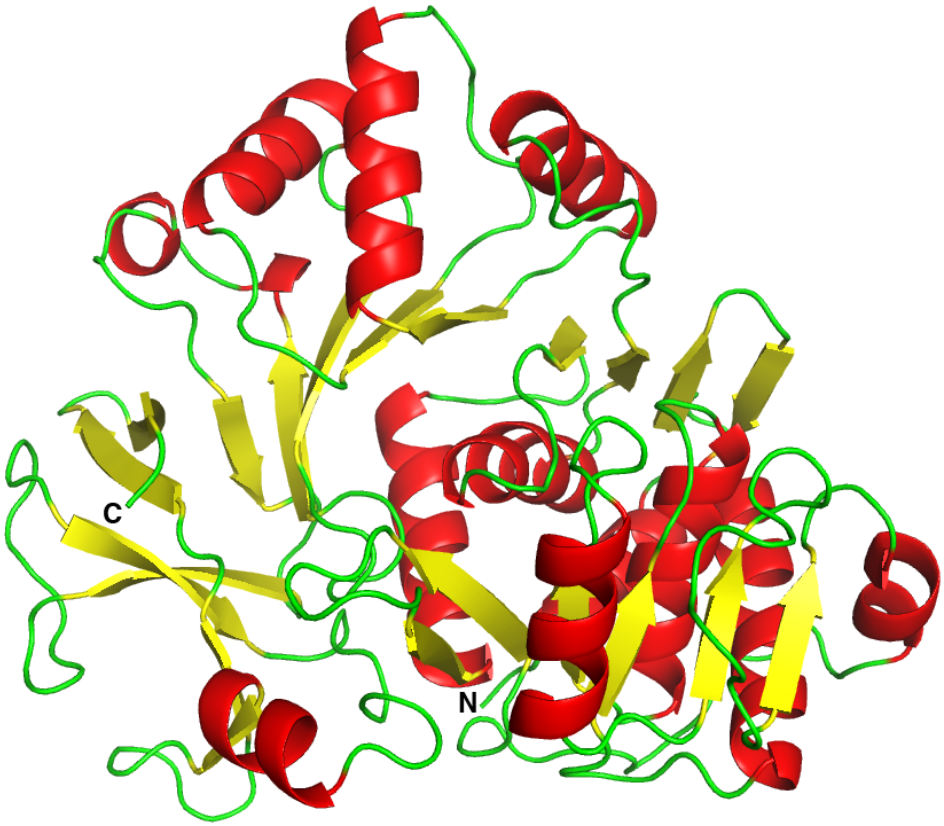

Figure 57. The crystals structure of apo-ApnA A_1 PCC7821. The structural composition of the enzyme has high similarity to the N-terminal domains of other adenylating enzymes. It can be divided into three subdomains, where two subdomains combrise a mix of α-helices and β-sheets and the third a distorded β-barrel core structure. The α-helices are depicted in red, the β-strands in yellow and the loops in green.

4.4.3.2 The ApnA A_1 PCC7821 complex structures

The main aim of the structural studies with ApnA A_1 PCC7821 was to elaborate the bi-specificity of the enzyme. To that end, the crystals of the apo-enzyme were soaked shortly in suitable cryo-solution with the substrates added as described in the materials and methods. The soaking experiments were successful and resulted in four different complex structures. Two of these were showed the incorporation of a non-hydrolyzable ATP analog, AMP-PNP both with and without L-arginine as co-substrate into the active site. These were expected to represent the enzymatic state of ApnA A_1 PCC7821 prior to adenylation. The other two structures were obtained in the presence of the reaction products, adenylated L-tyrosine and adenylated L-arginine, which were derived from the enzymatic activity in the presence of the natural substrates, ATP and the amino acid.

The four ApnA A_1 PCC7821 complex structures showed the conserved domain organization of the adenylating enzymes. They consisted of a large N-terminal domain with three subdomains (residues 2-408) and a significantly smaller C-terminal domain (residues 409-490), which is connected to the N-terminal domain by flexible hinge of four residues (Figure 58). The active site located at the interface of these two domains with the C-terminal domain functioning as lid to close the substrate-binding pocket. When compared to the full-length ApnA A_1 PCC7821, the obtained structure lacked the first N-terminal methionine residue and also about 60 residues (491 – 547) from the C-terminal end. Additionally, no electron density was observed for the residues 445 – 451, which locate in a putative flexible loop between two β-strands in the C-terminal domain.

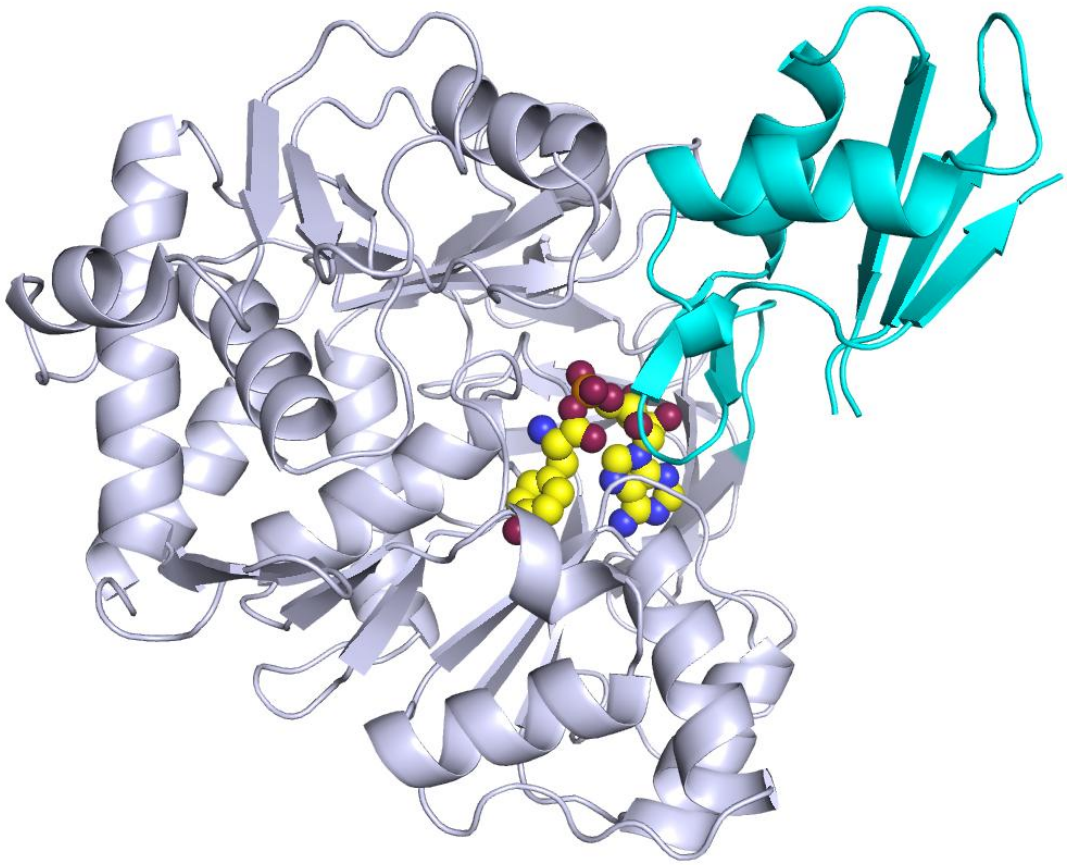

Figure 58. The overall structural composition of the ApnA A_1 PCC7821 in the presence of substrate, shown here with the adenylated L-tyrosine (in yellow) bound in the active site. This structure represents the common fold of adenylating enzymes from same superfamily, consisting of larger N-terminal domain (lilac) and smaller C-terminal one (light blue) linked together by a short flexible loop. The protein is represented as cartoon model with the bound adenylated L-tyrosine depicted as spheres.

The N-terminal domain in the ApnA A_1 PCC7821 complex structures showed the same structural characteristics as described for the apo form of the enzyme. Interestingly, in the substrate bound state a majority of C-terminal domain residues also became structured and could be built into the model. The C-terminal domain had adopted a similar conformation to the C-terminal domain of PheA with a r.ms.d.-value of 0.741 Å for C_α-atoms of 58 residues. It comprised a β-hairpin (residues 413 – 422) followed by a three-stranded β-sheet, where the first two adjacent β-strands are antiparallel and the third

β-strand parallel to its neighbour. Additionally there are two α-helices in the C-terminal domain that pack against the three-stranded β-sheet. The last missing C-terminal residues of the ApnA A_1 PCC7821 model form a long loop in PheA structure that folds against the three-stranded β-sheet.

The comparison of the ApnA A_1 PCC7821 models to the published crystal structure of PheA (Conti *et al*., 1997) revealed that the orientation of the C-terminal domain in respect to the N-terminal one differed in the two proteins (Figure 59). The conformation adopted by ApnA A_1 PCC7821 has been first described for the Acetyl-CoA synthase of *S. enterica* (Gulick *et al*., 2003) and is referred as the thioester-forming conformation. In this conformation the C-terminal domain is rotated about 140° compared to PheA. The rotational movement of the C-terminal domain is mediated by the flexible hinge region between the two domains. The residues forming the hinge are part of a conserved motif, which in adenylation domains has the consensus sequence GRxDxQVKIRGxRIELGEIE (residues 408 – 427 in ApnA A_1 PCC7821). This motif is shared among the superfamily of adenylating enzymes and contains conserved aspartate residue (Asp411 in ApnA A_1 PCC7821) with putative function as the hinge residue (Marahiel *et al*., 1997; Gulick, 2009).

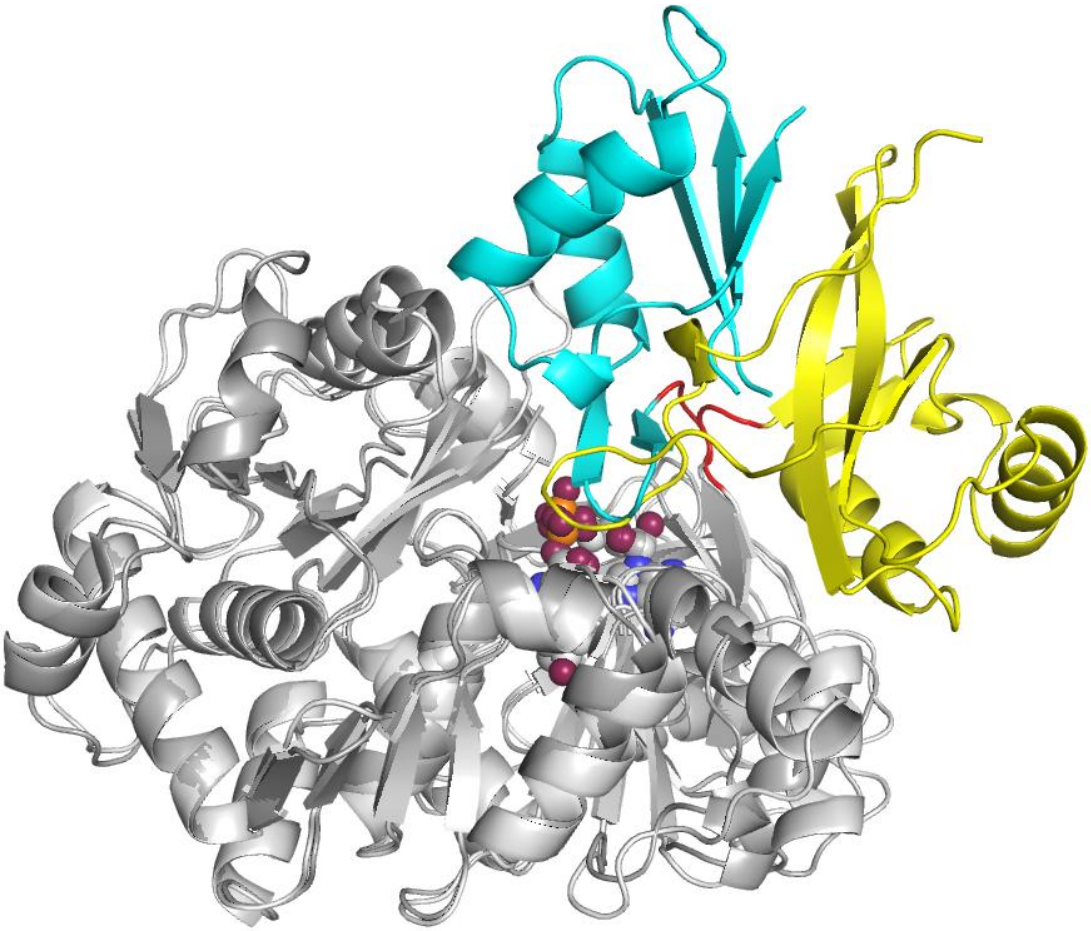

Figure 59. The structural alignment of ApnA A_1 PCC7821 complex structure with the PheA structure. The overlay shows the different orientation of the C-terminal domain in the two enzymes. The C-terminal domain of ApnA A_1 PCC7821 (in light blue) is rotated roughly 140° towards the N-terminal domain compared to the PheA (C-terminal domain depicted in yellow). The hinge region responsible for the rotational movement is highlighted in red.

The active site of ApnA A_1 PCC7821 is located at the cleft between the two domains and as mentioned above is mainly surrounded by the secondary structure elements from the subdomain B and C of the N-terminal domain. The C-terminal domain functions as a lid to the binding pocket and also contributes to the ligand-enzyme interaction network. In the AMP-PNP bound structures a clear electron density is observed for the adenosine unit and for major part of the phosphate moiety. However, the γ-phosphate group of AMP-PNP appeared to fluctuate between two possible

conformations. In one conformation the γ-phosphate group is pointing inside the N-terminal domain, where it interacts with the P-loop. In the other it is facing a cavity, which is connected to the surface of the enzyme. The examination of the binding site in the presence of the amino acid substrate (both in free and adenylated form) revealed that the side chain of this substrate is sandwiched between the α-helical linker of β-strands B3 and B4 from one side, and the β-strands B5 and B6 from the other as seen for phenylalanine in PheA. For the adenylated L-tyrosine and L-arginine, the calculation of omit map results in a uniform density is observed for the amino acid and AMP units thus confirming the presence of the adenylation reaction product (Figure 60A and B, respectively). In the presence of AMP-PNP no link between the amino acid ligand and the nucleotide is observed, as expected.

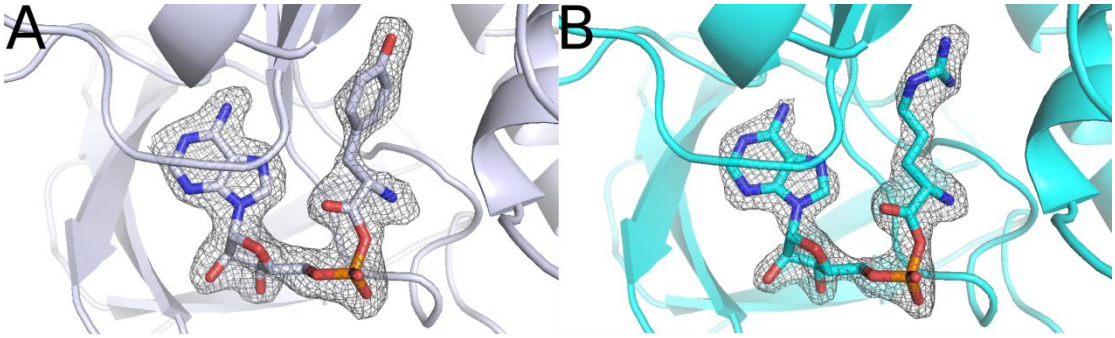

Figure 60. The bound adenylated amino acid at the active site of ApnA A_1 PCC7821. The difference electron density for the adenylate reaction intermediates of ApnA A_1 PCC7821 was calculated by omitting the ligands from the model. A) The map for tyrosyl adenylate intermediate contoured at 0.12464 e*Å^{-3} (σ = 2.0) and B) for the arginyl adenylate at 0.14328 e*Å^{-3} (σ = 3.0).

In order to get an overview of the accessibility of molecules to the active site, the surface models of the four ApnA A_1 PCC7821 complex structures were evaluated. Due to the position of the C-terminal domain in relation to the N-terminal one, the active site appeared to be virtually inaccessible for other than solvent molecules. However, a smaller

putative exit cavity for the pyrophosphate was identified. The presumable role of this cavity as exit route for the phosphates was supported by the observation that two phosphates were modeled in the tyrosyl adenylate structure at the proximity of this cavity. Additionally in one of the two conformations of AMP-PNP, the γ-phophate group points towards this cavity (Figure 61A and B). In this particular conformation the γ-phosphate group is also in the proximity of the carboxyl acid group of the protein when both AMP-PNP and L-Arginine are present in the active site. It was also noted that the size cavity of the putative phosphate exit cavity appeared to be larger in the structure with AMP-PNP bound than in the adenylate containing structure. A closer examination of the residues lining this cavity revelaed that two phenylalanine residues, Phe199 and Phe419 appeared to play an important role in modifying its size. These phenylalanines had a different conformation in the pre-reaction state of the enzyme (AMP-PNP bound) in comparison to the post-reaction state with adenylated amino acid bound inside (Figure 61C).

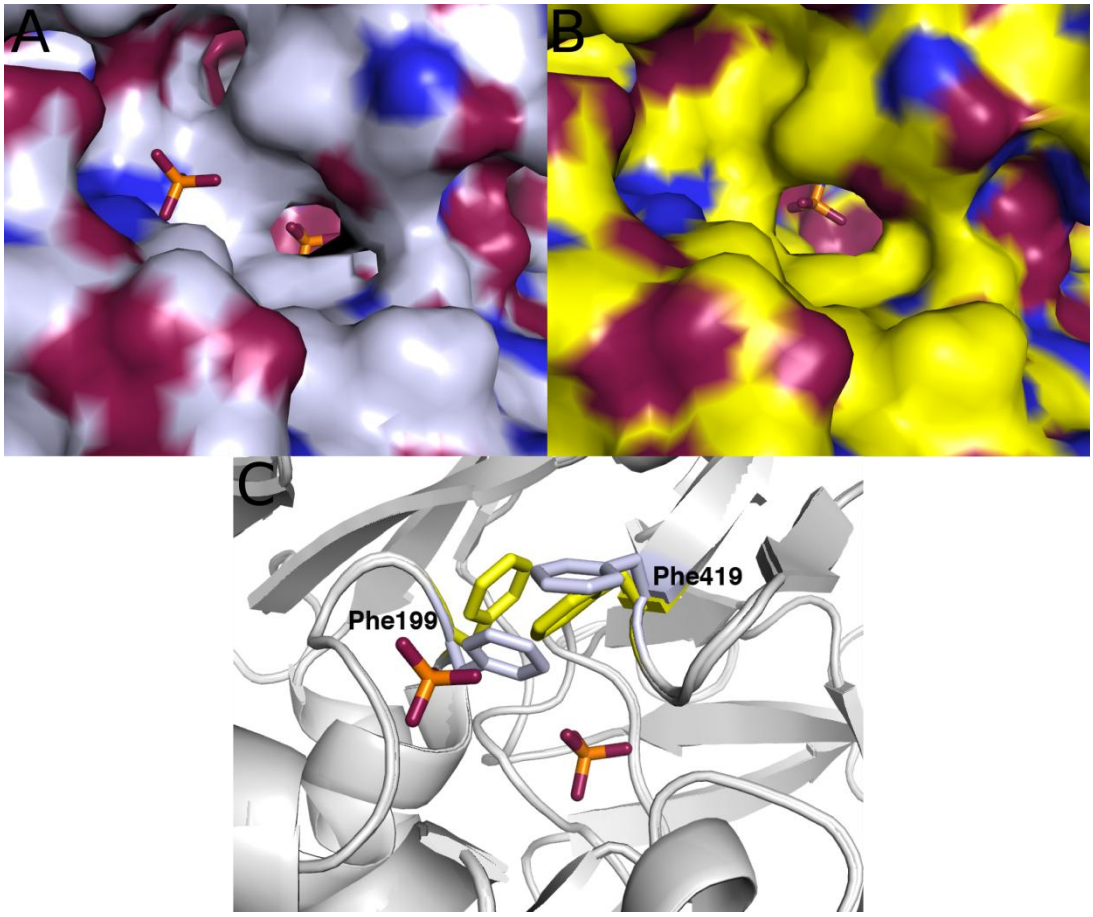

Figure 61. The putative exit cavity of ApnA A_1 PCC7821. A) – B) The surface models of the ApnA A_1 PCC7821 in complex with adenylated amino acid (in lilac) and in complex with AMP-PNP (in yellow) revealed a putative phosphate exit cavity. The size of this cavity is larger prior to the completion of the adenylation reaction based on the comparison of the two models. C) Two phenylalanine residues, Phe199 and Phe419 appeared to play a role in modifying the size of the cavity.

4.4.3.3 Nucleotide binding site

The active site of ApnA A_1 PCC7821 located at the interface of the N-terminal and C-terminal domains as observed for other adenylating enzymes with known three-dimensional structure. The nucleotide-binding site lies in the cleft between the N-terminal subdomains B and C with the ribose ring facing the β-hairpin of the C-terminal domain. Subsequently majority of the residues involved in the binding of the nucleotide arise from

the N-terminal domain. The four complex structures of ApnA A_1 PCC7821 were compared to identify residues that retain their conformation and interaction with the nucleotide substrate, independent of its existence as free molecule or as part of the adenylation product. These residues were considered to be important for the correct docking of the nucleotide substrate and included both amino acids with charged side chains and with hydrophobic or polar side chains (Figure 62).

The adenine ring is lined from one side by the main chain atoms of residues 274 – 276 whereas on the opposite site it is packed agaist the side chain of Tyr299. The bottom of the nucleotide-binding pocket is capped by the side chains of Ile328 and Tyr406. The Tyr299 is also forms a hydrogen bond 2'OH-group of the ribose unit through a highly conserved aspartate residue (Asp394). This aspartate is part of a conserved motif (S/T)GD that is shared among the members of the adenylating enzyme superfamily. It has a specific role in stabilizing the C3' endo conformation of the ribose unit by forming hydrogen bonds to its hydroxyl groups. These interactions occur between the OD1 and OD2 of Asp394 and 2' and 3' hydroxyl groups of the ribose ring as observed for the corresponding residue Asp413 in PheA. Other residues that are involved in the main hydrogen bond network in the nucleotide-binding pocket of ApnA A_1 PCC7821 include Asn297, Val298 and Thr302. Out of these Asn297 and Thr302 also seemed to have conserved role in the nucleotide binding since their position and coordination patterns overlap with their PheA counterparts, Asn321 and Thr326. Asn297 forms a hydrogen bond to the exocyclic aminogroup of the adenine ring through its OD1 atom and possible through this interaction enables the differentiation between adenine and guanine. Val298 also has a part in this hydrogen bond network through its main chain carbonyl oxygen.

Although most of the interactions with the nucleotide arise from the N-terminal domain, a few important residues from the C-terminal domain also contribute to the

nucleotide binding. A lysine residue (Lys415) from the C-terminal β-hairpin was found to take a central position as it formed hydrogen bonds both with the amino acid and the nucleotide. The position of Lys415 is invariant in the four ApnA A_1 PCC7821 complex structures. However, its interaction network varied slightly depending on whether the adenylation had occurred or not. In the obtained ApnA A_1 PCC7821 complex structures, Lys415 formed hydrogen bonds to O4' and O5' of the nucleotide. In adenylate bound structures there were additional hydrogen bonds to the O1P atom of the phosphate group and to the carbonyl oxygen of the amino acid. While the latter one also exist in the complex model with AMP-PNP and L-arginine, the former one is replace with hydrogen bond between Lys415 and the β-phosphate group when AMP-PNP is bound to the active site.

In addition to Lys415, two other residues, Arg409 and Arg420, from the C-terminal domain appeared to have a role in the nucleotide binding. The former participated in the hydrogen bond network via interaction with O2 of the α-phosphate group. Arg420 on the other hand interacted with the one of the α-phosphate oxygens and also forms a salt bridge with the Glu303.

