Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Structural analysis of protein complexes associated with DNA maintenance

Sandra Katharina Melanie Schuller
(geb. Schneller)

aus Ulm, Deutschland
2018

This thesis has been prepared from March 2013 to November 2017 in the laboratories of Prof. Dr. Karl-Peter Hopfner (Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München)

Zusammenfassung

Schäden am Genom stellen eine ernsthafte Bedrohung für jede Zelle dar. Daher ist es von entscheidender Bedeutung, dass die Zelle durch die Auslösung verschiedenster Signalwege schnell auf diese Schäden reagiert. Deren Aktivierung führt zur Rekrutierung unterschiedlicher Reparaturmaschinen und anderer Komplexe, die den Zugang zur dicht gepackten DNA ermöglichen. Der Verlust der genomischen Integrität führt zudem dazu, dass der Zellzyklus angehalten wird.

In Prokaryoten ist DisA (DNA Integrity Scanning Protein A) ein Hauptakteur, der nach genomischen Brüchen sucht. DisA ist ein homo-oktameres Enzym, das zyklisches di-AMP herstellt. Nach der Bindung an atypische DNA Strukturen wird die Synthese des zyklischen di-AMPs, welches als sekundärer Botenstoff wirkt, gehemmt, was die Sporulation unterbindet