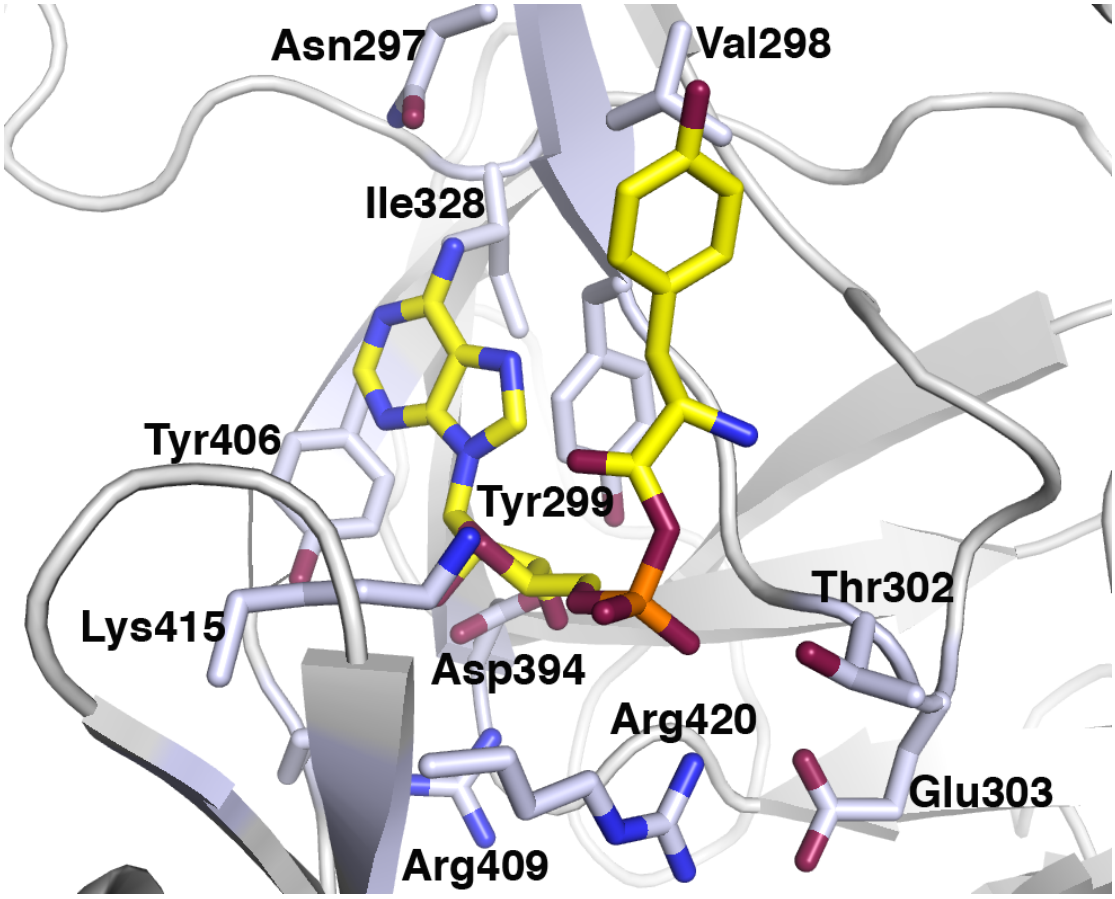

Figure 62. The residues involved in the nucleotide binding in ApnA A_1 PCC7821. Most of these residues are highly conserved and form specific interactions with the bound substrate. Their location in the active site is shown here in the presence of tyrosyl adenylate (in yellow). Both the tyrosyl adenylate and the active site residues are shown as sticks. The active site residues are coloured in lilac while the rest of the protein is shown in light gray (cartoon representation).

4.4.3.4 Amino acid binding pocket

The ApnA A_1 structures with amino acid substrate bound, both in adenylated form (L-arginine and L-tyrosine) and as free amino acid (only L-arginine) were used to identify the residues that have specific interaction with the substrate and thus contribute the most to the binding specificity of the enzyme. The majority of the residues that surround the amino acid in the active site of ApnA A_1 PCC7821 appeared to be hydrophobic or polar in nature. These were postulated to have a role in modifying the size of the binding

pocket since they interacted with the substrate mainly through hydrophobic interactions. Two valine residues, Val201 and Val298 lie in opposite sides of the active site with the side chain of the amino acid substrate sandwiched between them. Similarly, the side chain of Ala307 and the main chain atoms of Ile272 and Gly273 were positioned on the opposite faces of the amino acid binding pocket. The binding pocket is capped by the side chains of Ile271 and Tyr183. Tyr183 is also involved in the hydrogen bond network with the substrate through Glu204 to which it is coordinated via its OH-group. Two conserved residues Asp200 and Gly300 are coordinated to the α-aminogroup of the substrate in a similar manner than Asp235 and Gly324 in PheA. Asp200 forms hydrogen bond to the α-aminogroup through its side chain carboxyl oxygen OD1 whereas Gly300 interacts with the substrate via its main chain carbonyl oxygen. Also Ile306 participates to this network by forming a hydrogen bond to the α-aminogroup via its main chain carbonyl oxygen. In order to form this interaction, Ile306 has adopted an unfavorable conformation based on the dihedral angles that are outside the allowed region in Ramachandran plot (Ramakrishnan&Ramachandran, 1965). In tyrosyl adenylate, arginyl adenylate and in the AMP-PNP containing structures with and without the L-arginine, the corresponding ϕ angles for the Ile306 are 85°, 87°, 84° and 84°, and Ψ angles -63°, -59°, -69° and -65°, respectively. Same phenomenon was also observed for the corresponding isoleusine (Ile330) in PheA structure, which indicates that this residue has an important role in the amino acid binding. Finally, two residues lying at the bottom of the binding pocket were observed to have specific interaction with the side chain of the amino acid substrate. These glutamate (Glu204) and serine (Ser243) residues were both coordinated to the substrate through their side chain but had variations in the coordination chemistry depending on the bound substrate and also whether the adenylation reaction had occurred or not. Details of this will be described further. Figure 63 shows the composition of the

amino acid binding pocket in ApnA A_1 PCC7821 with the tyrosyl adenylate complex structure as an example.

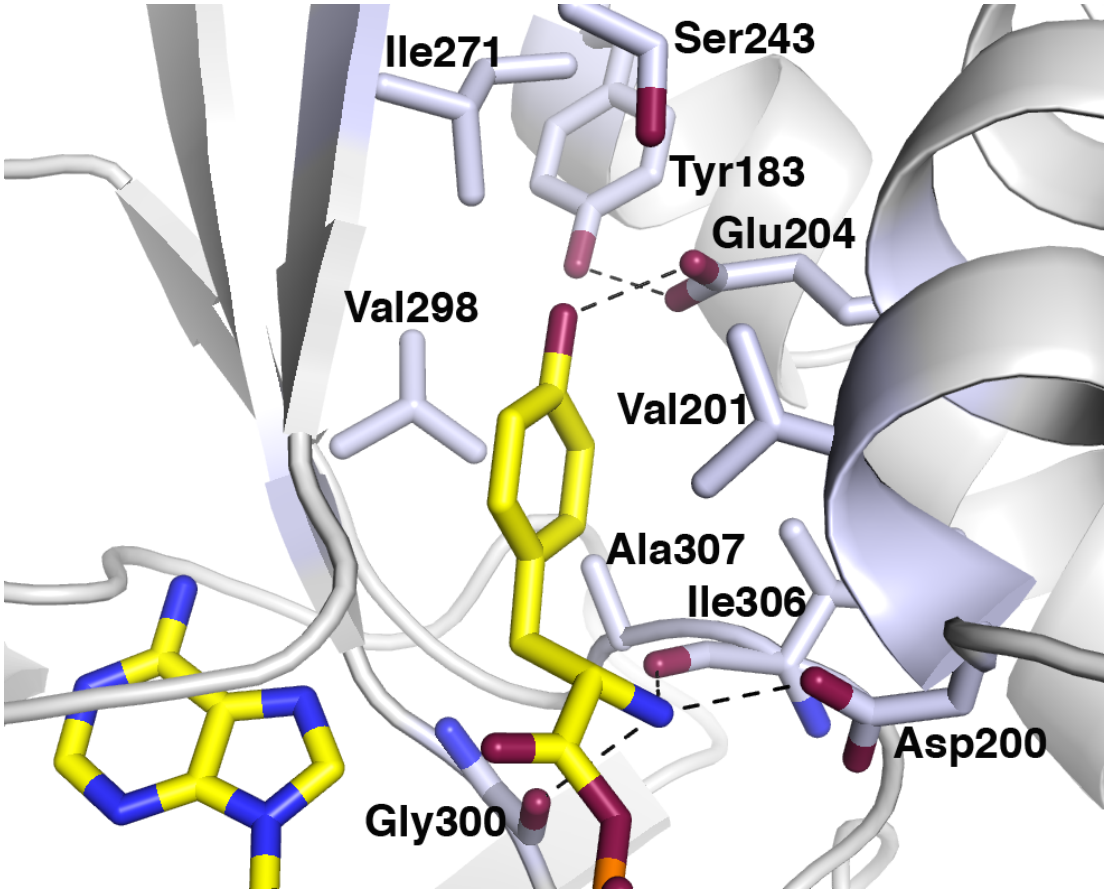

Figure 63. The amino acid binding pocket of ApnA A_1 PCC7821. The residues that in ApnA A_1 PCC7821 are involved in forming the amino acid binding pocket are shown here in the presence of tyrosyl adenylate (in yellow). A group of hydrophobic residues, namely Val201, Ile271, Val298 and Ala307 spread around the side chain of the amino acid substrate and putatively affect on the shape of the pocket. Asp200 and Gly300 are two conserved residues that interact with the α-aminogroup of the bound ligand. Two residues, Ser243 and Glu204 appear to have more unique role in the amino acid binding and in case of the tyrosyl adenylate, form hydrogen bonds to the OH-group of the tyrosyl side chain.

The recognition of the correct substrate lies in the residues surrounding it. The comparison of the three ApnA A_1 PCC7821 structures representing the binding of the amino acid prior to the adenylation reaction (for L-arginine) and after it (for both L-

arginine and L-tyrosine) two residues with putative role in recognition of the substrate were identified. These residues, Glu204 and Ser243 are located at the bottom of the amino acid binding pocket and appear to function as docking residues for the substrate side chain, Glu204 through its side chain carboxylic acid group and Ser243 through its hydroxyl group (Figure 64). Their interaction with the substrate appears to be dependent on both the nature of the substrate but also on the reaction state. In the case of Glu204 this difference arises form the rotation of its side chain carboxylic acid group whereas with Ser243 the changes in the side chain conformation of the arginine substrate in two separate states mediates the change in the coordination chemistry.

The most apparent difference in the coordination of Glu204 with the substrate is seen with the binding of the adenylated amino acid. In the presence of tyrosyl adenylate, the carboxylate group of Glu204 is rotated so that it forms hydrogen bond to the OH-group of tyrosine through its OE1 oxygen (Figure 64A). In this conformation Glu204 also forms a hydrogen bond to nearby tyrosine residue (Tyr183). However when arginyl adenylate is bound to the active site, the carboxyl group of Glu204 is rotated to enable the formation of a salt bridge with the guanidium group of the arginyl adenylate (Figure 64B). This interaction compensates the positive charge of the arginine substrate in otherwise relatively hydrophobic/polar environment, which is more adopted for tyrosine binding. It also abolishes the hydrogen bond between Tyr183 and Glu204. Surprisingly, the salt brigde does not seem to exist prior to the adenylation reaction as was observed for the L-arginine in the presence of AMP-PNP (Figure 64C). Instead the arginine substrate forms a hydrogen bond to the OE1 oxygen of Glu204 through its NH2 atom hydrogen. The observed change in the binding pattern between adenylated and free L-arginine results from the difference in the adopted rotamer in the two. Prior to the adenylation reaction an elongated conformation is observed for the arginine substrate allowing it to

protrude deeper into the binding pocket. Interestingly upon the formation of the arginyl adenylate the side chain of the amino acid substrate adopts a conformation that mimics the shape of the tyrosine side chain. In this conformation similar binding pattern with Ser243 is observed for both the arginyl adenylate and for the tyrosyl adenylate. The interaction with Ser243 occurs via water molecule through which it forms hydrogen bonds to the hydroxyl group of tyrosine and NH1 of arginine in the tyrosyl adenylate and arginyl adenylate structures, respectively. In the arginine and AMP-PNP containing complex the Ser243 forms a direct, though rather long (3.2 Å) hydrogen bond to NH2 of arginine.

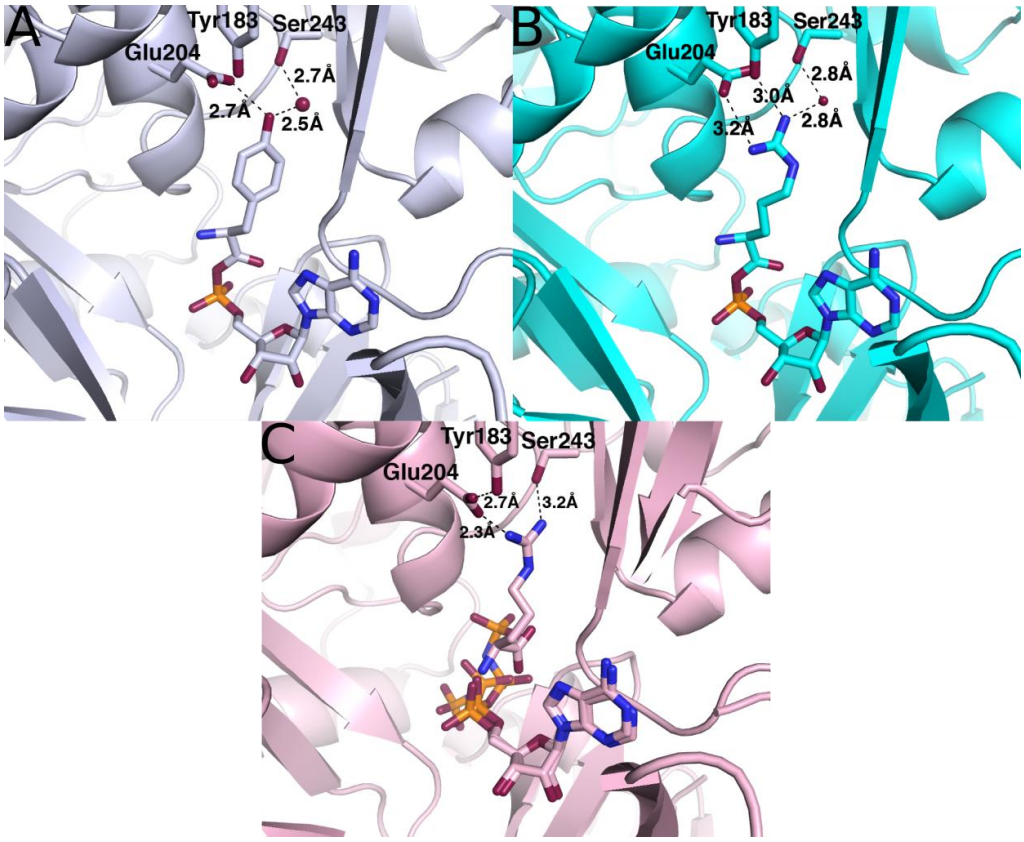

Figure 64. The key residues involved in the amino acid recognition and binding in ApnA A_1 PCC7821. A) – C) The binding network of Glu204 and Ser243 in the presence of the tyrosyl adenylate, arginyl adenylate and free arginine are shown. A nearby tyrosine residue (Tyr183) is also participating in the hydrogen bond network through interaction with Glu204.

In order to find residues that could play a role in the amino acid binding activity, the structure of ApnA A_1 PCC7821 was aligned with selected A-domain or A-domain like structures with both amino acid substrate and nucleotide bound to the active site. The selected structures were for D-Alanyl Carrier Protein DltA from *B. cereus* (PDB code: 3DHV, Du *et al*. 2008), for NRPS adenylation protein CytC1 from *Streptomyces* (PDB code: 3VNS, Ueki *et al.* publication not available) and for PheA. The DltA and CytC1 structures Special focus was paid on residues that interacted with the side chain of the bound target of adenylation. As a result two non-conserved residues, which correspond to Val201 and Ala307 in ApnA A_1 PCC7821 and flank the binding pocket, were noticed to have some correlation with the size of the substrate. In PheA that like ApnA A_1 PCC7821 has a substrate with a large side chain (phenylalanine) these residues were Ala236 and Cys331, i.e. residues with relatively compact side chain. On the other hand in the DltA, the amino acid binding pocket size was reduced by the replacement of leucine (Leu198) in the same position with Val201. However, the residue corresponding to Ala307 was also alanine (Ala302) in DltA. The most significant difference was in the structure of the D-Valine binding NRPS adenylation protein CytC1 where the active site size was noticeably reduced by the two flanking side chains of Phe209 and Phe315 corresponding to Val201 and Ala307, respectively. The overlay of the Val201 and Ala307 with the corresponding residues in DltA, CytC1 and PheA are shown in figure 65.

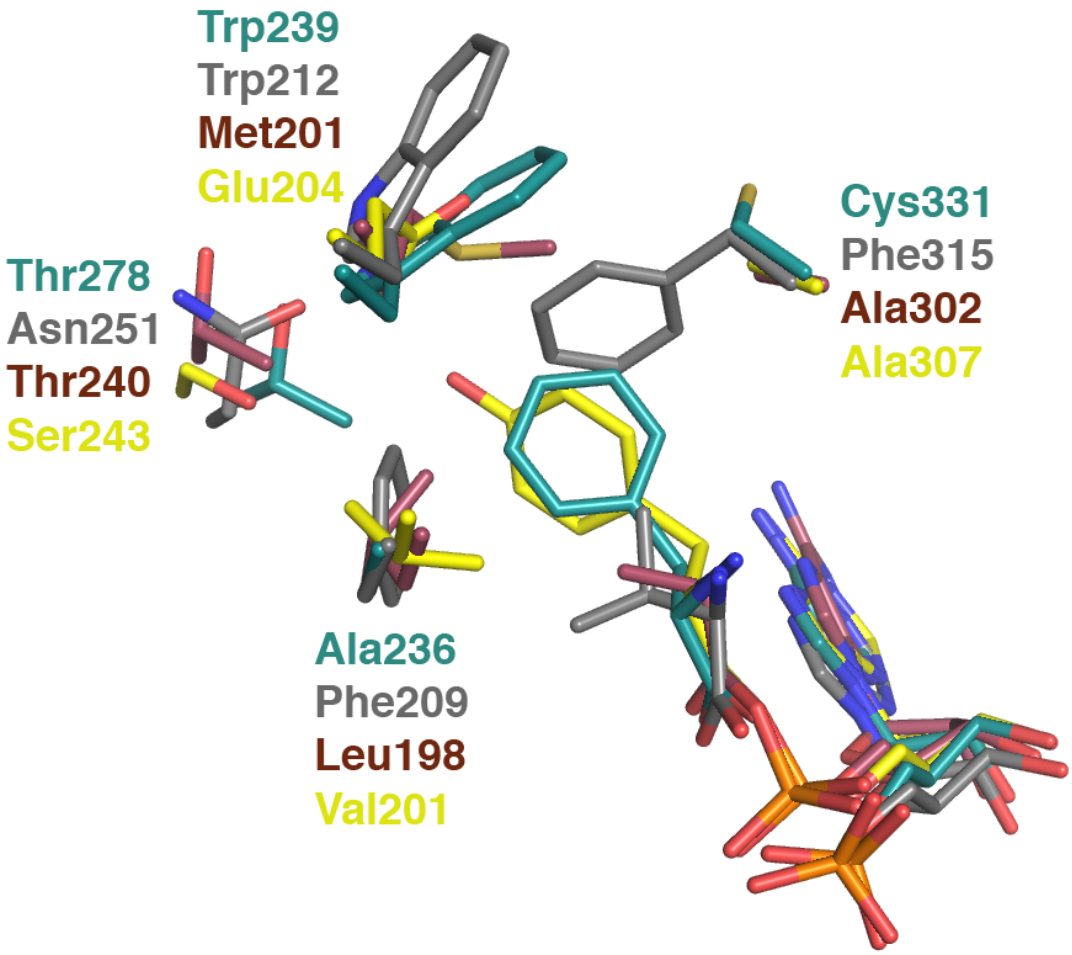

Figure 65. The structural alignment of selected four active site residues from ApnA A_1 PCC7821 (in yellow), D-Alanyl Carrier protein DltA (in brown, PDB code: 3DHV), NRPS adenylation protein CytC1 (in grey, PDB code: 3VNS) and PheA (in dark green, PDB code: 1AMU). The size of the residues at the positions 201 and 307 (ApnA A_1 PCC7821 numbering) appear to correlate with the size of the substrate. The residues and the bound substrates are shown as stick representation.

4.4.6 Activity assays for the ApnA A_1 PCC7821 mutants

4.4.6.1 Enzymatic activity with the proteinogenic amino acid substrates

The wild type ApnA A_1 from *P. agardhii* strain PCC7821 is bi-specific enzyme that uses both L-tyrosine and L-arginine as substrate. The analysis of the ApnA A_1 PCC7821 complex structures indicated that a serine (Ser243) and a glutamate residue

(Glu204) have an important role in defining the substrate specificity of the enzyme. Additionally an alanine residue (Ala307) was suggested to have a suitable position in the active site to create mutants with specificity towards smaller substrates as illustrated in the figure. The role of the residues at positions 243 and 307 in determination of substrate specificity had previously indicated when ApnA A_1 enzymes from different Planktothrix strains were characterized phylogenetically and biochemically (Christiansen *et al*., 2011). The authors of the study noted that enzymes with acidic residues at the positions 243 and 307 were explicitly using basic amino acid as substrate.

As an attempt to change the substrate specificity of ApnA A_1 PCC7821, several variants were made by substituting the active site residue at position 204, 243 or 307 with another amino acid. It was hypothesized that replacement of the residue either in position 243 or 307 with an acidic one would make the enzyme more suitable for basic amino acids and vice versa. As Glu204 was expected to be particularly important for the arginine binding, it was replaced with the small glycine to see if this would abolish the enzymatic activity for the arginine substrate. The ApnA A_1 PCC7821 mutants were then tested against the 20 canonical amino acids using the hydroxamine-trapping assay as described in the methods section. This assay has been originally developed to study the adenylation activity of NIS synthetases (Kadi & Challis, 2009) and was adopted to test the activity of the ApnA A_1 PCC7821 variants. The results of the assay are summarized in figure 66. The significance threshold for the measurements was calculated based on the absorbance value and corresponding standard deviation from the measurements with the wild type enzyme in absence of substrates. This resulted in the significance threshold value of 0.08 in the activity assays with the proteinogenic amino acids. The absorbance values in the final graphs are the average from five replicates for wild type enzyme while the values for the variants were derived from duplicate measurements.

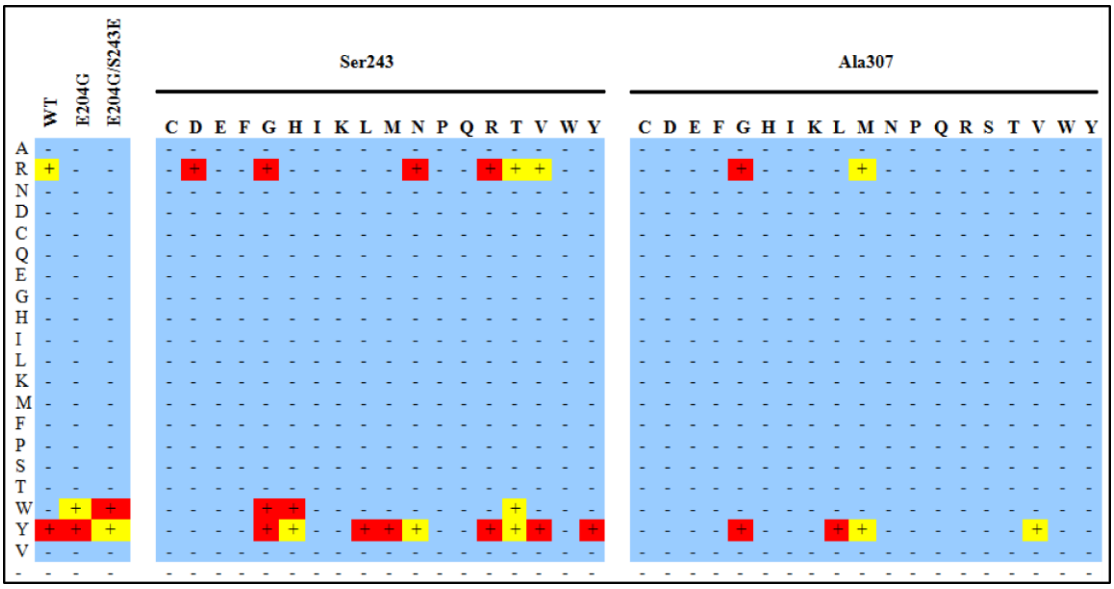

Figure 66. The summary of the hydroxylamine-trapping assay results for the wild type and the variants of ApnA A_1 PCC7821. The left bar indicates the tested amino acids, the upper bar the tested enzyme. The potential substrates are marked with + sign. Additionally the substrate with highest activity is marked with red colour whereas the substrate with detectable but lower activity is marked with yellow colour. Blue colour and – sign indicate very low or non-detectable activity.

The results for the wild type ApnA A_1 PCC7821 are shown in figure 67A. Similarly, the A307G, S243V and S243R variants showed activity only towards the two natural substrates, L-arginine and L-tyrosine (Figure 67B – D). The wild type behavior of the A307G variant is not suprising since the introduced mutation mainly provides more space in the active site rather than additional interaction partner for possible new substrates. However, the replacement of Ser243 with a large and the most bacic amino acid, arginine, was expected to have an effect on the catalytic activity. As it created an additional positive charge in the active site, it was postulated to drive the catalytic activity of the enzyme towards negatively charged residues or possibly inhibit the binding of the arginine substrate. One possible explanation for the observed wild type like behavior is the elimination of the added charge by a nearby glutamate residue (Glu205), which

locates near the putative catalytically important glutamate (Glu204). However, without structural information for the S243R variant, the reason for the observation remains speculative.

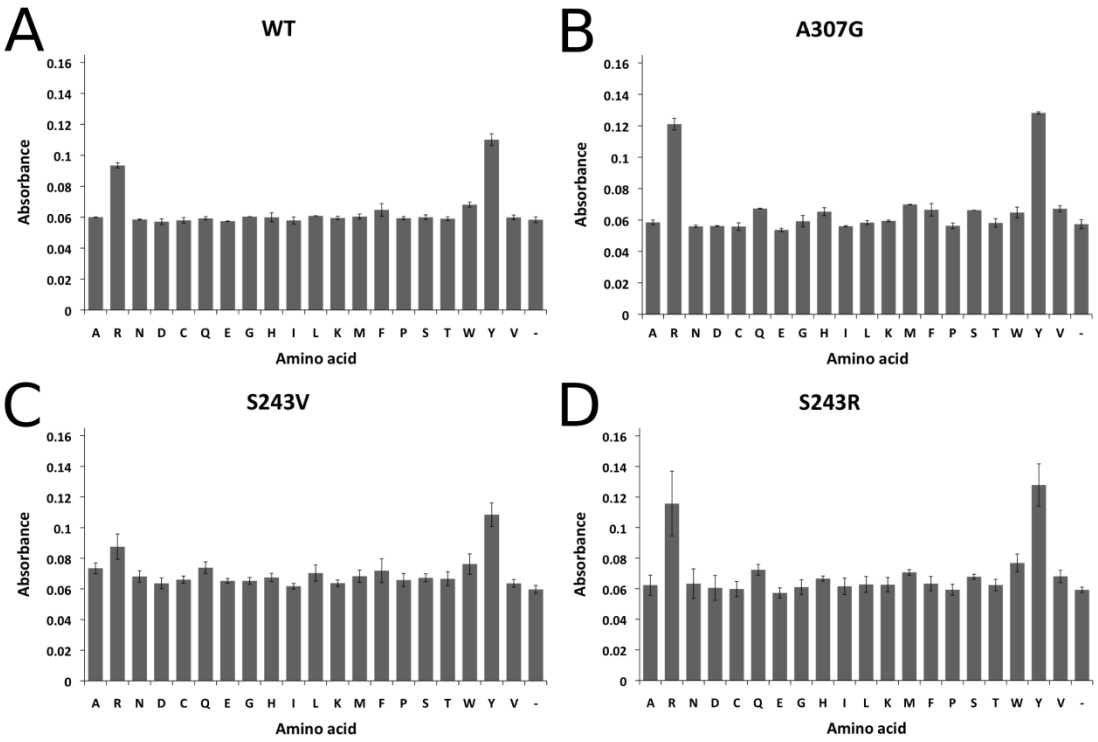

Figure 67. The results of the hydroxylamine-trapping assay for the wild type ApnA A_1 PCC7821 and the wild type like variants. A) As expected the wild type enzyme activated only L-arginine and L-tyrosine. B) - D) Similarly, three ApnA A_1 PCC7821 variants, A307G, S243V and S243R were observed to have WT-like behaviour and use L-tyrosine and L-arginine as substrates.

Several ApnA A_1 PCC7821 variants were found to be inactive or were considered as inactive due to a significantly reduced enzymatic activity. Especially the substitutions to position 307 often resulted in loss of enzymatic activity. These included the A307E variant, which was expected to prefer more basic amino acids, such as arginine and lysine as substrate based on the observations on the ApnA A_1 enzymes in other *Planktothrix* strains (Figure 68A). Also the substitution of Ala307 with a larger polar residue did not

make the enzyme more prone to use smaller polar substrates as was observed with the A307F variant (Figure 68B). A possible explanation could be the loss of the correct protein folding or creation of environment that abolishes the substrate binding. The results for the rest of the A307 variants are shown in appendix 5. In addition to the number of Ala307 mutants, six of the tested Ser243 variants appeared as inactive. The results for S243E and S243K are shown as example in the figure 68C and D, respectively. A possible explanation for the inactivity of the former variant might be in the repulsion between the mutated residue and the two nearby glutamate residues (Glu204 and Glu205), which disables the folding of the protein. The abolished activity of the S243K variant was surprising as the similar S243R variant behaved like the wild type enzyme. The results for the four other inactive variants (S243C, S243I, S243Q and S243W) are presented in the appendix 5.

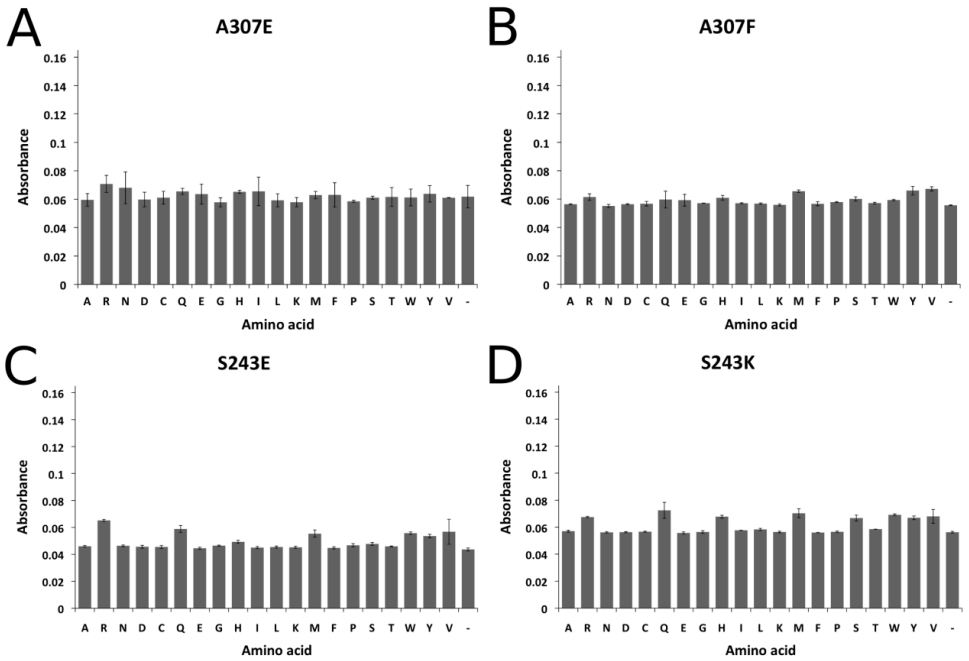

Figure 68. Inactive variants of ApnA A_1 PCC7821. The substitutions to position 307 and in some case to position 243 in ApnA A1 resulted in the loss of the enzymatic activity according to the hydroxylamine-trapping assay. The results for the A307E, A307F, S243E and S243K variants are shown as example.

Interestingly, several ApnA A1 mutants showed altered substrate specificity according to the hydroxylamine-trapping assay. For a majority of these variants, the mutation had resulted to a shift from bi-specific to monospecific towards one of the original substrates, i.e. either towards L-tyrosine or L-arginine. From the tested variants, S243D and S243N appeared to prefer L-arginine as substrate (Figure 69A and B). This observation was in line with the initial hypothesis that more acidic substrate-binding pocket would be better adapted for basic amino acid substrates. The S243D variant also showed residual activity towards lysine, which further supported the theory. Why S243D variant was active and similar S243E variant was not, could be due to the longer side chain of glutamate when compared to aspartate. The one carbon difference is possibly enough to avoid the repulsion between the closely spaced negatively charged side chains.

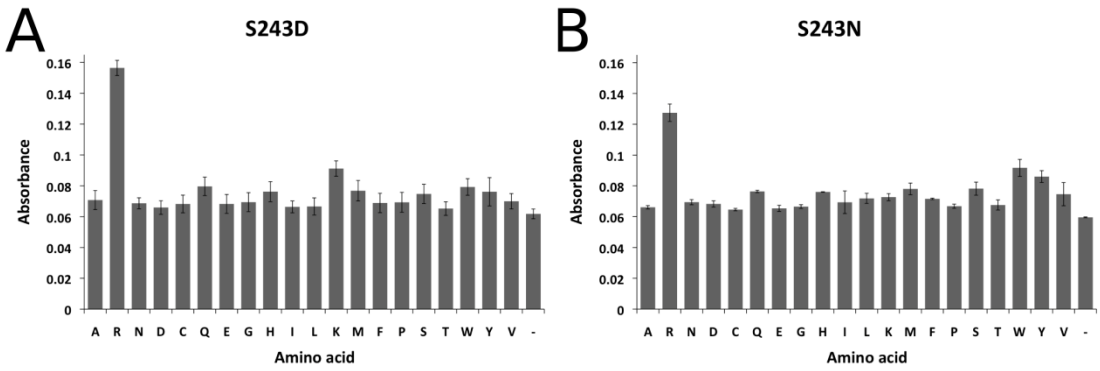

Figure 69. The hydroxylamine-trapping assay results for the S243D and S243N variants with specificity towards L-arginine. The S243D mutant appears to have residual activity for L-lysine as well, which is in line with the original hypothesis that added negative charge would create an enzyme more suitable for positively charged amino acids.

The monospecificity to L-arginine is not unusual since it has been observed with ApnA A_1 enzymes from other Planktothrix strains as indicated in the study by Christiansen *et al*. (2011). However, an ApnA A_1 enzyme that would explicitly use L-tyrosine as substrate has not been isolated so far according to the current knowledge.

Interestingly, a majority of the active ApnA A_1 PCC7821 variants activated solely L-tyrosine. These included A307V and A307L from the Ala307 variants and S243L, S243M, S243P and S243Y from the Ser243 variants (Figure 70A – F). The common factor for these monospecific variants was that the introduced mutation increased the hydrophobic surface in the amino acid binding pocket. This presumably provides a more optimal enviroment for the aromatic side chain of L-tyrosine and thus enhanced the enzymatic activity for this substrate.

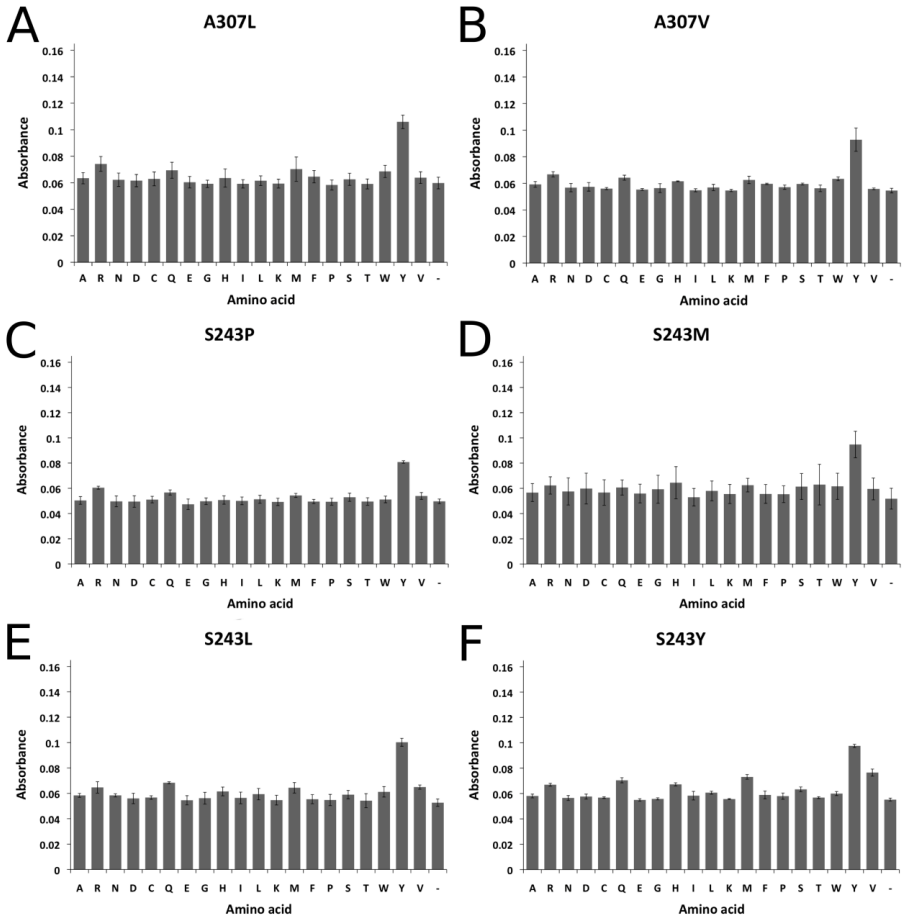

Figure 70. The ApnA A_1 PCC7821 variants that preferred L-tyrosine as a substrate according to the hydroxylamine-trapping assay. A) – B) The substitution of Ala307 with slightly larger aliphatic amino acid provided additional hydrophobic surface in the active site and thus more favorable enviroment for L-tyrosine substrate. C) – F) Similarly the introduced hydrophobic or polar residue with larger side chain at position 243 appeared to favor L-tyrosine activation.

The most interesting group of ApnA A_1 PCC7821 variants appeared to have acquired capability to activate an amino acid other than L-tyrosine or L-Arginine. Some still possessed WT-like activity as well but a few clearly preferred the new amino acid substrate. The mutants of the former type included S243G, S243T and E204G. The two serine mutants appeared to be able to adenylate in addition to the wild type substrates also L-tryptophan (Figure 71A and B). The E204G variant differed from S243G and S243T in that it appeared to be unable to activate L-arginine (Figure 71C). The most plausible explanation for this observation was the loss of the important salt bridge between the L-arginine substrate and the glutamate residue. Interestingly, in one of the ApnA A_1 PCC7821 variants the introduced mutation resulted in loss of wild type activity and instead L-tryptophan was preferred as a substrate. In this variant the active site serine (Ser243) had been substituted with histidine (Figure 71D). The change in the specificity could be exlained through favorable interaction between the histidine and the substrate via ring stacking and specific hydrogen bonds. However, without structural information the reason for the observed activity can only be speculated.

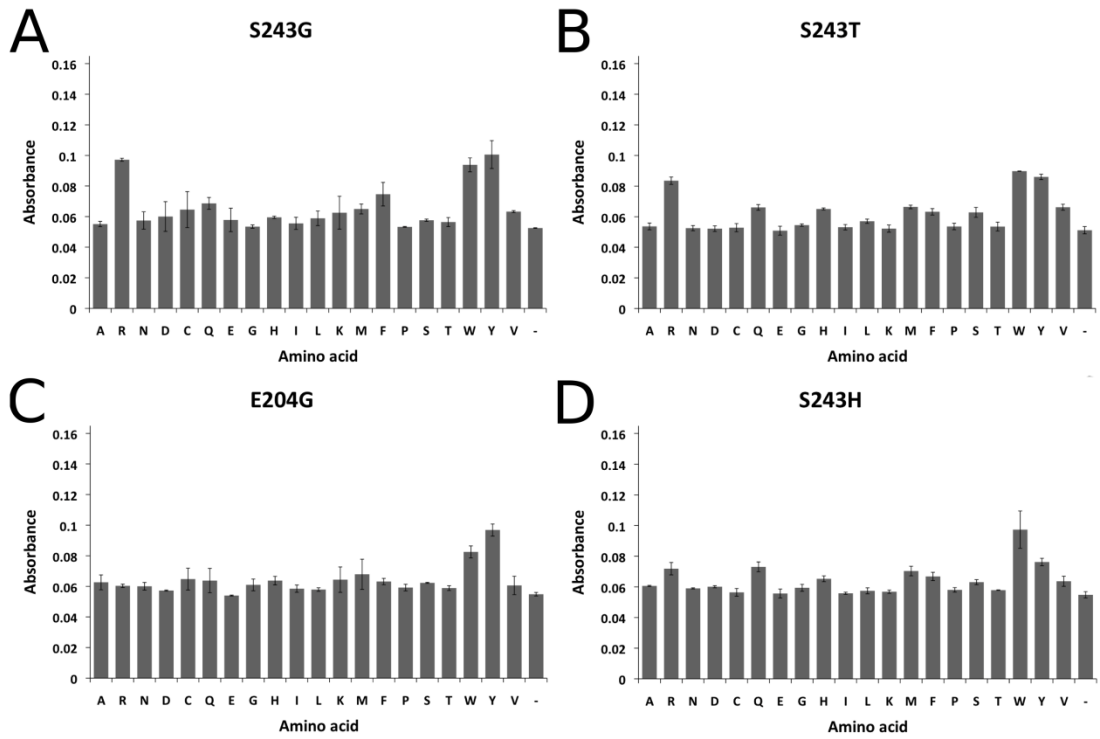

Figure 71. The ApnA A_1 PCC7821 variants that were observed to activate L-tryptophan in addition or instead of the native substrates, L-arginine and L-tyrosine.

As described earlier, the S243E variant was found to be inactive althought based on initial predictions it was assumed to activate basic amino acids, i.e. L-arginine and L-lysine. The inactivity was proposed to result from the repulsion between the introduced negatively charged glutamate and the nearby active site glutmates (Glu204 and Glu205). Threfore a double mutant, where Glu204 was substituted with a glycine and Ser243 with a glutamate, was made. Surprisingly, the double mutant was found to activate mainly L-tryptophan (Figure 72). A partial explanation for the observed behaviour could be an increase in the size of the binding pocket due to the replacement of the glutamate residue with the small glycine.

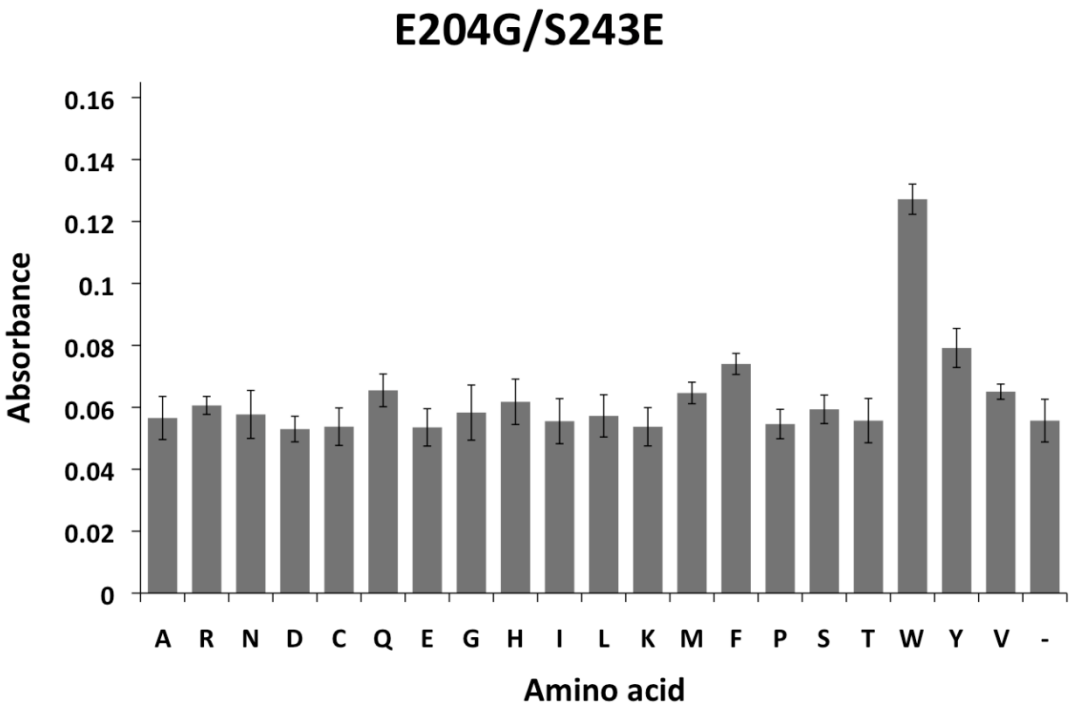

Figure 72. The hydroxylamine-trapping assay results for the E204G/S243E double mutant of ApnA A_1 PCC7821. The double variant appeared to activate explicitly L-tryptophan as shown here.

4.4.6.2 Unnatural amino acids as substrate candidates

The conducted hydroxylamine-trapping assays for the ApnA A_1 PCC7821 variants showed that the substrate specificity of the enzyme could be altered. In few cases the ApnA A_1 PCC7821 variant was able to adenylate a new substrate, namely L-tryptophan either in addition to the native substrates or as the only substrate. However, the activity of the mutants was tested only for the canonical proteinogenic amino acids. As the A-domains of NRPSs are known to use also non-proteinogenic residues as substrates, it was interesting to investigate whether the specific ApnA A_1 PCC7821 variants with altered specificity would be able to use unnatural amino acids as substrate. Two unnatural amino acids, 4-Azido-phenylalanine (Az) and 4-Fluoro-phenylalanine (Fl) (Figure 73A and B) were chosen since they possess an aromatic ring and thus shared

resemblance with the natural substrates of the studied enzyme. The ApnA A_1 PCC7821 mutants chosen for testing were E204G, E204G/S243E, S243H, S243T and S243N. The first four were capable of adenylating L-tryptophan, whereas S243N was monospecific for L-arginine. The WT enzyme was also tested for comparison and to define the significance threshold for these measurements. The calculated significance threshold value was 0.06. The WT ApnA A_1 PCC7821 and the chosen variants were also tested against L-tyrosine, L-arginine, L-tryptophan and L-glycine as a control.

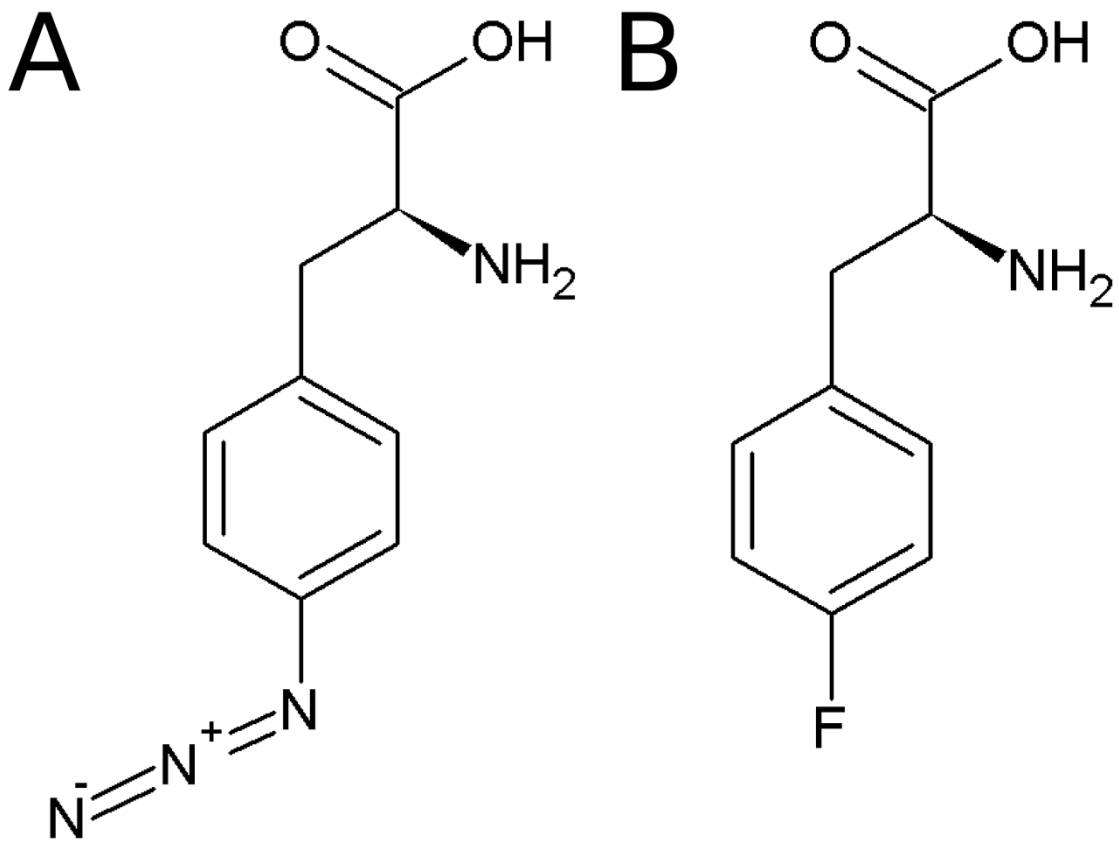

Figure 73. Two unnatural amino acids were chosen as substrate candidates for a selected set of ApnA A_1 PCC7821 variants. A) The 4-Azido-phenylalanine (Az) resembles the native substrates as it has aromatic ring and positively charged group at the end of the side chain. B) 4-Fluoro-phenylalanine is similar to Az with the exception of the missing positive charge at the side chain.

As expected, the WT ApnA A_1 PCC7821 activated mainly its native substrates (Figure 74A). Interestingly, the tested variants appeared to accept the unnatural amino acids as substrates with the 4-azido-phenylalanine as a better substrate candidate. For the E204G variant 4-azido-phenylalanine appeared as equally good and preferred substrate with L-tyrosine (Figure 74B). The enzyme variant was also capable of activating 4-Fluoro-phenylalanine at the comparable level with L-tryptophan. The E204G/S243E double variant showed nearly equal activity towards the tested unnatural amino acids as towards the preferred L-tryptophan substrate (Figure 74C). In the case of S243N, the enzyme had an apparent preference for L-arginine but from the tested aromatic substrates, 4-azido-phenylalanine came out as the strongest candidate (Figure 74D). The S243T mutant appeared to have most relaxed specificity as it activated the tested unnatural amino acids, L-tryptophan and the two native substrates at comparable levels based on the measured absorbance values (Figure 74E). Finally, S243H showed very interesting behaviour. As described in the previous chapter, this enzyme variant was virtually monospecific for L-tryptophan when tested against the 20 proteinogenic amino acids. However, it appeared that the unnatural amino acid 4-Azido-phenylalanine was even slightly better substrate for this S243H variant than L-tryptophan (Figure 74F). Additionally, the enzyme appeared to be able to adenylate 4-Fluoro-phenylalanine but not as efficiently as the two substrates mentioned above.

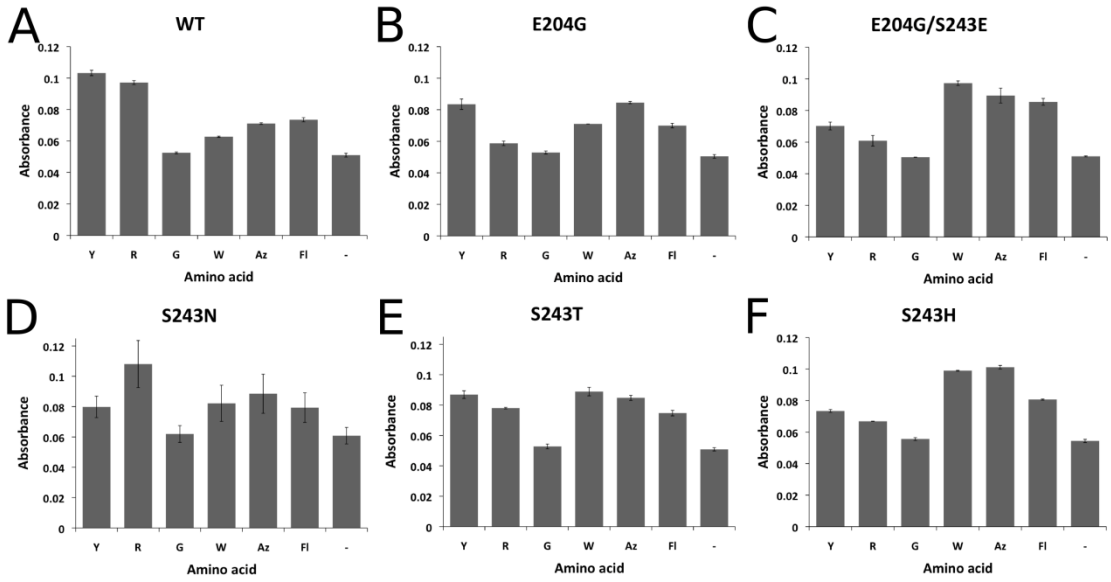

Figure 74. A selection of ApnA A_1 PCC7821 variants was tested against unnatural amino acids 4-Azido-phenylalanine (Az) and 4-Fluoro-phenylalanine (Fl) alongside with the wild type enzyme. A) The wild type enzyme preferred as expected its natural substrates based on the measured A_{540}-values. B) - F) All of the five tested ApnA A_1 PCC7821 variants showed activity towards the unnatural amino acids. 4-Azido-phenylalanine was the preferred substrate out of the two.

4.4.6.3 Active site lysine

Previously it has been shown that the catalytically important lysine residue arises from the conserved sequence motif PxxxxGKxx(R/K), located at the end of the C-terminal domain in adenylation subunits (Marahiel *et al.*, 1997; Gulick, 2009). The corresponding lysine (Lys500) of ApnA A_1 PCC7821 was part of a putative C-terminal loop with no observable electron density even in the presence of substrate. Furthermore, due to the orientation of the C-terminal domain in ApnA A_1 PCC7821, another conserved lysine (Lys415) had taken the position of the presumably catalytic lysine (Figure 75). The Lys415 (Lys434 in PheA) is part of the conserved motif that also includes the hinge region and forms similar hydrogen bond network in the active site as the Lys517 in PheA. In PheA structure a lysine (Lys517) corresponding to Lys500 of ApnA A_1 PCC7821 is

the central coordinating residue in the active site and interacts both with the AMP and the phenylalanine molecule. To study the possibility that Lys415 would take the role of Lys500, single substitutions to these two positions were made by replacing the lysine with an alanine. Additionally, a double variant where both lysines were replaced with alanine was created. The hypothesis was that if these lysines would have ability to replace each other then only the double mutant should be completely inactive. The activity of these mutants was studied using the hydroxylamine-trapping assay as described in materials and methods.

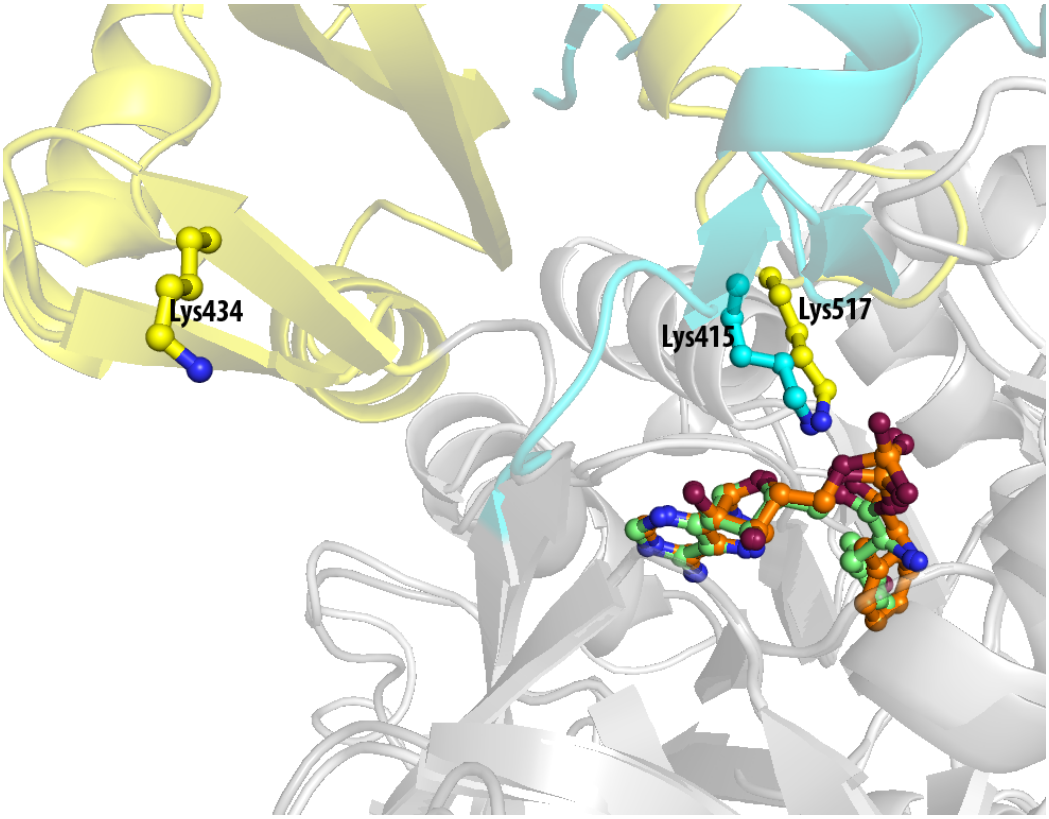

Figure 75. The two conserved lysines. The conserved lysine of the C-terminal domain, corresponding to Lys517 in PheA and Lys500 in ApnA A_1 has shown to be important for the catalytic activity of the adenylating enzymes. The different C-terminal domain orientation in ApnA A_1 PCC7821 results the replacement of this vital lysine with another conserved lysine residue, Lys415. The corresponding lysine in PheA, Lys434, is located further away from the active site. The Lys500 of ApnA A_1 is on the other hand part of the putative C-terminal loop that appered to be disordered and could not be modeled.

The activity of the three lysine mutants, K415A, K500A and K415A/K500A was tested against the 20 canonical proteinogenic amino acids and compared to the activity of the wild type enzyme. As expected, the double mutant was shown to be inactive (Figure 76A). However, when the single variants were tested it was found that Lys500 was indeed critical for the catalytic activity of the enzyme and the loss of this lysine could not be recovered by the lysine at position 415 (Figure 76B and C). The K500A variant had no enzymatic activity whereas the K415A variant was fully active and behaved as the wild type enzyme.

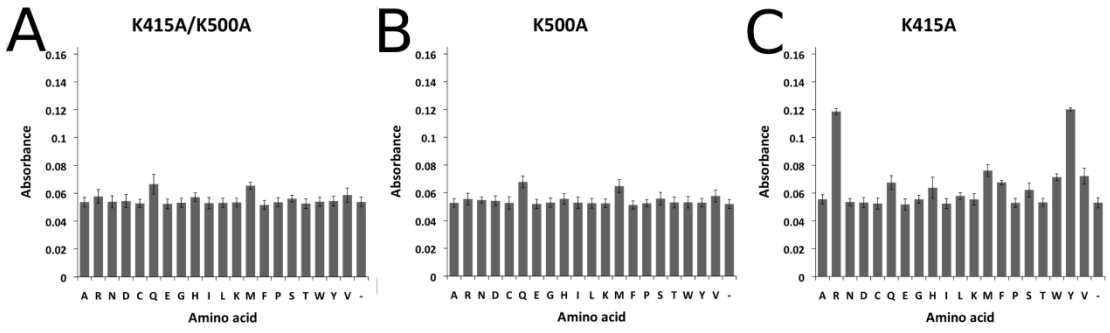

Figure 76. The results for the lysine variants of ApnA A_1 PCC7821 from the hydroxylamine-trapping assay. A) The K415A/K500A double mutant showed the expected lack of enzymatic activity. B) The K500A variant was also inactive, which supports its critical role as a catalytic residue. Lys415 is not able to recover the loss of catalytic activity. C) K415A mutant of ApnA A_1 PCC7821 shows similar activity to wild type, which indicates that this lysine is not required for the adenylation reaction to occur.

5. DISCUSSION

5.1 Ethylene receptor 1

5.1.1 The expression and isolation of the full-length ETR1

The plant hormone ethylene and its signaling pathway have been the topic of intense research for years. This can be attributed to the central role of ethylene as a regulator of several important developmental processes and responses to both biological and environmental stress factors in plants. The well-known effect of ethylene in fruit ripening has also made it an amenable target for industrial applications. Ethylene is known to bind to ER-located receptors, which resemble the prokaryotic sensor histidine kinases. Like ethylene receptors, these sensor histidine kinases are involved in mediating responses to environmental changes. In the model plant *Arabidopsis thaliana,* a group of five ethylene receptors initiate a signaling cascade that is known to include a putative MAPKKK called CTR1 and a member of NRAMP metal ion transporter family, EIN2. Ethylene signaling represents an interesting signaling pathway where prokaryotic-like sensor histidine kinase at the ER-membrane regulates its pathway together with a MAPKKK, a canonical eukaryotic signaling component.

Despite the accumulated knowledge on the molecular determinants in ethylene signaling pathway, very little is known how the signal is mediated from the transmembrane domain of the receptor to its cytosolic domains and further to the downstream components such as CTR1. The structural characterization of a full-length ethylene receptor would provide the means for better understanding of these initial steps in the signaling pathway. Since the ethylene receptor ETR1 from *A. thaliana* is the best-characterized member of its family and shorter constructs of it had been studied in the

group before, it was chosen as a main target for the structural studies in this thesis work. A published protocol for the expression and purification of ETR1 (Voet-van-Vormizeele & Groth, 2008) was used as a starting point.

The expression of the full-length ETR1 in *E. coli* was attempted using the published protocol. To this end, both the gene for the receptor from *A. thaliana* cDNA library and the synthetic ETR1 gene codon-optimized for *E. coli* were cloned into the *E. coli* expression vector pET15b, which added an N-terminal His_6-tag to the target protein for affinity purification. Additionally, both the native gene and the synthetic one were cloned into two other *E. coli* expression vectors (pETM-11/LIC and pETM11SUMO3GFP) to examine if this would have influence on the expression. The expression tests were mainly conducted in the *E. coli* C41(DE3) and C43(DE3) strains, which are more suited for membrane protein expression (Miroux & Walker, 1996) and according to Voet-van-Vormizeele and Groth (2008) were successfully used to produce ETR1. However, the attempts to express the receptor protein in *E. coli* were not successful. As a stable expression using baculovirus expression vector system was achieved, the production of the protein in *E. coli* was not pursued further.

When the expression of ETR1 in *E.coli* was proven to be challenging, alternative expression systems more adapted to eukaryotic protein expression were investigated. As Dr. Imre Berger was a member of the Thesis Advisory Committee (TAC), this provided the opportunity to test ETR1 expression in the Multibac expression system developed in the Berger lab at EMBL Grenoble. The multibac expression system is a baculovirus expression vector system (BEVS) that is especially suited for production of large eukaryotic complexes (Berger *et al.*, 2004; Fitzgerald *et al.*, 2006; Bieniossek *et al.*, 2008). For the expression of ETR1 in Multibac system, the synthetic *E. coli* codon-optimized gene of the receptor protein in a pFastBac HTa vector was ordered form

GenScript (GenScript USA Inc., Piscataway, USA). The integration of the target gene into the baculoviral genome was carried out in *E. coli* stain DH10EMBacY, which carries the baculaviral DNA as an artificial chromosome. The production of the initial virus (V_0), virus amplification as well as the expression of the target protein was performed in *Spodoptea frugiperda* Sf21 cells using HyClone culture medium. For the protein expression, the insect cells were infected with the amplified (V_1) virus and kept in constant density until the day of proliferation arrest after which they started to express the target protein. The protein levels of the simultaneously expressed YFP were monitored to determine when the maximal expression level for ETR1 was reached. During the initial virus amplification, it was noticed that 48 hours after dpa the cells started to disrupt resulting in significant loss of viable cells. The most probable cause for this was suggested to be the insertion of the overexpressed target protein into the cell membrane. Therefore, in the large-scale expression trials the cells were harvested at the 48^{th} hour after dpa. The successful expression of the target protein was confirmed by SDS-PAGE analysis and Western blot.

After the expression of ETR1 was established, the next step was to isolate and purify the receptor from the insect cells. A modified protocol for membrane protein isolation from *E. coli* cells was used. The buffers in the initial experiments included TRIS pH 7.6 and 100 mM NaCl as the main components and 0.5% n-dodecyl-β-D-maltoside (DDM) as the detergent in the solubilization step. The non-ionic DDM was chosen as the detergent since it commonly used in isolation of membrane proteins and thus provided a good starting point. In the first isolation trials, the majority of the protein appeared to remain in the insoluble protein fraction. This indicated that the composition of the buffers was not optimal as the pH of the buffer and the salt concentration can have an effect on the solubilization capacity of the detergent. Therefore as an attempt to improve the

isolation process the pH and NaCl concentration were varied. Both increased and decreased pH was tested in combination with lower and higher salt concentration. Additionally, the DDM concentration in the solubilization step was increased from 0.5 % to 1.0 %. More protein was recovered in the initial lysis step when TRIS at pH 8 and NaCl concentration 20 mM was used but the solubilization was still not successful.

The full-length ETR1 was successfully expressed in insect cells during this thesis work. Additionally encouraging results were obtained from the initial isolation and solubilization attempts. The main bottleneck in this step appeared to be in the final overnight solubilization step, which indicates that the chosen detergent might not be suitable. Therefore for future prospects several detergents should be screened. Possible candidates for testing would be other non-ionic detergents such as octyl-β-D-glucopyranoside (β-OG), polyoxyethylene-(9,5)-p-t-octylphenol (Triton X-100) or Tween 20/40/80, or ionic ones such as N,N-Dimethyldodecylamine-N-oxide (LDAO) or 3-[(3-Cholamidopropyl)dimethylammonium]-1-propanesulfonate (CHAPS). The main challenge is to find a detergent that would solubilize the protein without affecting its stability. Ionic detergents are usually relatively harsh and might denaturate the protein when used at higher concentrations. Therefore the detergent screening would require not only investigation of suitable detergents but also their concentration.

5.1.2 The SAXS model of ETR1-ΔTM

The cytosolic portion of ETR1 (ETR1-ΔTM) was considered an intruiging target for structural studies since it provides the necessary elements for the interaction with the downstream signaling components, such as CTR1. Furthermore, an available expression and purification protocol existed for the ETR1-ΔTM construct (Mayerhofer, 2011). The purified ETR1-ΔTM was prone to aggregate, which was problematic concerning

structural studies. Previously, the addition of NDSB was shown to stabilize the protein. However, in the structural characterization of ETR1-ΔTM using SAXS, the presence of stabilizing NDSB component caused additional scattering effect and subsequently a buffer mismatch at higher angles of the scattering curve. The SAXS models in the presence of the NDSB compound indicated dimeric arrangement of the protein with an overall dumbbell shape (Mayerhofer, 2011). In the current study, a more SAXS compatible stabilizing agent was found from glycerol. The final SAXS data obtained in the presence of glycerol was of good quality and no additional scattering from the buffer was observed. The calculated D_{max}-value of 190 Å as well as MM estimation from Porod volume (161 kDa) agreed with the MM of the dimeric molecule (130 kDa) while the obtained MM_{SAXS}-value (53 kDa) from comparison with BSA standard was in line with the MM of the monomer (65 kDa). Therefore, based on the SAXS parameters determined from the scattering data, ETR1-ΔTM appered to exist in monomer-dimer equilibrium in the solution as was seen in the presence of NDSB. Both *ab initio* bead modeling and the rigid body refinement using crystallographic high-resolution structures resulted in the expected dumbbell shaped dimer with the GAF domains at the top and the CA and receiver domains at the bottom of the central helical bundle. The overall dimensions of the dimer in presence of glycerol were slightly larger than in the presence of NDSB, which appeared to result from the flexible movement of the receiver domain. To this end, a flexibility analysis was performed using EOM approach. The results from this analysis indicated a possible switching between a more extended and more compact conformations with R_g-values of 54 Å and 51 Å, respectively. The more abundant species has the R_g of 54 Å, which correlates well with the R_g-value of 56 Å from the rigid body modeling.

In the current study, the main aim for the ETR1-ΔTM construct was to find stabilizing purification conditions that would allow the collection of SAXS data without scattering effect from the buffer components. This was achieved by replacing the NDSB compound with glycerol and resulted in better fit at the higher angles in the structural modeling. The obtained *ab initio* and rigid body models confirmed the dimeric arrangement of ETR1-ΔTM with an overall dumbbell shape as was seen in the presence of NDSB. The observed flexibility at the C-terminal end of the construct could possibly be result from the lack of the cytosolic interaction partner, CTR1. Therefore for future prospects, structural studies in the presence of CTR1 should be done.

5.2 EDR1

The mitogen-activated protein kinases (MAPKs) play a crucial role in wide variety of cellular responses in eukaryotic organisms. They are involved in signaling cascades that regulate developmental processes, stress responses and apoptosis, among others. The protein kinase Enhanced Disease Resistance 1 (EDR1) from *Arabidopsis thaliana* is a Raf-like MAPKKK that acts as negative regulator of plant defence and programmed cell death. The Arabidopsis plants that carry a mutation in the *edr1* gene display enhanced disease resistance to powdery mildew, which is caused by the fungus *Erysiphe cichoracearum* (Frye & Innes, 1998, Frye *et al*., 2001). EDR1 has been linked to several plant signaling cascades, such as abscisic acid, jasmonic acid and ethylene signaling pathways (Frye *et al*., 2001). It shares high sequence identity with another *A. thaliana* MAPKKK, CTR1, which negative regulates ethylene-signaling pathway. EDR1 has been shown to be involved in regulation of ethylene-induced senescence but not via the main ethylene signaling pathway based on the triple response assays (Frye *et al*., 2001).

5.2.1 Autophosphorylation activity of EDR1 kinase domain

The *A. thaliana* EDR1 has previously been shown to possess Ser/Thr kinase activity and furthermore to be able autophosphorylate *in trans*. The autophosphorylation is not an uncommon phenomenon among protein kinases and is linked to the activation of the enzyme through phosphorylation of flexible activation segment. Furthermore, the activation of protein kinases through phosphorylation has been shown to be in some cases dimerization dependent. Recent example of dimerization-dependent activation is the kinase domain *A. thaliana* CTR1, which was shown to form specific "back-to-back" dimers in the active, phosphorylated state (Mayerhofer *et al*., 2013). Mutations on the important dimer-interface residues led to inactivation of the enzyme. Due to its similarity with CTR1, EDR1 was also postulated to show similar dimerization-dependent behaviour. To this end, the kinase domain of EDR1 was cloned, expressed and purified for crystallization and subsequent structural characterization.

The wild type EDR1 kinase domain was resistant to crystallization despite of the vigourous attempts. The SDS-PAGE analysis of the purified sample indicated the presence of a protein impurity, which had molecular mass near the target protein and could therefore not be removed by SEC. Based on the accurate mass analysis with ESI-MS and peptide-fingerprinting experiments the purified enzyme sample contained multiple protein species in different phosphorylation states. The main protein species appeared to possess 3, 4 or 5 phosphorylated residues. The peptide-fingerprinting MS analysis with trypsin and chymotrypsin digested samples led to the identification of nine putative phosphorylation sites, which corresponded to the amount of predicted sites. As was observed with CTR1-kd, also the phosphorylations sites of EDR1 kinase domain concentrated on the activation segment.

The structural studies on the wild type EDR1 kinase domain were hampered due to the heterogeneity of the protein sample. As differently phosphorylated protein species appeared to be one source of the heterogeneity, a kinase inactive version of the enzyme was made. To this end, the catalytically important aspartate (Asp792) was substituted with asparagine. This aspartate is part of a conserved HRD (sometimes referred as HxD motif) and indicated to function as catalytic base that accepts the proton from the hydroxyl group of the target residue in phosphotransfer reaction (Hanks & Hunter, 1995). The EDR1-D792N construct was expressed and purified under similar conditions as the wild type enzyme. Based on the SDS-PAGE analysis the purified protein sample contained only the target protein and no additional protein species were detected. Additionally, an accurate mass analysis with ESI-MS was performed for EDR1-D792N to confirm the homogeneity of the purified sample. As expected, the most abundant protein species detected had a molecular mass of 32,008 Da, which corresponds very accurately the non-phosphorylated EDR1-D72N construct. To confirm the inactive state of EDR1-D792N, a coupled-kinase assay was performed. In this assay the consumption of ATP is directly proportional to the oxidation rate of NADH, which could be measured. The EDR1-D792N showed no activity in this assay while for the wild type kinase domain appeared to be fully active and allowed the calculation of kinetics parameters according to Michaelis-Menten equation.

5.2.1.1 EDR1 not activated by dimerization?

The kinase inactive EDR1-D792N was successfully crystallized in the presence of nucleotide analog AMP-PNP. The crystals belonged to the space group $P3_221$ and diffracted x-rays to 2.55 Å resolution. The asymmetric unit was found to contain two

molecules, which led to investigate the possibility whether similar back-back dimer would be observed as seen for CTR1. In the back-to-back dimer of CTR1, the central interactions at the interface are formed by three residues, Arg604, Phe611 and Met612. The corresponding residues of EDR1 (Arg722, Phe729 and Leu730) were found also to locate at the interface of the putative EDR1-D792N dimer. However, they were positioned at the periphery of the interface and were not interacting with their partner residues in the neighbouring molecule. Further examination of the interface by PISA revealed that the covered surface area was significantly smaller than expected for a back-to-back interface. Therefore it became clear that the dimeric arrangement observed in the crystal structure of EDR1-D792N differed from the expected back-to-back interface. This difference was also easy to detect by visual inspection of the surface models for the EDR1-D792N crystal structure and the CTR1-kd dimer (Figure 77A and B). It is questionable whether the inactive state of the enzyme is the reason for the observed packing as the kinase inactive CTR1-D676 also formed back-to-back dimer in the crystal structure. Furthermore, when the oligomerization state of EDR1-kd and EDR1-D792N was examined using SEC-LS/UV/RI, it was found that both the wild type kinase domain and its inactive version were primarily in monomeric form. Therefore it was concluded that the autophosphorylation activity of EDR1-kd is unlikely to be dependent on homodimerization.

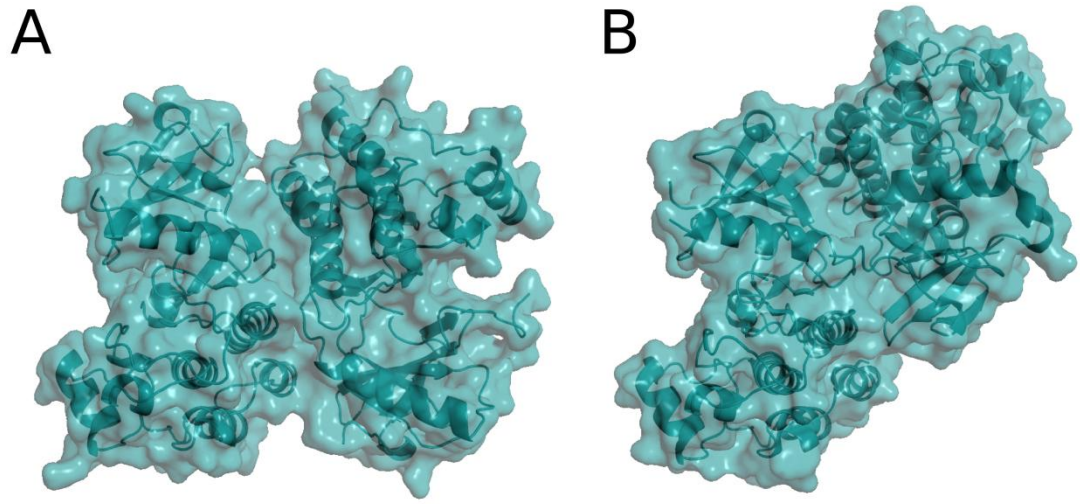

Figure 77. The dimeric arrangement observed in the asymmetric unit of EDR1-D792N differed from the back-to-back dimer of CTR1. A) The surface representation of EDR1-D792N dimer as observed in the crystal structure. B) The back-to-back dimer of CTR1-kd (PDB code 3PPZ) shown as surface.

5.2.1.2 AMP-PNP binding site

The crystal structure of EDR1-D792N exhibited the expected bi-lobal organization of protein kinase. The active site of the enzyme located at the conserved position between the N-terminal and C-terminal lobe of the enzyme. Both copies of EDR1-D792N were found to have AMP-PNP bound to the active site. The position and the binding mode for the adenosine ring of the nucleotide analog were virtually identical in the two chains (Figure 78). Some minor differences could were found in the coordination network of the triphosphate unit due to the altered rotamer of Lys696 and Arg753 in the chain B. Lys696 was found to participate in salt bridge formation with Glu714 in chain B while in chain A it provided hydrogen bond to the oxygen of the β-phosphate group. Similarly, Arg753 was only participating in the nucleotide binding in the chain A while in the chain B it was facing away from the active site and towards a EDR1-D792N molecule from adjacent asymmetric unit.

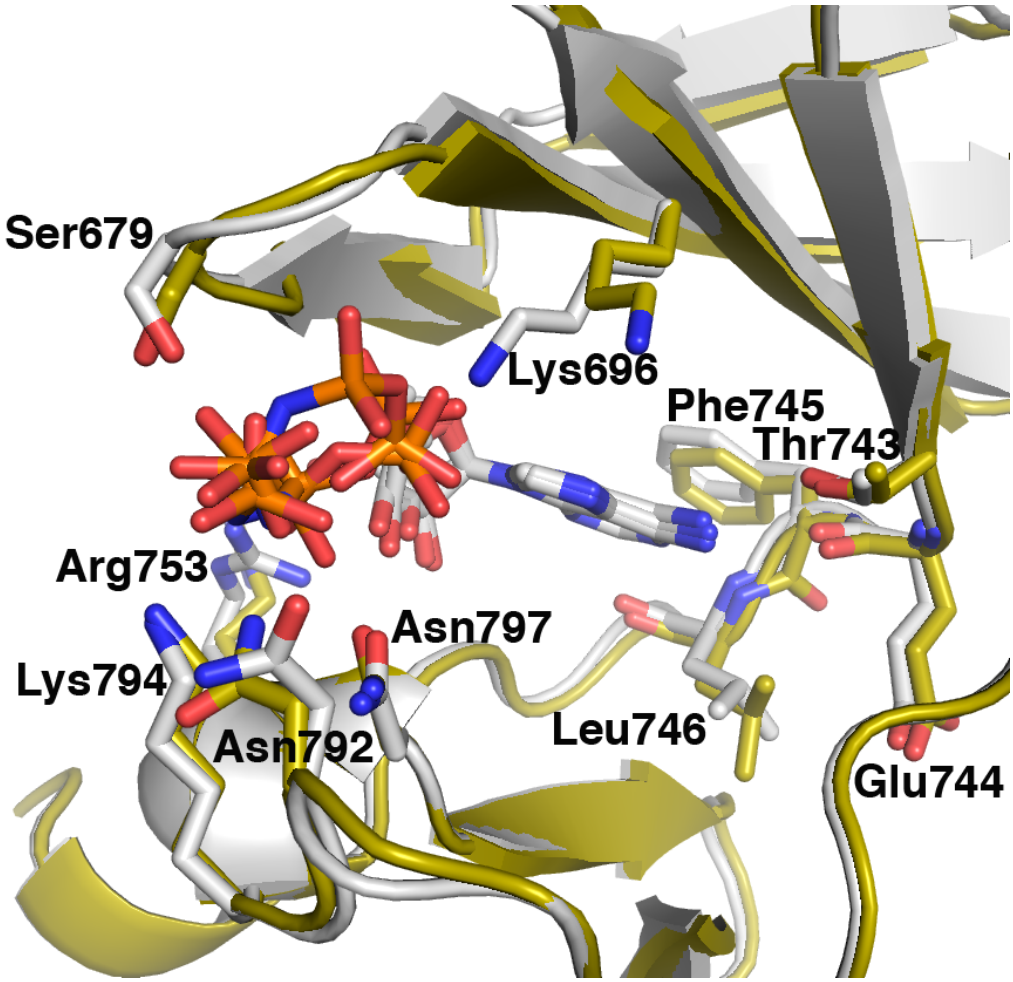

Figure 78. The overlay of AMP-PNP binding pocket for the EDR1-D792N chains A (in grey) and B (in gold). A highly similar binding mode for the nucleotide between the two chains was observed.

5.2.1.3 Enzyme-substrate complex of EDR1-D792N

The overall fold of EDR1-D792N shared closest resemblance with the crystal structure of CTR1-D676N, which was in line with the high sequence homology between the two enzymes and also with the inactive state. The hallmark features that separate active kinase from an inactive one include a salt bridge between conserved lysine and glutamate residue (Lys696 and Glu714 in EDR1) and structural rearrangement of the activation segment and αC-helix (Kornev *et al.*, 2006). The phosphorylation induced folding of the activation segment allows the formation of a short two-stranded β-sheet

(strands β6 and β9), which doesn't exit in an inactive kinase, often due to the lack of stable conformation for the activation segment. The chain A in EDR1-D792N crystal structure was found to have the characteristics of an inactive kinase. Due to the orientation of the αC-helix, the glutamate residue (Glu714) was pulled away form Lys696 thus disabling the formation of the conserved salt bridge. Additionally, part of the activation segment was disordered and lacked stabile 3D organization. This comprised four residues preceding the P+1 loop at the C-terminal end of the activation segment. Surprisingly, the second molecule (chain B) in the asymmetric unit of EDR1-D792N was found to have fully folded activation segment. Furthermore, two threonine residues appeared to be phosphorylated despite the expected non-phosphorylated and inactive state of the protein. These phosphorylations located at positions 820 and 827 correlating with putative phosphorylation sites from the peptide fingerprinting analysis with the wild type enzyme. TPO820 aligned with known phosphorylation site of CTR1-kd (TPO704) while TPO827 adopted similar position with an assumed primary phosphorylation site of CTR1 (SEP710) (Figure 79). The phosphate group of TPO827 was accommodated within the so-called basic pocket as seen for primary phosphorylation sites in other kinases.

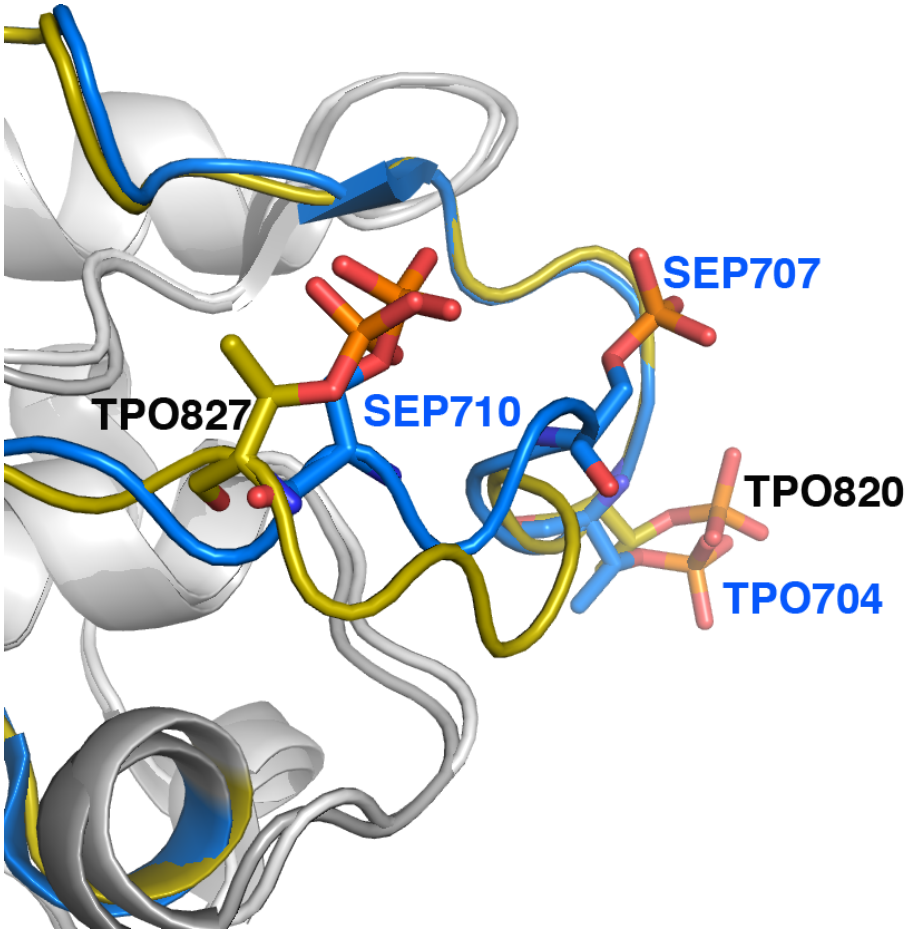

Figure 79. The phosphorylation of the activation loop in the active-like EDR1-D792N molecule. EDR1-D792N was observed to contain two phosphorylations at one of the molecules in the asymmetric unit. These phosphorylations located in the activation loop and had nearly identical position with two confirmed phosphorylation sites of CTR1-kd (numering in blue colour). The activation segment of CTR1-kd is shown in blue and of EDR1-D792N in gold with the phosphorylated residues as sticks for both.

The observed difference in the phosphorylation state between the two molecules in the crystal structure of EDR1-D792N led to more detailed comparison. The structural alignment of the two chains revealed that the phosphorylated EDR1-D792N molecule possessed the described characteristics of an active kinase. Recently, similar phenomenon was observed in the crystal structure of the catalytically inactive PAK1 kinase domain (PAK1-KD$^{K299R/D389N}$) where the asymmetric unit also contained two copies of the enzyme with one in inactive and the other in active conformation (Wang *et al*., 2011).

However, the observed dimeric arrangement for PAK1-KD$^{K299R/D389N}$ represented the expected trans-autophosphorylation complex of protein kinases. The importance of similar transautophosphorylation complex had been previously shown for the receptor Ser/Thr kinase PknB from *M. tuberculosis* (Mieczkowski *et al*., 2008). However, the enzymes in the PknB complex were both in an inactive state. Interestingly, when the interfaces between EDR1-D792N molecules across different asymmetric units were examined, a putative enzyme-substrate complex was found. In this complex the partially disordered activation segment of the inactive EDR1-D792N molecule functioned as the substrate and protruded the active site of the neighbouring molecule in the active site. A serine residue (Ser824) was optimally positioned near the γ-phosphate group of the bound AMP-PNP and the side chain carboxyl oxygen of the mutated aspartate (Asp792). This appeared to mimic the situation prior to the phosphotransfer reaction, which involves deprotonation of the target hydroxyl group by the aspartate and subsequent nucleophilic attack of the deprotonated hydroxyl group to the γ-phosphate. Similar phenomenon was observed in the crystal structure of PAK1-KD$^{K299R/D389N}$, but for a residue from a different part of the activation loop. The reason for this appeared to be in the different folding pattern of the activation loop in EDR1-D792N when compared to PAK1-KD$^{K299R/D389N}$. Additionally, the orientation of the protein molecules in the ES-complex differed in the two, based on the structural alignment of the EDR1-D792N enzyme-substrate dimer with the PAK1-KD$^{K299R/D389N}$ structure (Figure 80). It has been previously proposed that the rotation of the molecules in relation to one another helps to orient the phosphorylation target to correct position in the active site. This could also explain the observed difference between the ES-complex of EDR1-D792N and PAK1-KD$^{K299R/D389N}$.

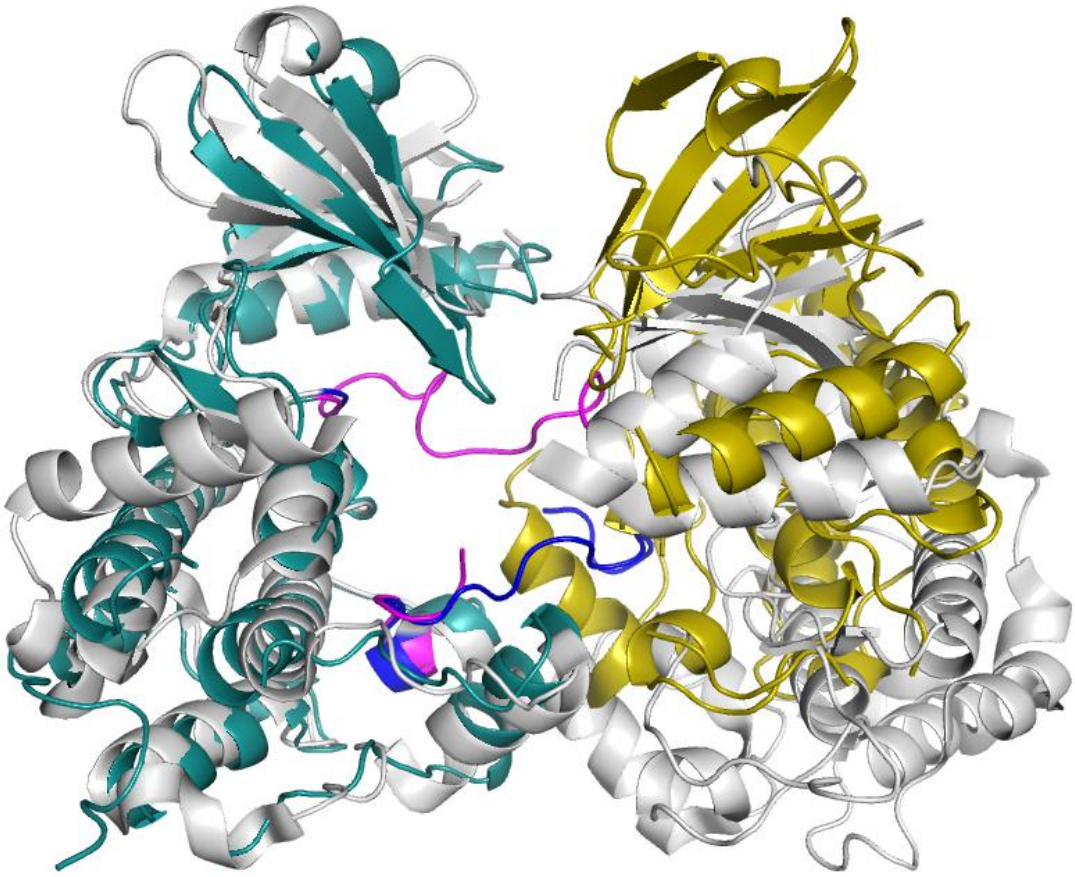

Figure 80. The structural alignment of EDR1-D792N ES-complex with the corresponding one for PAK1-KD$^{K299R/D389N}$. Based on the overlay, the orientation of the molecules in the complex differed between the two. The inactive and active-like EDR1-D792N molecules are shown as dark green and gold cartoon representatios. The PAK1-KD$^{K299R/D389N}$ ES-complex is shown in gray cartoon model. The figure also shows the different orientation of the partially unfolded activation loops for the two enzymes. These loops are coloured in purple and blue for EDR1-D792N and PAK1-KD$^{K299R/D389N}$, respectively.

To confirm the authentity of the ES-complex seen between the EDR1-D792N molecules from adjacent asymmetric units, the features of the interface were investigated. The residues from the αG-helix appeared to form a zipper-like organization that promoted the interaction between the two molecules. The importance of αG-helices in the ES-complex formation was shown for the PknB protein kinase where mutations that disrupted the interaction between the αG-helices resulted in an inactive enzyme. Using

the mutational studies with PknB as an example, a phenylalanine (Phe881) and a valine (Val876) from the αG-helix of EDR1-kd were substituted with alanine. Addtionally a double alanine substitution was made. Indeed, the three αG-helix mutants of EDR1-kd were shown to be inactive in the coupled kinase assay. This supported the observations from the structural studies and gave a clear indication that the transautophosphorylation requires also in EDR1 the formation of αG-helix mediated ES-complex.

5.3 ApnA A_1 PCC7821

5.3.1 Structural basis for the substrate recognition

The A-domains of Non-ribosomal peptide synthetases have been extensively studied the past decades in order to understand the underlying features that define their substrate specificity. A combination of genetical and structural studies has been employed to gain information about the key residues at the active site of these enzymes. The main aim for these studies has been to obtain enzyme variants with altered substrate specificity that could then be used to create new biologically active peptides for a variety of industrial applications. The first breakthrough was the first crystal structure of PheA and the subsequent discovery of the so-called specificity conferring code ("Stachelhaus code"), which defined a set of ten important active site residues (Stachelhaus *et al*., 1999). This code allowed the prediction of the catalytic activity for newly found A-domains but also to some level to modify the activity of A-domains with known subtsrate specificity (Eppelmann *et al*., 2002; Stevens *et al*., 2006; Chen *et al*., 2009; Villiers & Hollfelder, 2011; Thirlway *et al*., 2012). However, the engineering of the known A-domains has mainly led to variants with activity towards substrates that are similar to the

natural one. Structural information from A-domains with different substrate specificities would therefore provide the knowledge on the binding of the specific amino acids.

Recently, a bi-specific A-domain was discovered from a cyanobacterium *Planktotrix agardhii* (Christiansen *et al*., 2011). The enzyme ApnA A_1 is part of the initiation module in an Anabaenopeptin synthetase and is an interesting enzyme of its class since it is able to use both tyrosine and arginine as natural substrates. The side chains of these two substrates have very different chemical properties, which made it difficult to understand how the active site of the enzyme accommodates both of them. To understand the bi-specificity of this enzyme, X-ray crystallography was used to investigate ApnA A1 from a *P. agardhii* strain PCC7821. The plasmid with the gene for the wild type enzyme with an N-terminal His_6-tag was obtained from G. Christiansen and the expression and subsequent purification was conducted in Hamburg University in the lab of Prof. A. Rentmeister. The crystallization experiments and data collection were done at EMBL Hamburg.

The first structure for ApnA A_1 PCC7821 was obtained in the unliganded form. It was noticed that structure consisted of residues 2 – 410, which covered mainly the N-terminal domain of the enzyme. Based on a structural alignment, the fold of the unliganded ApnA A_1 PCC7821 resembled the N-terminal domain of PheA (Figure 81). The lack of interpretable elecron density for the C-terminal residues 411-547 indicated that they were disordered, presumably due to a lack of stabilizing interaction between the N-terminal and C-terminal domain in the absence of the bound substrate. Indeed, when the native crystals of ApnA A_1 PCC7821 were soaked in the presence of AMP-PNP, a non-hydrolyzable ATP analog, the C-terminal domain residues 411-444 and 452-492 could be built into the model. The topology of the C-terminal domain was similar to the C-terminal domain of PheA. However, the overall conformation of ApnA A_1 PCC7821

differed from PheA. Instead, ApnA A_1 PCC7821 had adopted the so-called closed conformation similar to the one seen for the Acetyl-coenzyme A Synthetase from *S. enterica* (PDB code: 1PG4) and for DltA from *B. subtilis* (PDB code: 3E7W). Two arginine residues (Arg409 and Arg420) and a lysine (Lys415) from the C-terminal domain of ApnA A_1 PCC7821 were found to interact with the bound nucleotide and thus promoted the observed overall conformation. The arginine residues and the lysine formed hydrogen bonds to the α- and β-phosphate groups of the AMP-PNP, respectively. The expected catalytic lysine (Lys500) from a conserved motif at the C-terminal domain was found not to locate at the active site but in fact was part of the unstructured C-terminus. Since Lys415 had taken the position of Lys500, it was initially hypothesized to be an alternative catalytic lysine. However, the substitution of Lys500 with alanine rendered the enzyme completely inactive while the variant with similar mutation at position 415 behaved like the wild type enzyme. Therefore it was confirmed that Lys500 is also required for the catalytic activity in ApnA A_1 PCC7821 while Lys415 is not.

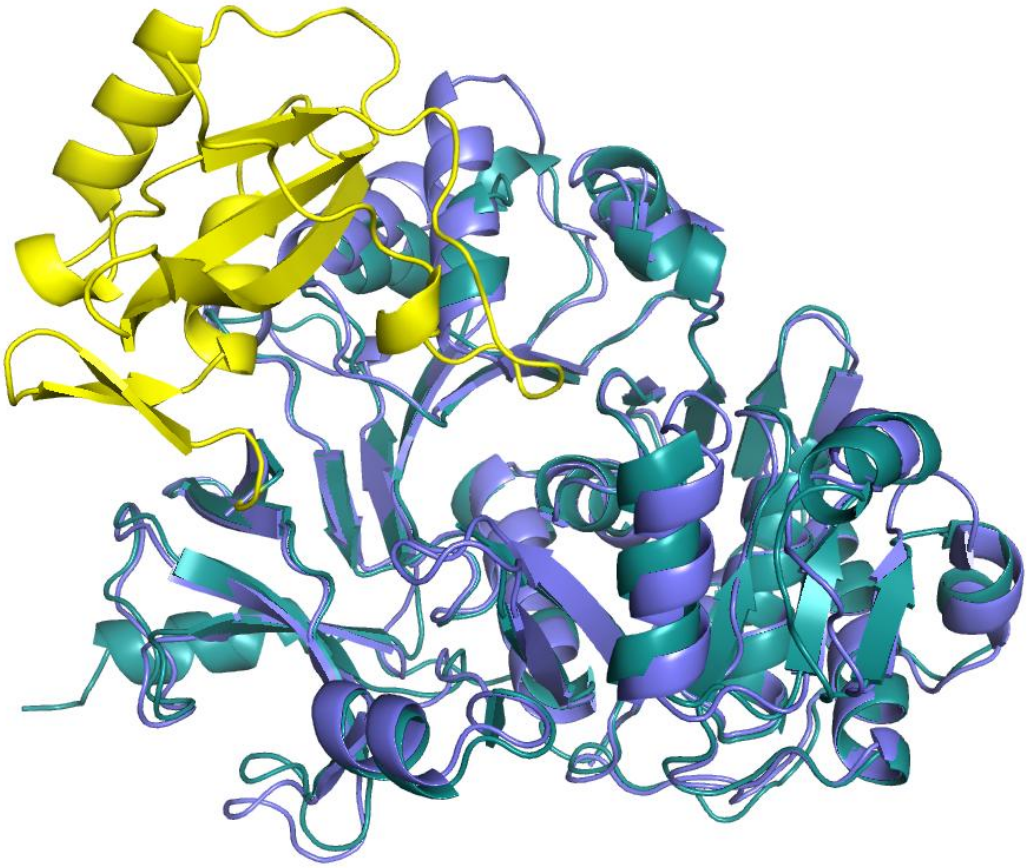

Figure 81. The structural alignment of the unliganded ApnA A_1 PCC7821 with PheA. The structure of ApnA A_1 PCC7821 included only the N-terminal domain of the enzyme and had similar overall fold with the N-terminal domain of PheA. The two protein structures are shown as cartoon representation, ApnA A_1 PCC7821 in dark lilac, the PheA N-terminal domain in dark green and the C-terminal domain in yellow.

The structure of the ApnA A_1 PCC7821 in complex with AMP-PNP revealed that the nucleotide binding site located in a cleft between the two domains and shared the conserved elements with other adenylating enzymes. To gain insights into the amino acid binding and subsequently into the bi-specificity of ApnA A_1 PCC7821, the crystal structure of the enzyme was solved in complex with AMP-PNP and L-arginine, and also in complex with the adenylated amino acid substrates, L-tyrosine and L-arginine. Despite the apparent difference between the side chains of these amino acids, it was found that both were surprisingly bound within the same binding pocket. The amino acid binding

pocket of ApnA A_1 PCC7821 located at conserved position within the N-terminal domain as seen with other adenyalting enzymes and comprised a set of residues, which included most of the specificity-conferring code residues (Figure 82). The exception was the catalytic lysine (Lys500), the position of which was taken by Lys415 due to the alternative orientation of the C-terminal domain as decribed for the AMP-PNP complex. Three active site residues at positions 204, 243 and 307 were implied to have specific role in accommodating the side chains of the amino acid substrates. The importance of the residue at position 307 in discriminating between naturally occurring monospecific and bi-specific ApnA A_1 variants had already previously been indicated when the enzymes from different *Planktothrix* species were studied (Christiansen *et al*., 2011). The corresponding alanine residue of ApnA A_1 PCC7821 located at the edge of the binding pocket to allow van der Waals interactions with the side chain of the substrate. Additionally, based on the structural alignment of ApnA A_1 PCC7821 with three known A-domain (PDB codes: 3DHV, 3VNS and 1AMU), there appeared to be some correlation between the side chain size of the residue at identical position with Ala307 and the amino acid substrate. The glutamate and serine residues at positions 204 and 243 at the base of the substrate-binding pocket were found to provide the key hydrogen bond partners for the bound substrate. The side chain carboxy group of Glu204 was observed to tilt slightly to allow formation of a bidentate salt bridge with the guanidium group of the arginine substrate. With the hydroxyl group of L-tyrosine substrate Glu204 formed a single hydrogen bond. The interactions with Ser243 occurred via a water molecule that functioned as a bridge between the substrate and the hydroxyl group of Ser243. Interestingly, in order to achieve the described coordination network and fit the shape of the pocket better, the side chain of the adenylated arginine adopted a conformation that mimicked the shape of the tyrosine side chain (Figure 83).

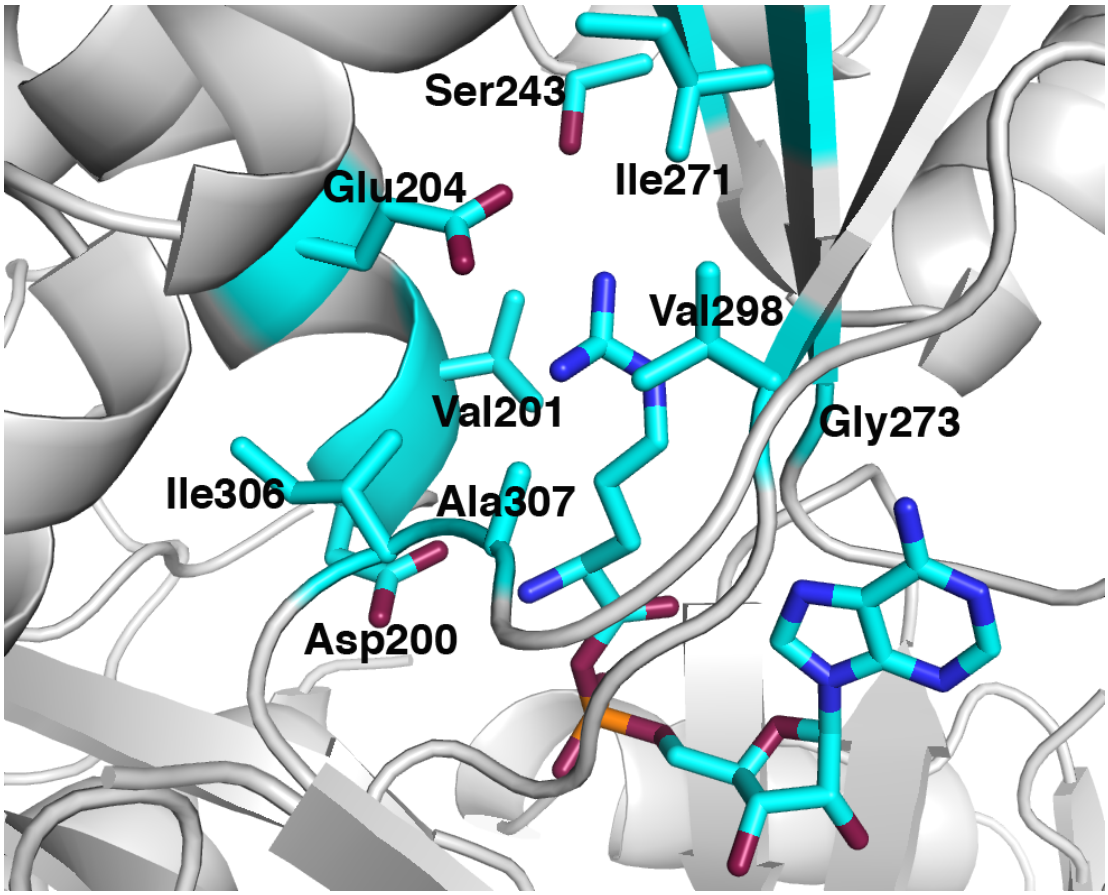

Figure 82. The specificity-conferring code residues at the active site of ApnA A_1 PCC7821. The catalytic and invariant lysine residue (Lys500) is not shown since it was part of unstructured loop, which could not be built into the model. The residues as well as the bound arginyl adenylate are shown as sticks in cyan while the rest of the protein is shown as cartoon representation in light grey.

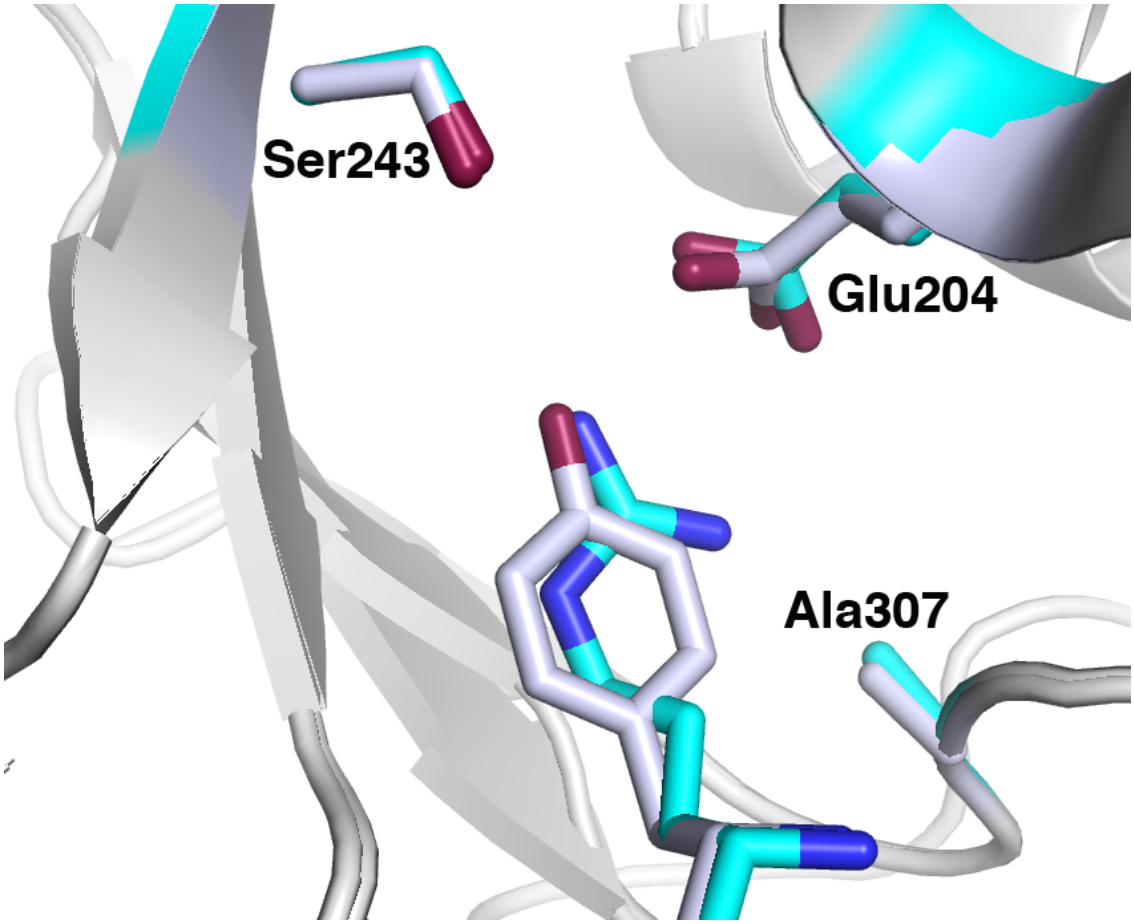

Figure 83. The side chain rotamer of the adenylated arginine mimicked the shape of the aromatic ring of tyrosine. The three active site residues, Glu204, Ser243 and Ala307 were considered to be important for the substrate binding of ApnA A_1 PCC7821 based on structural and phylogenetic analysis. The alanine residue is packed against the side chain of the amino acid substrate, as shown here for the adenylated arginine. The Glu204 is either hydrogen bonded to the substrate, as seen with tyrosine or in the case of arginine forms a salt bridge to the guanidium group of arginine through its side chain carboxylate.

The active site residues Glu204, Ser243 and Ala307 were reasoned to be optimal targets for mutational studies as their location and interaction with substrate indicated a high importance for the substrate binding and activity of the enzyme.

5.3.2 Alteration of the ApnA A_1 PCC7821 substrate specificity

Previous studies on the ApnA A_1 enzymes from seven different Planktothrix strains indicated that naturally occurring amino acid substitutions at specific positions

have an effect on the substrate specificity of the enzyme (Christiansen *et al*., 2011). More specifically, it was noted that the enzymes with acidic residue at the positions 243 and 307 activated only basic amino acid substrates (arginine, homoarginine and lysine) while the enzymes with more neutral residues, namely serine and alanine at corresponding positions were able to use both basic and aromatic amino acids as substrates. The structural studies on a bi-specific ApnA A_1 from *P. agardhii* strain PCC7821 supported the important role of both Ala307 and Ser243 in the substrate binding. Additionally, Glu204 was found to be involved in the central hydrogen bond network with the side chain of the amino acid substrate. Based on the combined information from the studies by Christiansen *et al.* (2011) and the obtained results from the structural analysis for ApnA A_1 PCC7821, a number of substitutions to the positions 204, 243 and 307 were made with the aim to change the substrate specificity of the enzyme. It was hypothesized that variants with more basic binding pocket would provide more optimal enviroment for the activation of acidic amino acids and vice versa. Additionally, the removal of hydrogen bond donor at the position 204 or 243 and its replacement with aliphatic amino acid were expected to favor binding of more hydrophobic/aliphatic amino acids. The substitution of the residue at position 307 with a larger hydrophobic residue was expected to create an enzyme variant with activity towards smaller hydrophobic/polar substrates.

The standard method to study the activation of amino acids in adenylating enzymes is the ATP-pyrophosphate assay, which requires the use of radiolabeled ATP analog. However, the ApnA A_1 PCC7821 variants were tested using a new assay known as hydroxylamine-trapping assay (see materials and methods for a detailed description), which has been successfully employed to determine the adenylation activity of NIS synthetases (Kadi & Challis, 2009). The assay was proven to be suitable for the activity studies of the target protein as the wild type enzyme showed the expected activation of L-

tyrosine and L-arginine. In few instances also the enzyme variant behaved as the wild type enzyme presumably either due to negligible change in the active site by the introduced mutation (A307G) or due to compensation by the nearby residues (S243R). Interestingly, the majority of the active variants were found to be monospecific for L-tyrosine. Substitution of the alanine at position 307 with a slightly larger non-polar residue (Val, Leu) eliminated the activation of L-arginine but retained the activity for L-tyrosine. Similarly, the replacement of Ser243 with a non-polar (Leu, Met, Pro) or a larger polar residue (Tyr) resulted in an enzyme, which appeared explicitly activate L-tyrosine. It was reasoned that a non-polar residue with a side chain of moderate size at position 307 provided additional surface for van der Waals interactions with the aromatic ring of tyrosine thus promoting its binding while the bulkier side chain of arginine can no longer be accommodated. The tyrosine substrate appeared to be better adjusted to the loss of a hydrogen bonding parther at position 243 when additional hydrophobic interactions were provided by the modified amino acid.

The substitution to position 243 or 307 with acidic amino acids (Asp, Glu) was expected to create a variant that would activate mainly basic amino acids (Lys, Arg). Indeed, the S243D variant was shown to mainly activate arginine but also lysine but to a far less extent. Surprisingly, substitution of asparagine to position 243 also had a similar effect and the corresponding enzyme variant appeared monospecific for arginine. However, it was noticed that replacing Ala307 with charged amino acids resulted in an inactive enzyme, probably because it inpaired proper folding of the enzyme. Similarly, the S243E variant was found to be inactive.

Most of the active ApnA A1 PCC7821 variants did not acquire new substrate specificity but rather lost the ability to activate one of the native substrates. However, in few instances the substitution led to activity for L-tryptophan. This often correlated with

the increased space in the active site, as was the case with the substitution of glycine either to position 204 or 243. Similarly, the introduced threonine into position 243 appeared to create more space and also putatively additional hydrophobic contact for the large side chain of the tryptophan substrate. The S234G and S243T variants were also retained their wild type activity while the E204G variant no longer activated arginine. The latter observation indicated that this active site glutamate was critical for the binding of the arginine substrate. An interesting change in the activity was seen when histidine was substituted to position 243. The corresponding enzyme variant accepted only L-tryptophan as substrate, possible through optimal ring stacking between the active site residue and the tryptophan substrate.

5.3.3 From natural to un-natural amino acids

The *P. agardhii* PCC7821 ApnA A_1 provided a good starting point for alteration of the substrate specificity, as it is capable of adenylating two amino acids with very different chemical properties. Indeed, the mutants showed that the substrate specificity of the enzyme could be altered. However, the active variants appeared to prefer one of the native substrates or tryptophan, which similarly to tyrosine has a large, aromatic side chain. The attempts to change the active site suitable for smaller and/or acidic residues were not successful. As the A-domains of other NRPSs are known to use also non-proteinogenic amino acids as substrates, the possibility that one of the ApnA A_1 PCC7821 variants would have also acquired specificity towards a non-proteinogenic amino acid was examined. Two unnatural amino acids, 4-azidophenylalanine and 4-fluorophenylalanine with applications in bio-orthogonal click chemistry (Baskin & Bertozzi, 2007) were chosen as both had similar chemical properties with the native

substrate. Both 4-azidophenylalanine and 4-fluorophenylalanine contain a side chain with aromatic ring similar to tyrosine.

Four single substitution variants (S243H, S243N, S243T, E204G) and the E204G/S243E double variant of ApnA A_1 PCC7821 were tested as these were expected to have large enough binding pocket or the suitable environment for the activation of the unnatural amino acids based on the initial activation assays with the proteinogenic amino acids. All five were found to be able to activate the tested unnatural amino acids but to variable levels. 4-azidophenylalanine was better accepted as substrate candidate than 4-fluorophenylalanine. Its activation was in general at equal or comparable level with the preferred proteinogenic substrate for the tested variants. However, the S243N and S243H variants appeared to be exceptions to this rule. The former showed lower levels of activation for the unnatural amino acid than for its preferred proteinogenic substrates L-arginine. Surprisingly, the latter variant had a minor but still noticeable preference for 4-azidophenylalanine over the tryptophan substrate.

The NRPS system provides an efficient way to produce biologically active peptides for a number of purposes, including in pharmaceutical industry. The central role of A domains in defining which amino acid is incorporated to the peptide product, has made them a target of intense research the past decades. The discovery of specificity-conferring code (Stachelhaus *et al*., 1999) was a breakthrough as it allowed the prediction and alteration of the substrate specificity of A domains. Also in the current study of the bi-specific cyanobacterial enzyme ApnA A_1 PCC7821, it was found that the important substrate-binding residues were part of the specificity-conferring code. As in the previous studies with other A domains, a radical alteration of the substrate specificity of ApnA A_1 PCC7821 was proven to be challenging. However, specific mutants of ApnA A_1 PCC7821 were found to accept un-natural amino acids (4-fluorophenylalanine and 4-

azidophenylalanine) as substrates, which has not been shown for other A domain according to current knowledge. Whether the unnatural amino acid can be incorporated into the final peptide product should be tested. As the use of recombinant DNA technology is possible in *Planktothrix*, this would provide possibility to test the functionality of the mutant A domain as part of the NRPS enzyme complex within the host organism. A possible problem might lie in the assumption that A domain is the only one carrying amino acid selection capacity, as recent studies have indicated that also the C domain has substrate specificity to some extent. However, if the incorporation of 4-azidophenylalanine by the ApnA A_1 PCC7821 variant would succeed, this could provide a mean to study the binding targets and localization of the produced anabaenopeptin within the living cell through click chemistry mediated labeling.

ACKNOWLEDGEMENTS

I want to thank my first EMBL supervisor Jochen Müller-Dieckmann for providing me the opportunity to work in his group. I'm also very grateful for his continuous support and encouragement and for the help with various issues during my time as predoctoral fellow. I would also like to thank Rob Meijers both as a member of my Thesis Advisory Committee (TAC) and as my supervisor. I'm very thankful for all the great advices he gave me during our numerous discussions and also for introducing me to the interesting world of non-ribosomal peptide synthetases. I also want to acknowledge my TAC members Imre Berger from EMBL Grenoble and Professor Axel Scheidig from University of Kiel for their support during the thesis work.

I'm very thankful for all the amazing friends that I acquired during the time I was at EMBL. First, I want to dedicate a special thank to a former EMBL member Jon Rapley, a great office mate, colleague and friend of mine, who taught me a lot both in and outside of the wetlab. I enjoyed the many conversations that we had about life and life sciences. I also want to thank my lovely office girls Diana, Natasha and Selina for being always so supportive and creating such a warmhearted and friendly working environment. I owe a big thank you also for all the EMBL Hamburg wetlab people who helped me with various problems and made the wetlab a great place to work. Especially thanks to Anna, Nina, Alex, Magda, Matthew and Sandra for their support during this thesis project and also for all the great nights out. Most of all I want to thank my two dear friends Anne and Jana, who are always there for me. I value our friendship more than words can express.

I want to acknowledge EMBL for providing the funding and excellent scientific environment. I would also like to acknowledge the synchrotron facility and the beam line scientist at EMBL Hamburg. I especially want to thank Thomas Schneider for providing

the crystallographic beam lines (P13 and P14) and Dmitri Svergun for providing the SAXS beam line P12 for the use in this thesis work. I also want to acknowledge Gleb Bourenkov, Michele Cianti, Johanna Kallio and Guillaume Pompidor for their help and advices at the crystallographic beam lines. I'm also very grateful to Anne Tuukkanen for the great help during the SAXS measurements and also for the data analysis. Regarding the ApnA A_1 PCC7821 project, I would like to thank Professor Andrea Rentmeister and Guntram Christiansen for the fruitful collaboration. Additionally, I want to acknowledge Stephan Schiefelbein, Julia Sandberg-Meinhardt and Sandra Kozak for their work on the ApnA A_1 PCC7821 project.

Last but not least I want to thank my family for all their support throughout my whole life. I would not be here if it wasn't for you.

LIST OF ABBREVIATIONS

A domain	Adenylation domain
ACC	1-aminocyclopropane-1-carboxylic acid
ACO	ACC oxidase
ACS	ACC synthase
Acs	Acetyl-CoA synthase
ACV	δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine
ATP	Adenosine triphosphate
BIIC	baculovirus infected insect cell
BSA	Bovine serum albumin
CA	catalytic ATP-binding
CoA	Coenzyme A
CTR1	Constitutive triple response 1
DAMP	Damage-associated molecular patterns
DHB	2,3-dihydroxybenzoate
DhbE	2,3-dihydroxybenzoate (DHB) activating adenylation domain
DHp	dimerization histidine phosphotransfer
DltA	D-Alanine carrier protein
dpa	day of proliferation arrest
EBF1 & 2	EIN3-targeting F-box proteins 1 & 2
EDR1	Enhanced disease resistance 1
EIL	EIN3-like
EIN2	Ethylene Insensitive 2
EIN3	Ethylene Insensitive 3

EIN4	Ethylene Insensitive 4
EOL1	ETO1-like1
EOM	Ensemble optimization method
EREBP	Ethylene-Response-Element-Binding-Protein
erf1	ethylene response factor 1
ERS1 & 2	Ethylene Response Sensor 1 & 2
ESI	electrospray ionization
ETP1 & 2	EIN2 targeting protein 1 & 2
ETO1	Ethylene overproducer 1
ETR1 & 2	Ethylene receptor 1 & 2
ETR1-ΔTM	The cytosolic portion of ETR1
GAF	cGMP-specific phosphodiesterase, adenylyl cyclases and FhlA
GFP	Green Fluorescence Protein
HR	Hypersensitive response
IPTG	isopropyl-β-D-thiogalactopyranoside
ISR	Induced systemic resistance
LB	Luria-Bertani
MALDI	matrix-assisted laser desorption/ionization
MAMP	Microbe-associated molecular patterns
MAPK	Mitogen-activated protein kinase
MAPKKK	Mitogen-activated protein kinase kinase kinase
MS	mass spectrometry
MTA	5'-methylthioadenosine
NADPH	Nicotinamide adenine dinucleotide phosphate
NDSB	non-detergent sulfobetaine

NIS	NRPS-independent siderophores
NMR	Nuclear Magnetic Resonance
NRAMP	Natural resistance-associated macrophage protein
NRPS	Non-ribosomal peptide synthetase
PAMP	Pathogen associated molecular patterns
PCP	Peptide carrier protein
PCR	Polymerase Chain Reaction
PEG	Polyethylene glycol
PheA	Phenylalanine adenylation subunit of Gramicidin S synthetase 1
PPTase	Phosphopanthetheinyl transferase
PR	Pathogenesis-related
PRR	Pattern recognition receptors
PTI	PAMP-triggered immunity
RAN1	Responsive-to-Antagonist 1
SA	Salicylic acid
SAR	Systemic acquired resistance
SAXS	Small Angle X-ray Scattering
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
SOC	Super optimal broth with catabolite repression
S-Adomet	S-Adenosyl-L-methionine
TB	Terrififc Broth
YFP	Yellow Fluorescence Protein

Amino acids

Alanine	Ala	A
Cysteine	Cys	C
Aspartate	Asp	D
Glutamate	Glu	E
Phenylalanine	Phe	F
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Leucine	Leu	L
Methionine	Met	M
Asparagine	Asn	N
Proline	Pro	P
Glutamine	Gln	Q
Arginine	Arg	R
Serine	Ser	S
Threonine	Thr	T
Valine	Val	V
Tryptophan	Trp	W
Tyrosine	Tyr	Y

BIBLIOGRAPHY

Abel, S., Nguyen, M.D., Chow, W. & Theologis, A. (1995) ACS4, a primary indoleacetic acid-responsive gene encoding 1-aminocyclopropane-1-carboxylate synthase in Arabidopsis thaliana. Structural characterization, expression in Escherichia coli, and expression characteristics in response to auxin [corrected]. *J Biol Chem*, **270**, 19093-19099.

Adams, D.O. & Yang, S.F. (1977) Methionine metabolism in apple tissue: implication of s-adenosylmethionine as an intermediate in the conversion of methionine to ethylene. *Plant Physiol*, **60**, 892-896.

Adams, D.O. & Yang, S.F. (1979) Ethylene biosynthesis: Identification of 1-aminocyclopropane-1-carboxylic acid as an intermediate in the conversion of methionine to ethylene. *Proc Natl Acad Sci U S A*, **76**, 170-174.

Aebersold, R. & Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*, **422**, 198-207.

Ahuja, I., Kissen, R. & Bones, A.M. (2012) Phytoalexins in defense against pathogens. *Trends Plant Sci*, **17**, 73-90.

Alexander, L. & Grierson, D. (2002) Ethylene biosynthesis and action in tomato: a model for climacteric fruit ripening. *J Exp Bot*, **53**, 2039-2055.

Alonso, J.M., Hirayama, T., Roman, G., Nourizadeh, S. & Ecker, J.R. (1999) EIN2, a bifunctional transducer of ethylene and stress responses in Arabidopsis. *Science*, **284**, 2148-2152.

Babbitt, P.C., Kenyon, G.L., Martin, B.M., Charest, H., Slyvestre, M., Scholten, J.D., Chang, K.H., Liang, P.H. & Dunaway-Mariano, D. (1992) Ancestry of the 4-chlorobenzoate dehalogenase: analysis of amino acid sequence identities among families of acyl:adenyl ligases, enoyl-CoA hydratases/isomerases, and acyl-CoA thioesterases. *Biochemistry*, **31**, 5594-5604.

Bednarek, P. & Osbourn, A. (2009) Plant-microbe interactions: chemical diversity in plant defense. *Science*, **324**, 746-748.

Belshaw, P.J., Walsh, C.T. & Stachelhaus, T. (1999) Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*, **284**, 486-489.

Bennett, R.N. & Wallsgrove, R.M. (1994) Secondary metabolites in plant defence mechanisms. *New Phytologist*, **127**, 617-633.

Berger, I., Fitzgerald, D.J. & Richmond, T.J. (2004) Baculovirus expression system for heterologous multiprotein complexes. *Nat Biotechnol*, **22**, 1583-1587.

Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. & Svergun, D.I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc*, **129**, 5656-5664.

Bieniossek, C., Richmond, T.J. & Berger, I. (2008) MultiBac: multigene baculovirus-based eukaryotic protein complex production. *Curr Protoc Protein Sci*, **Chapter 5**, Unit 5.20.

Bierer, B.E., Hollander, G., Fruman, D. & Burakoff, S.J. (1993) Cyclosporin A and FK506: molecular mechanisms of immunosuppression and probes for transplantation biology. *Curr Opin Immunol*, **5**, 763-773.

Bisson, M.M. & Groth, G. (2010) New insight in ethylene signaling: autokinase activity of ETR1 modulates the interaction of receptors and EIN2. *Mol Plant*, **3**, 882-889.

Bleecker, A.B. & Kende, H. (2000) Ethylene: a gaseous signal molecule in plants. *Annu Rev Cell Dev Biol*, **16**, 1-18.

Blom, N., Gammeltoft, S. & Brunak, S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol*, **294**, 1351-1362.

Boller, T. & He, S.Y. (2009) Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens. *Science*, **324**, 742-744.

Challis, G.L. & Naismith, J.H. (2004) Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struct Biol*, **14**, 748-756.

Chang, C. & Bleecker, A.B. (2004) Ethylene biology. More than a gas. *Plant Physiol*, **136**, 2895-2899.

Chang, C., Kwok, S.F., Bleecker, A.B. & Meyerowitz, E.M. (1993) Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators. *Science*, **262**, 539-544.

Chang, C. & Stadler, R. (2001) Ethylene hormone receptor action in Arabidopsis. *Bioessays*, **23**, 619-627.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. & Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **66**, 12-21.

Chen, Y.F., Randlett, M.D., Findell, J.L. & Schaller, G.E. (2002) Localization of the ethylene receptor ETR1 to the endoplasmic reticulum of Arabidopsis. *J Biol Chem*, **277**, 19861-19866.

Christiansen, G., Philmus, B., Hemscheidt, T. & Kurmayer, R. (2011) Genetic variation of adenylation domains of the anabaenopeptin synthesis operon and evolution of substrate promiscuity. *J Bacteriol*, **193**, 3822-3831.

Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **50**, 760-763.

Conti, E., Franks, N.P. & Brick, P. (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure*, **4**, 287-298.

Conti, E., Stachelhaus, T., Marahiel, M.A. & Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *Embo j*, **16**, 4174-4183.

Cook, P.F., Neville, M.E., Jr., Vrana, K.E., Hartl, F.T. & Roskoski, R., Jr. (1982) Adenosine cyclic 3',5'-monophosphate dependent protein kinase: kinetic mechanism for the bovine skeletal muscle catalytic subunit. *Biochemistry*, **21**, 5794-5799.

Cox, J. & Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech*, **26**, 1367-1372.

de Wit, P.J. (2007) How plants recognize pathogens and defend themselves. *Cell Mol Life Sci*, **64**, 2726-2732.

Du, L., He, Y. & Luo, Y. (2008) Crystal structure and enantiomer selection by D-alanyl carrier protein ligase DltA from Bacillus cereus. *Biochemistry*, **47**, 11473-11480.

Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, **66**, 486-501.

Eppelmann, K., Stachelhaus, T. & Marahiel, M.A. (2002) Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry*, **41**, 9718-9726.

Eschenfeldt, W.H., Lucy, S., Millard, C.S., Joachimiak, A. & Mark, I.D. (2009) A family of LIC vectors for high-throughput cloning and purification of proteins. *Methods Mol Biol*, **498**, 105-115.

Finking, R. & Marahiel, M.A. (2004) Biosynthesis of nonribosomal peptides1. *Annu Rev Microbiol*, **58**, 453-488.

Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J. & Berger, I. (2006) Protein complex expression by using multigene baculoviral vectors. *Nat Methods*, **3**, 1021-1032.

Frye, C.A. & Innes, R.W. (1998) An Arabidopsis mutant with enhanced resistance to powdery mildew. *Plant Cell*, **10**, 947-956.

Frye, C.A., Tang, D. & Innes, R.W. (2001) Negative regulation of defense responses in plants by a conserved MAPKK kinase. *Proc Natl Acad Sci U S A*, **98**, 373-378.

Fu, Z.Q. & Dong, X. (2013) Systemic acquired resistance: turning local infection into global defense. *Annu Rev Plant Biol*, **64**, 839-863.

Gaspar, T., Kevers, C., Penel, C., Greppin, H., Reid, D. & Thorpe, T. (1996) Plant hormones and plant growth regulators in plant tissue culture. *In Vitro Cellular & Developmental Biology - Plant*, **32**, 272-289.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.e., Wilkins, M., Appel, R. & Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In Walker, J. (ed) *The Proteomics Protocols Handbook*. Humana Press, pp. 571-607.

Gewolb, J. (2002) Bioengineering. Working outside the protein-synthesis rules. *Science*, **295**, 2205-2207.

Gomez-Gomez, L. & Boller, T. (2000) FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. *Mol Cell*, **5**, 1003-1011.

Gu, Y. & Innes, R.W. (2011) The KEEP ON GOING protein of Arabidopsis recruits the ENHANCED DISEASE RESISTANCE1 protein to trans-Golgi network/early endosome vesicles. *Plant Physiol*, **155**, 1827-1838.

Gulick, A.M. (2009) Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem Biol*, **4**, 811-827.

Gulick, A.M., Starai, V.J., Horswill, A.R., Homick, K.M. & Escalante-Semerena, J.C. (2003) The 1.75 A crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A. *Biochemistry*, **42**, 2866-2873.

Guo, H. & Ecker, J.R. (2004) The ethylene signaling pathway: new insights. *Curr Opin Plant Biol*, **7**, 40-49.

Guzman, P. & Ecker, J.R. (1990) Exploiting the triple response of Arabidopsis to identify ethylene-related mutants. *Plant Cell*, **2**, 513-523.

Hirayama, T. & Alonso, J.M. (2000) Ethylene captures a metal! Metal ions are involved in ethylene perception and signal transduction. *Plant Cell Physiol*, **41**, 548-555.

Hirayama, T., Kieber, J.J., Hirayama, N., Kogan, M., Guzman, P., Nourizadeh, S., Alonso, J.M., Dailey, W.P., Dancis, A. & Ecker, J.R. (1999) RESPONSIVE-TO-ANTAGONIST1, a Menkes/Wilson disease-related copper transporter, is required for ethylene signaling in Arabidopsis. *Cell*, **97**, 383-393.

Holm, L. & Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res*, **38**, W545-549.

Hu, H., Xue, J., Swarts, B.M., Wang, Q., Wu, Q. & Guo, Z. (2009) Synthesis and antibacterial activities of N-glycosylated derivatives of tyrocidine A, a macrocyclic peptide antibiotic. *J Med Chem*, **52**, 2052-2059.

Hua, J., Sakai, H., Nourizadeh, S., Chen, Q.G., Bleecker, A.B., Ecker, J.R. & Meyerowitz, E.M. (1998) EIN4 and ERS2 are members of the putative ethylene receptor gene family in Arabidopsis. *Plant Cell*, **10**, 1321-1332.

Hur, G.H., Vickery, C.R. & Burkart, M.D. (2012) Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep*, **29**, 1074-1098.

Huse, M. & Kuriyan, J. (2002) The conformational plasticity of protein kinases. *Cell*, **109**, 275-282.

Jabs, T., Tschope, M., Colling, C., Hahlbrock, K. & Scheel, D. (1997) Elicitor-stimulated ion fluxes and O2- from the oxidative burst are essential components in triggering defense gene activation and phytoalexin synthesis in parsley. *Proc Natl Acad Sci U S A*, **94**, 4800-4805.

Johnson, P.R. & Ecker, J.R. (1998) The ethylene gas signal transduction pathway: a molecular perspective. *Annu Rev Genet*, **32**, 227-254.

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

Kabsch, W. (2010) XDS. *Acta Crystallogr D Biol Crystallogr*, **66**, 125-132.

Kadi, N. & Challis, G.L. (2009) Chapter 17. Siderophore biosynthesis a substrate specificity assay for nonribosomal peptide synthetase-independent siderophore synthetases involving trapping of acyl-adenylate intermediates with hydroxylamine. *Methods Enzymol*, **458**, 431-457.

Kantardjieff, K.A. & Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci*, **12**, 1865-1871.

Keating, T.A., Marshall, C.G., Walsh, C.T. & Keating, A.E. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol*, **9**, 522-526.

Kende, H. (1993) Ethylene Biosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology*, **44**, 283-307.

Kende, H. & Boller, T. (1981) Wound ethylene and 1-aminocyclopropane-1-carboxylate synthase in ripening tomato fruit. *Planta*, **151**, 476-481.

Kieber, J.J., Rothenberg, M., Roman, G., Feldmann, K.A. & Ecker, J.R. (1993) CTR1, a negative regulator of the ethylene response pathway in Arabidopsis, encodes a member of the raf family of protein kinases. *Cell*, **72**, 427-441.

Kleinkauf, H. (1979) Antibiotic polypeptides--biosynthesis on multifunctional protein templates. *Planta Med*, **35**, 1-18.

Kleinkauf, H. & von Dohren, H. (1990) Nonribosomal biosynthesis of peptide antibiotics. *Eur J Biochem*, **192**, 1-15.

Koglin, A., Mofid, M.R., Lohr, F., Schafer, B., Rogov, V.V., Blum, M.M., Mittag, T., Marahiel, M.A., Bernhard, F. & Dotsch, V. (2006) Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science*, **312**, 273-276.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. & Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, **36**, 1277-1282.

Konz, D., Klens, A., Schorgendorfer, K. & Marahiel, M.A. (1997) The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem Biol*, **4**, 927-937.

Konz, D. & Marahiel, M.A. (1999) How do peptide synthetases generate structural diversity? *Chem Biol*, **6**, R39-48.

Krissinel, E. & Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, **372**, 774-797.

Kunz, J. & Hall, M.N. (1993) Cyclosporin A, FK506 and rapamycin: more than just immunosuppression. *Trends Biochem Sci*, **18**, 334-338.

Lai, J.R., Koglin, A. & Walsh, C.T. (2006) Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis. *Biochemistry*, **45**, 14869-14879.

Langer, G., Cohen, S.X., Lamzin, V.S. & Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc*, **3**, 1171-1179.

Lautru, S. & Challis, G.L. (2004) Substrate recognition by nonribosomal peptide synthetase multi-enzymes. *Microbiology*, **150**, 1629-1636.

Lebedev, A.A., Young, P., Isupov, M.N., Moroz, O.V., Vagin, A.A. & Murshudov, G.N. (2012) JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D Biol Crystallogr*, **68**, 431-440.

Lee, B.N., Kroken, S., Chou, D.Y., Robbertse, B., Yoder, O.C. & Turgeon, B.G. (2005) Functional analysis of all nonribosomal peptide synthetases in Cochliobolus heterostrophus reveals a factor, NPS6, involved in virulence and resistance to oxidative stress. *Eukaryot Cell*, **4**, 545-555.

Lin, Z., Zhong, S. & Grierson, D. (2009) Recent advances in ethylene research. *J Exp Bot*, **60**, 3311-3336.

Linne, U. & Marahiel, M.A. (2000) Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry*, **39**, 10439-10447.

Lipmann, F. & Tuttle, L.C. (1945) The detection of activated carboxyl groups with hydroxylamine as interceptor. *J Biol Chem*, **161**, 415.

Marahiel, M.A. (1997) Protein templates for the biosynthesis of peptide antibiotics. *Chem Biol*, **4**, 561-567.

Marahiel, M.A. & Essen, L.O. (2009) Chapter 13. Nonribosomal peptide synthetases mechanistic and structural aspects of essential domains. *Methods Enzymol*, **458**, 337-351.

Marahiel, M.A., Stachelhaus, T. & Mootz, H.D. (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev*, **97**, 2651-2674.

Mason, M.G. & Schaller, G.E. (2005) Histidine kinase activity and the regulation of ethylene signal transduction. *Canadian Journal of Botany*, **83**, 563-570.

Matthews, B.W. (1968) Solvent content of protein crystals. *J Mol Biol*, **33**, 491-497.

May, J.J., Kessler, N., Marahiel, M.A. & Stubbs, M.T. (2002) Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci U S A*, **99**, 12120-12125.

Mayerhofer, H. (2011). Structural studies of the Arabidopsis thaliana ethylene signal transduction pathway (Dissertation), Heidelberg University. Retrieved from http://www.ub.uni-heidelberg.de/archiv/12902

Mayerhofer, H., Panneerselvam, S. & Mueller-Dieckmann, J. (2012) Protein kinase domain of CTR1 from Arabidopsis thaliana promotes ethylene receptor cross talk. *J Mol Biol*, **415**, 768-779.

McElroy, W.D., DeLuca, M. & Travis, J. (1967) Molecular uniformity in biological catalyses. The enzymes concerned with firefly luciferin, amino acid, and fatty acid utilization are compared. *Science*, **157**, 150-160.

Mieczkowski, C., Iavarone, A.T. & Alber, T. (2008) Auto-activation mechanism of the Mycobacterium tuberculosis PknB receptor Ser/Thr kinase. *Embo j*, **27**, 3186-3197.

Miroux, B. & Walker, J.E. (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol*, **260**, 289-298.

Mootz, H.D., Schwarzer, D. & Marahiel, M.A. (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem*, **3**, 490-504.

Moussatche, P. & Klee, H.J. (2004) Autophosphorylation activity of the Arabidopsis ethylene receptor multigene family. *J Biol Chem*, **279**, 48734-48741.

Mueller-Dieckmann, J. (2006) The open-access high-throughput crystallization facility at EMBL Hamburg. *Acta Crystallogr D Biol Crystallogr*, **62**, 1446-1452.

Muller-Dieckmann, H.J., Grantz, A.A. & Kim, S.H. (1999) The structure of the signal receiver domain of the Arabidopsis thaliana ethylene receptor ETR1. *Structure*, **7**, 1547-1556.

Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, **53**, 240-255.

Muthamilarasan, M. & Prasad, M. (2013) Plant innate immunity: an updated insight into defense mechanism. *J Biosci*, **38**, 433-449.

Nagarajan, R. (1991) Antibacterial activities and modes of action of vancomycin and related glycopeptides. *Antimicrob Agents Chemother*, **35**, 605-609.

Nolen, B., Taylor, S. & Ghosh, G. (2004) Regulation of protein kinases; controlling activity through activation segment conformation. *Mol Cell*, **15**, 661-675.

Novelli, G.D. (1967) Amino acid activation for protein synthesis. *Annu Rev Biochem*, **36**, 449-484.

Nurnberger, T., Brunner, F., Kemmerling, B. & Piater, L. (2004) Innate immunity in plants and animals: striking similarities and obvious differences. *Immunol Rev*, **198**, 249-266.

Olmedo, G., Guo, H., Gregory, B.D., Nourizadeh, S.D., Aguilar-Henonin, L., Li, H., An, F., Guzman, P. & Ecker, J.R. (2006) ETHYLENE-INSENSITIVE5 encodes a 5'-->3' exoribonuclease required for regulation of the EIN3-targeting F-box proteins EBF1/2. *Proc Natl Acad Sci U S A*, **103**, 13286-13293.

Petoukhov, M.V., Konarev, P.V., Kikhney, A.G. & Svergun, D.I. (2007) ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. *Journal of Applied Crystallography*, **40**, s223-s228.

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

Pieterse, C.M., Leon-Reyes, A., Van der Ent, S. & Van Wees, S.C. (2009) Networking by small-molecule hormones in plant immunity. *Nat Chem Biol*, **5**, 308-316.

Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr*, **59**, 1131-1137.

Rajakulendran, T., Sahmi, M., Lefrancois, M., Sicheri, F. & Therrien, M. (2009) A dimerization-dependent mechanism drives RAF catalytic activation. *Nature*, **461**, 542-545.

Ramos, H.C., Rumbo, M. & Sirard, J.C. (2004) Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends Microbiol*, **12**, 509-517.

Ravanel, S., Gakiere, B., Job, D. & Douce, R. (1998) The specific features of methionine biosynthesis and metabolism in plants. *Proc Natl Acad Sci U S A*, **95**, 7805-7812.

Riov, J. & Yang, S.F. (1982) Autoinhibition of Ethylene Production in Citrus Peel Discs : SUPPRESSION OF 1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID SYNTHESIS. *Plant Physiol*, **69**, 687-690.

Roessle, M.W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D., Gehrmann, T., Konarev, P., Round, A., Fiedler, S., Hermes, C. & Svergun, D. (2007) Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *Journal of Applied Crystallography*, **40**, s190-s194.

Roman, G., Lubarsky, B., Kieber, J.J., Rothenberg, M. & Ecker, J.R. (1995) Genetic analysis of ethylene signal transduction in Arabidopsis thaliana: five novel mutant loci integrated into a stress response pathway. *Genetics*, **139**, 1393-1409.

Sahl, H.G., Jack, R.W. & Bierbaum, G. (1995) Biosynthesis and biological activities of lantibiotics with unique post-translational modifications. *Eur J Biochem*, **230**, 827-853.

Samel, S.A., Schoenafinger, G., Knappe, T.A., Marahiel, M.A. & Essen, L.O. (2007) Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure*, **15**, 781-792.

Santner, A., Calderon-Villalobos, L.I. & Estelle, M. (2009) Plant hormones are versatile chemical regulators of plant growth. *Nat Chem Biol*, **5**, 301-307.

Schaller, G.E., Ladd, A.N., Lanahan, M.B., Spanbauer, J.M. & Bleecker, A.B. (1995) The ethylene response mediator ETR1 from Arabidopsis forms a disulfide-linked dimer. *J Biol Chem*, **270**, 12526-12530.

Scheel, D. (1998) Resistance response physiology and signal transduction. *Curr Opin Plant Biol*, **1**, 305-310.

Schmelz, S. & Naismith, J.H. (2009) Adenylate-forming enzymes. *Curr Opin Struct Biol*, **19**, 666-671.

Schwarzer, D., Mootz, H.D., Linne, U. & Marahiel, M.A. (2002) Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc Natl Acad Sci U S A*, **99**, 14083-14088.

Shen, Q.T., Chen, X.L., Sun, C.Y. & Zhang, Y.Z. (2004) Dissecting and exploiting nonribosomal peptide synthetases. *Acta Biochim Biophys Sin (Shanghai)*, **36**, 243-249.

Sieber, S.A. & Marahiel, M.A. (2003) Learning from nature's drug factories: nonribosomal synthesis of macrocyclic peptides. *J Bacteriol*, **185**, 7036-7043.

Sieber, S.A. & Marahiel, M.A. (2005) Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev*, **105**, 715-738.

Stachelhaus, T. & Marahiel, M.A. (1995) Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. *FEMS Microbiol Lett*, **125**, 3-14.

Stachelhaus, T., Mootz, H.D. & Marahiel, M.A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol*, **6**, 493-505.

Stein, T., Vater, J., Kruft, V., Otto, A., Wittmann-Liebold, B., Franke, P., Panico, M., McDowell, R. & Morris, H.R. (1996) The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *J Biol Chem*, **271**, 15428-15435.

Strieker, M., Tanovic, A. & Marahiel, M.A. (2010) Nonribosomal peptide synthetases: structures and dynamics. *Curr Opin Struct Biol*, **20**, 234-240.

Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*, **41**, 207-234.

Suzuki, H., Nagai, K., Yamaki, H., Tanaka, N. & Umezawa, H. (1968) Mechanism of action of bleomycin. Studies with the growing culture of bacterial and tumor cells. *J Antibiot (Tokyo)*, **21**, 379-386.

Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography*, **25**, 495-503.

Svergun, D., Barberato, C. & Koch, M.H.J. (1995) CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography*, **28**, 768-773.

Svergun, D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J*, **76**, 2879-2886.

Tang, D., Christiansen, K.M. & Innes, R.W. (2005) Regulation of plant disease resistance, stress responses, cell death, and ethylene signaling in Arabidopsis by the EDR1 protein kinase. *Plant Physiol*, **138**, 1018-1026.

Tans-Kersten, J., Huang, H. & Allen, C. (2001) Ralstonia solanacearum needs motility for invasive virulence on tomato. *J Bacteriol*, **183**, 3597-3605.

Tsuchisaka, A. & Theologis, A. (2004) Unique and overlapping expression patterns among the Arabidopsis 1-amino-cyclopropane-1-carboxylate synthase gene family members. *Plant Physiol*, **136**, 2982-3000.

Turgay, K., Krause, M. & Marahiel, M.A. (1992a) Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol Microbiol*, **6**, 2743-2744.

Turgay, K., Krause, M. & Marahiel, M.A. (1992b) Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol Microbiol*, **6**, 529-546.

Vagenende, V., Yap, M.G. & Trout, B.L. (2009) Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry*, **48**, 11084-11096.

Vagin, A. & Teplyakov, A. (1997) MOLREP: an Automated Program for Molecular Replacement. *Journal of Applied Crystallography*, **30**, 1022-1025.

Voet-van-Vormizeele, J. & Groth, G. (2003) High-level expression of the Arabidopsis thaliana ethylene receptor protein ETR1 in Escherichia coli and purification of the recombinant protein. *Protein Expr Purif*, **32**, 89-94.

Vogel, J.P., Woeste, K.E., Theologis, A. & Kieber, J.J. (1998) Recessive and dominant mutations in the ethylene biosynthetic gene ACS5 of Arabidopsis confer cytokinin insensitivity and ethylene overproduction, respectively. *Proc Natl Acad Sci U S A*, **95**, 4766-4771.

Wagner, J.R., Zhang, J., Brunzelle, J.S., Vierstra, R.D. & Forest, K.T. (2007) High resolution structure of Deinococcus bacteriophytochrome yields new insights into phytochrome architecture and evolution. *J Biol Chem*, **282**, 12298-12309.

Walsh, C.T. (2004) Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science*, **303**, 1805-1810.

Walsh, C.T. (2008) The chemical versatility of natural-product assembly lines. *Acc Chem Res*, **41**, 4-10.

Wang, J., Wu, J.W. & Wang, Z.X. (2011) Structural insights into the autoactivation mechanism of p21-activated protein kinase. *Structure*, **19**, 1752-1761.

Wang, K.L., Li, H. & Ecker, J.R. (2002) Ethylene biosynthesis and signaling networks. *Plant Cell*, **14 Suppl**, S131-151.

Wang, X., Kong, H. & Ma, H. (2009) F-box proteins regulate ethylene signaling and more. *Genes Dev*, **23**, 391-396.

Wawrzynska, A., Christiansen, K.M., Lan, Y., Rodibaugh, N.L. & Innes, R.W. (2008) Powdery mildew resistance conferred by loss of the ENHANCED DISEASE RESISTANCE1 protein kinase is suppressed by a missense mutation in KEEP ON GOING, a regulator of abscisic acid signaling. *Plant Physiol*, **148**, 1510-1522.

Wen, J., Arakawa, T. & Philo, J.S. (1996) Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. *Anal Biochem*, **240**, 155-166.

Williams, D.H., Stone, M.J., Hauck, P.R. & Rahman, S.K. (1989) Why are secondary metabolites (natural products) biosynthesized? *J Nat Prod*, **52**, 1189-1208.

Wittmann, M., Linne, U., Pohlmann, V. & Marahiel, M.A. (2008) Role of DptE and DptF in the lipidation reaction of daptomycin. *Febs j*, **275**, 5343-5354.

Yang, S.F. & Hoffman, N.E. (1984) Ethylene Biosynthesis and its Regulation in Higher Plants. *Annual Review of Plant Physiology*, **35**, 155-189.

Yonus, H., Neumann, P., Zimmermann, S., May, J.J., Marahiel, M.A. & Stubbs, M.T. (2008) Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. *J Biol Chem*, **283**, 32484-32491.

Zhao, C., Nie, H., Shen, Q., Zhang, S., Lukowitz, W. & Tang, D. (2014) EDR1 physically interacts with MKK4/MKK5 and negatively regulates a MAP kinase cascade to modulate plant innate immunity. *PLoS Genet*, **10**, e1004389.

Zhao, Q. & Guo, H.W. (2011) Paradigms and paradox in the ethylene signaling pathway and interaction network. *Mol Plant*, **4**, 626-634.

Zipfel, C. & Felix, G. (2005) Plants and animals: a different taste for microbes? *Curr Opin Plant Biol*, **8**, 353-360.

LIST OF FIGURES AND TABLES

APPENDIX

Appendix 1 – The table of primers

Table A1. The list of primers used to clone the main constructs. The abbreviation ETR1s stands for the synthetic gene of the receptor. The overlap with the inserted gene is underlined.

Primer Type	Target*	Cloning site	Primer sequence (5' → 3')
Forward	ETR1	LIC	CAGGGCGCCATGATGGAAGTCTGCAATTGTATTGAACCG
Reverse	ETR1	LIC	GACCCGACGCGGTTACATGCCCTCGTACAGTACCCG
Forward	ETR1	NdeI	CATATGGAAGTCTGCAATTGTATTGAACCGC
Reverse	ETR1	BamHI	GGATCCTTATTACATGCCCTCGTACAGTACCCG
Forward	ETR1	NcoI	CCATGGAAGTCTGCAATTGTATTGAACCGC
Reverse	ETR1s	LIC	GACCCGACGCGGTTATTACATGCCTTCATACAGCACACGCGG
Forward	ETR1s	LIC	CAGGGCGCCATGGAAGTGTGTAATTGCATTGAACCGC
Forward	ETR1s	NdeI	GATCATATGGAAGTGTGTAATTGCATTGAACCGC
Reverse	ETR1s	BamHI	GACAGGATCCTCATTACATGCCTTCATACAGCACAC
Forward	ETR1s	NcoI	GATCCATGGAAGTGTGTAATTGCATTGAACC
Forward	EDR1	LIC	TACTTCCAATCCAATGCCGATGATGCAGATGTTGGTGAATGTGAAATTCC
Reverse	EDR1	LIC	TTATCCACTTCCAATGTTATTGTGGTGTAGGAAGTACAAGCCGGTTC

* ETR1 gene with NCBI accession no. NM_105305 and EDR1 gene with NCBI accession no. NM_100745 used for primer design.

Appendix 2 – Nucleotide sequence for the ApnA A_1 PCC7821 construct

```
1       ATGGTCAAGC AATCTATCCA CCAGTTATTT GAAACCCAAG TTGAACGCAC    50
51      ACCAGAAGCA GTGGCCGTCC TTTCAGAACA AGGGCAATTG ACCTATGAGG   100
101     AATTAAATAC TAAAGCGAAT CAACTAGCCC ATTATTTGCG AACCCTGGGG   150
151     GTTAAGTCCG AGACATTGGT GGGAGTTTGT GTTGATCGCT CCCTAGAGAT   200
201     GGTCATTGGC TTATTGGCTA TTCTCAAGGC GGGGGGAGCC TATGTGCCGC   250
251     TCGATCCAAC TTATCCCAGG GAAAGATTGA CCTACATGGT GCAGGATGCT   300
301     CAAATATCCG TGTTGGTCAC GCAAACGCAA TGGTCTAACT TAATTTCCGA   350
351     CTATCAAGGA CAGGTGATTT GCTTAGATAG TCAATGGGCA AAAATTGCCA   400
401     GTTATAGTCA AGAAAATCTG GTTAATACGG TAAATCCTGA GAATTTAGCC   450
451     TATGTCATCT ATACATCGGG ATCAACGGGA AAACCCAAGG GGGTAATGAT   500
501     TGAACACCAA TCCTTGGTAA ACTTCACTAA ACTAGCGATC GCCCAATATC   550
551     AGATCACGAC AAGCGATCGC ACTCTTCAAT TTGTCTCTAT CAGTTTTGAT   600
601     GTGGCCGCAG AGGAAATCTA TGTTACTCTT TGTTCTGGTG CTACCTTAAT   650
651     CTTACGAACT GAAGAAATGA TCAGTTCTAT TCCTTCATTT GTTCAAAAAT   700
701     CCCAAGATTG GCAAATAACT GTTTGGAGCT TACCCACCGC TTACTGGCAT   750
751     TTATTAGTCA ATGAATTAGT TAAGAGTAAA ATAGCATTAC CAGACAGCTT   800
801     GCGATTAGTC ATTATCGGCG GAGAAAGAGT TCAACCAGAA TTGGTGAGAA   850
851     TGTGGTTTAA AAATGTAGGA AATTTTCCTG AATTAATTAA TGTTTATGGC   900
901     CCTACGGAAG GGACAATTGC GGTAAGCCTT TGTCGCCTAT CACAACTAAC   950
951     AGAAAGTCAA CGCAACAGGA CAGAAATTCC TATTGGAAAG AGTTTAGGAG  1000
1001    AAAATATTTC AGTTTATGTT TTAGATGAAA CCTTAAAAAC CGTTCCTCCT  1050
1051    GAAACACCAG GAGAAATTTA TATCGGGGGT ACTGCGCTTG CCAGGGGTTA  1100
1101    TTTGAATCGT CCAGAATTAA CTGCTCAAAA ATTTATTCAA GATCCCTTTA  1150
1151    GCCCCTCAGA AAGATTATAT AAAACCGGAG ATTTAGGCAG ATATTTAGCC  1200
1201    GACGGTAATT TAGAGTATTT AGGAAGAGTC GATCATCAAG TTAAAATCAA  1250
1251    CGGTTTTAGA GTTGAATTAG GAGAAATTGA AACCGTTTTA CTTCAACATC  1300
1301    ATCAAGTCGC TCAAGCCGTT GTTATTGATC GAGAAGATCC TCTGGGAAAT  1350
1351    AAAAGATTAG TTGCTTATCT TGTTCCCCAT TCAACCGAGG AAAATTTAAC  1400
1401    CGTAACTCTG CAACAATTTT TGAAGAATAA GCTTCCGAGC TATATGATTC  1450
1451    CTGCGACTTT TGTTGTATTA AATGAGTTGC CCCTTAGTCC CAATGGAAAA  1500
1501    ATTGATCGTA AAGCTTTACC CATACCAGAC TATGACGGAA ACGAGCGTCA  1550
1551    AACCCCCTTT ATTGCCCCTC GTAATCACCA AGAAGAAAAA TTAGCAAATA  1600
1601    TTTGGCATCA GGTATTTGGA TTAGAAAAAA TTGGGGTCTA A          1641
```

Appendix 3 – The Thermofluor screens of EMBL Hamburg

Table A2. Thermofluor Buffer screen developed at EMBL Hamburg (Boivin *et al.* 2013).

	1	2	3	4	5	6	7	8	9	10	11	12
A	Water	Citric acid pH 4.0	Na acetate pH 4.5	Citric acid pH 5.0	MES pH 6.0	K Phosphate pH 6.0	Citric acid pH 6.0	Bis-Tris pH 6.5	Na Cacodylate pH 6.5	Na Phosphate pH 7.0	K Phosphate pH 7.0	HEPES pH 7.0
B	MOPS pH 7.0	Ammonium acetate Ph7.3	Tris-HCl pH 7.5	Na Phosphate pH 7.5	Imidazole pH 8.0	HEPES pH 8.0	Tris-HCl pH 8.0	Tricine pH 8.0	Bicine pH 8.0	Bicine pH 8.5	Tris-HCl pH8.5	Bicine pH 9.0
C	250 mM NaCl	Citric acid 250 mM NaCl pH 4.0	Na acetate 250 mM NaCl pH 4.5	Citric acid 250 mM NaCl pH 5.0	MES 250 mM NaCl pH 6.0	K Phosphate 250 mM NaCl pH 6.0	Citric acid 250 mM NaCl pH 6.0	Bis-Tris 250 mM NaCl pH 6.5	Na cacodylate 250 mM NaCl pH 6.5	Na phosphate 250 mM NaCl pH 7.0	K Phosphate 250 mM NaCl pH 7.0	HEPES 250 mM NaCl pH 7.0
D	MOPS 250 mM NaCl pH 7.0	Ammonium acetate 250 mM NaCl pH 7.3	Tris-HCl 250 mM NaCl pH 7.5	Na Phosphate 250 mM NaCl pH 7.6	Imidazole 250 mM NaCl pH 8.0	HEPES 250 mM NaCl pH 8.0	Tris-HCl 250 mM NaCl pH 8.0	Tricine 250 mM NaCl pH 8.0	Bicine 250 mM NaCl pH 8.0	Bicine 250 mM NaCl pH 8.5	Tris-HCl 250 mM NaCl pH 8.5	Bicine 250 mM NaCl pH 9.0
E	Buffer screen "A" pH 4.0	Buffer screen "A" pH 4.78	Buffer screen "A" pH 5.21	Buffer screen "A" pH 5.62	Buffer screen "A" pH 5.95	Buffer screen "A" pH 6.23	Buffer screen "A" pH 6.53	Buffer screen "A" pH 6.81	Buffer screen "A" pH 7.16	Buffer screen "A" pH 7.80	Buffer screen "A" pH 9.0	Buffer screen "A" pH 10
F	Buffer screen "B" pH 4.0	Buffer screen "B" pH 4.22	Buffer screen "B" pH 4.62	Buffer screen "B" pH 5.06	Buffer screen "B" pH 5.62	Buffer screen "B" pH 6.49	Buffer screen "B" pH 7.25	Buffer screen "B" pH 7.76	Buffer screen "B" pH 8.20	Buffer screen "B" pH 8.75	Buffer screen "B" pH 9.0	Buffer screen "B" pH 10
G	10 mM HEPES pH 7.5	50 mM HEPES pH 7.5	100 mM HEPES pH 7.5	250 mM HEPES pH 7.5	10 mM NaPO4 pH 7.5	50 mM NaPO4 pH 7.5	100 mM NaPO4 pH 7.5	200 mM NaPO4 pH 7.5	10 mM Tris-HCl pH 8.0	50 mM Tris-HCl pH 8.0	100 mM Tris-HCl pH 8.0	250 mM Tris-HCl pH 8.0
H	50 mM HEPES 50 mM NaCl pH 7.5	50 mM HEPES 125 mM NaCl pH 7.5	50 mM HEPES 250 mM NaCl pH 7.5	50 mM HEPES 500 mM NaCl pH 7.5	50 mM HEPES 750 mM NaCl pH 7.5	50 mM HEPES 1000 mM NaCl pH 7.5	50 mM Tris-HCl 50 mM NaCl pH 8.0	50 mM Tris-HCl 125 mM NaCl pH 8.0	50 mM Tris-HCl 250 mM NaCl pH 8.0	50 mM Tris-HCl 500 mM NaCl pH 8.0	50 mM Tris-HCl 750 mM NaCl pH 8.0	50 mM Tris-HCl 1000 mM NaCl pH 8.0

Buffers from lane A to F are at concentration 100 mM
Buffer Screens: A) Succinic Acid/NaPO4/Glycine (2:7:7), B) Citric acid/CHES/HEPES (2:4:3)

Table A3. Additive screen for thermofluor experiments (Boivin *et al.* 2013).

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.1 M Urea	0.5 M Urea	1 M Urea	2 M Urea	4 M Urea	6 M Urea	150 mM Guanidine-HCl	500 mM Guanidine-HCl	3% (v/v) DMSO	100 mM $NaHCO_3$	100 mM $KHCO_2$	100 mM NH_4HCO_2
B	100 mM $NaC_2H_3O_2$	100 mM $Ca(C_2H_3O_2)_2$	100 mM CH_3CO_2K	100 mM CH_3CO_2NH	100 mM Na_2SO_4	100 mM Mg_2SO_4	100 mM K_2SO_4	100 mM $(NH_4)_2SO_4$	100 mM Na_2HPO_4	100 mM NaH_2PO_4	100 mM K_2HPO_4	100 mM KH_2PO_4
C	100 mM $Na_2C_4H_4O$	100 mM $Na_3C_6H_5O_7$	100 mM $Na_2C_3H_2O_4$	100 mM $NaNO_3$	10 mM DTT	1 mM TCEP	5 mM TCEP	20 mM TCEP	100 mM Trimethylamine-HCl	1 mM Spermidine	1 mM Spermidine-HCl	5 mM EDTA
D	10 mM Betaine	50 mM Imidazole	100 mM Imidazole	250 mM Imidazole	500 mM Imidazole	10 mM $MgCl_2$	10 mM $CaCl_2$	1 mM $MnCl_2$	1 mM $NiCl_2$	1 mM $FeCl_3$	1 mM $ZnCl_2$	1 mM $CoCl_2$
E	100 mM NaF	100 mM KF	100 mM NH_4F	100 mM LiCl	100 mM NaCl	100 mM KCl	100 mM NH_4Cl	100 mM NaI	100 mM KI	100 mM NaBr	2 mM CHAPS	1 mM Octyl-β-D-Glucoside
F	3% (v/v) Ethylene glycol	1% (v/v) Glycerol	5% (v/v) Glycerol	10% (v/v) Glycerol	20% (v/v) Glycerol	5% (v/v) PEG 400	5% (v/v) PEG 1000	5% (v/v) PEG 4000	10 mM L-Proline	50 mM Taurine	50 mM L-Glycine	50 mM L-Histidine
G	50 mM L-Arginine	50 mM (each) L-Arg + L-Glu	500 mM (each) L-Arg + L-Glu	50 mM L-Glutamate	50 mM L-Glutamine	50 mM L-Lysine	100 mM D-Glucose	100 mM Sucrose	100 mM Maltose	100 mM D-Sorbitol	2 mM NADH (5 mM $MgCl_2$)	2 mM ATP (5 mM $MgCl_2$)
H	2 mM ADP (5 mM $MgCl_2$)	2 mM cAMP (5 mM $MgCl_2$)	2 mM GTP (5 mM $MgCl_2$)	2 mM GDP (5 mM $MgCl_2$)	2 mM cGMP (5 mM $MgCl_2$)	Buffer	Custom	Custom	Custom	Custom	Custom	Custom

Appendix 4 – MS protocol of Proteomics Core Facility at EMBL Heidelberg

LC-MS for the determination of molecular weight

The samples were treated with 1% trifluoroacetic acid (TFA) prior to injection into the Acquity UPLC System equipped with ACQUITY UPLC Protein BEH C4 Column (300 Å, 1.7 µm, 2.1 mm X 150 mm, Waters Corporation, Milford, USA) for protein separation. The amount of injected sample was approximately 10 µg. The samples eluting out from the column were directly subjected to mass spectrometry analysis using a Q-TOF II mass spectrometer (Waters Corporation, Milford, USA). The sample ionization was conducted using a standard ESI source in positive ion mode. The solvents used in the protein separation contained 0.1 % formic acid (aqueous solution, referred as solvent A) and 0.1% formic acid in acetonitrile (solvent B). The samples were loaded at a constant flow of 0.2 mL/min in a solvent mixture containing 96% solvent A and 4% B. The column was held at 4% B for 5 min before ramping to 25% B by 6 min. This was followed by a linear gradient to 80% B until 17 min.

For the Q-Tof, a spray voltage of 3.5 kV was applied with a cone voltage of 35 V and extraction cone at 5 V. The desolvation of the samples was conducted at 350°C while the source temperature was set to 120°C. Nitrogen was used as the desolvation gas at a flow rate of 500 l/min. Collision energy was set at 5 eV with argon in the collision cell at a pressure of 4.5 e-5 mbar. Data was acquired in continuum mode over the mass range 500-3500 m/z with a scan time of 0.5 s and an interscan delay of 0.1 s. Prior to sample data acquisition, an external calibration against a reference standard of intact myoglobin was performed. Spectra from the chromatogram protein peak were summed and the intact

mass was calculated using the MaxEnt1 maximum entropy algorithm (Waters/Micromass, Waters Corporation, Milford, USA). The molecular weight was obtained by deconvoluting to zero charge.

Sample preparation for in-solution digestion

For trypsin digestion the used reagents were dissolved in 50 mM ammonium bicarbonate while for chymotrypsin digestion buffer containing 100mM TRIS-HCl at pH 8.0, 10mM $CaCl_2$ were used. The digestion was performed using approximately 100 µg of protein, which was diluted to 1 µg/µL in the suitable buffer (50 mM ammonium bicarbonate for trypsin, 100mM Tris-HCl pH 8.0 and 10mM $CaCl_2$ for chymotrypsin). Prior to the digestion, the protein was reduced by treatment with DTT (2.5 µL of 50 mM DTT/100 µl solution) at 56 °C for 30 minutes and alkylated by treatment with iodoacetamide (2.5 µL of 110 mM stock) at room temperature protected from light for 20 minutes. The trypsin digestion (1 µL of a 1 µg/µL solution) was carried out overnight at 37 °C and the chymotrypsin digestion at 25°C.

LC-MS/MS of the digested protein

The separation of the peptides was conducted using the nanoAcquity UPLC system (Waters Corporation, Milford, USA) equipped with a nanoAcquity Symmetry C18 trapping column (5µm, 180 µm x 20 mm, Waters Corporation, Milford, USA) and a nanoAcquity BEH C18 analytical column (1.7µm, 75µm x 200mm, Waters Corporation, Milford, USA). The outlet of the analytical column was connected directly to an LTQ Orbitrap Velos Pro hybrid mass spectrometer (Thermo Fisher Scientific) using the

Proxeon nanospray source (ESI-source). Solvent A was 0.1 % formic acid in water and solvent B 0.1 % formic acid in acetonitrile. The samples were loaded with a constant flow of solvent A at 5 µL/min onto the trapping column using a trapping time of 6 minutes. The elution of peptides via the analytical column was conducted at a constant flow of 0.3 µL/min. During the elution step, the percentage of solvent B was increased in a linear fashion from 3 % to 7 % in 10 minutes, then to 25 % in 20 minutes and finally to 40% in 10 minutes. The peptides were injected into the mass spectrometer via a Pico-Tip Emitter (360 µm OD x 20 µm ID; 10 µm tip; New Objective, Inc., Woburn, MA, USA) by applying a spray voltage of 2.2 kV. The capillary temperature was set to 300 °C. Full scan MS spectra with mass range 300-1700 m/z were acquired in profile mode in the FT with resolution of 30000. The maximal filling time was set at 500 ms with limitation of 10^6 ions. Up to 15 of the most intense ions from the full scan MS were selected for sequencing in the LTQ. Normalized collision energy of 40 % was used. The fragmentation was carried out after accumulation of 3 x 10^4 ions or after filling time of 100 ms for each precursor ion depending on which occurred first. MS/MS data was acquired in centroid mode. Only multiply charged (2+, 3+, 4+) precursor ions were selected for MS/MS. The dynamic exclusion list was restricted to 500 entries with maximum retention period of 30s and relative mass window of 10 ppm. A lock mass correction was applied using a background ion (m/z 445.12003) in order to improve the mass accuracy.

Data analysis

The filtering of the data was performed using MaxQuant software (version 1.1.1.36, Cox & Mann, 2008). The same software was used to create the .mgf files, which

were needed for searching in MASCOT version 2.2.03 (Matrix Science). The data were searched against a species-specific (*Arabidopsis thaliana*) Uniprot database with a list of common contaminants included. The data were searched with the following modifications: Carbamidomethyl (C) (Fixed) and Oxidation (M) (Variable), Phospho (ST). The mass error tolerance for the full scan MS spectra was set at 20 ppm and for the MS/MS spectra at 0.5 Da. A maximum of 1 missed cleavage was allowed. Only peptides with a mascot score greater than 20 were reported.

Appendix 5 – Hydroxylamine assay results for inactive ApnA A1 PCC7821 variants

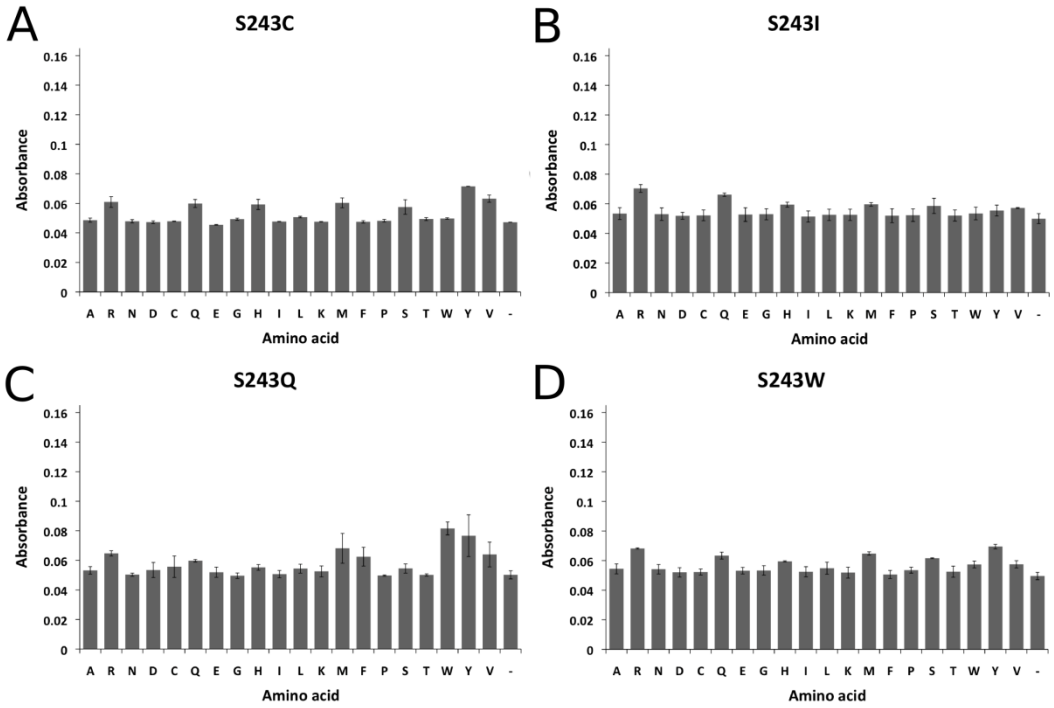

Figure A1. The rest of the hydroxylamine-trapping assay results for the inactive ApnA A1 PCC7821 variants with the substitution at position 243. A) – D) The substitution of Ser243 with larger polar or hydrophobic residue (cysteine, isoleusine, glutamine or tryptophan) resulted in inactive enzyme.

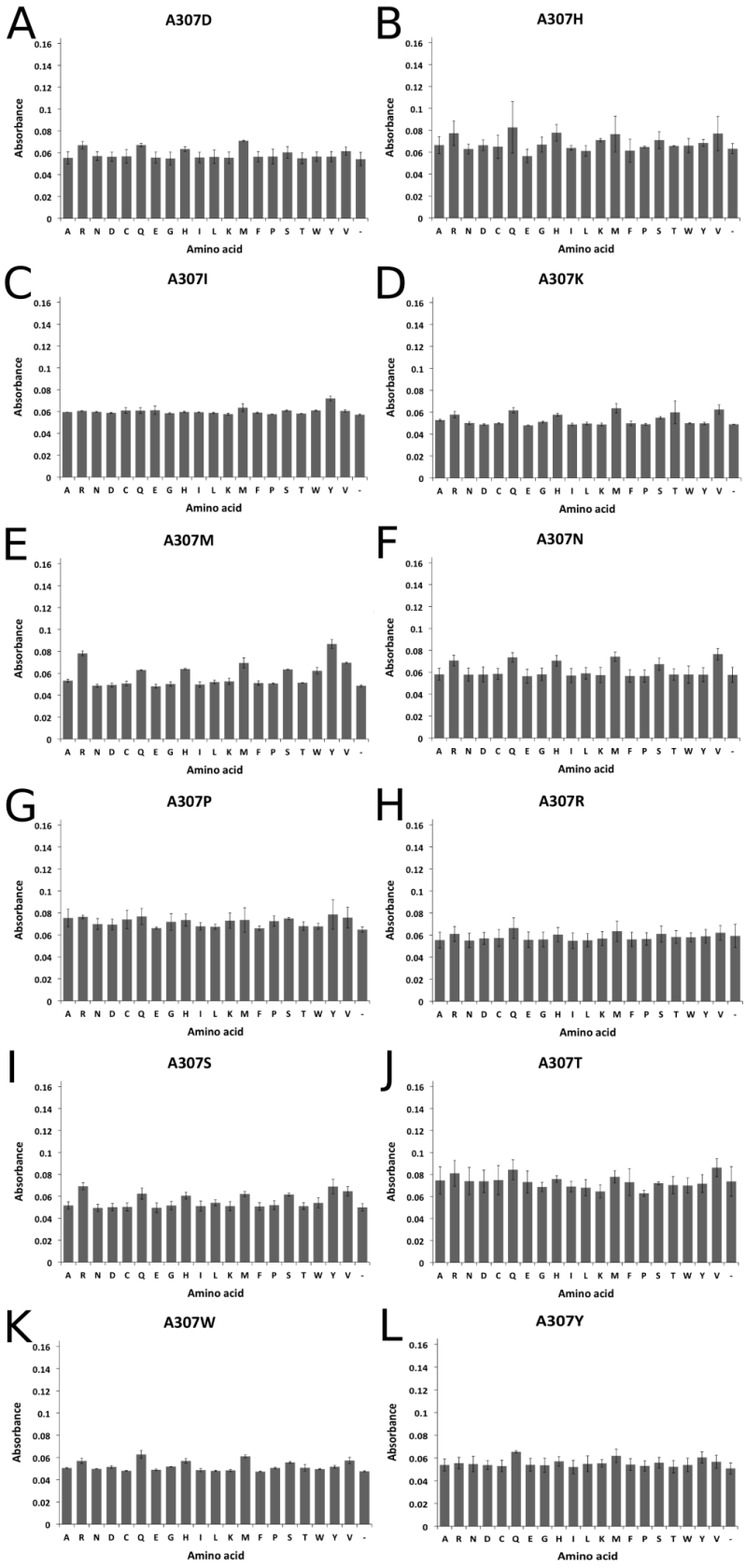

Figure A2. A) – L) The rest of the hydroxylamine-trapping assay results for the inactive ApnA A1 PCC7821 variants with the substitution at position 307.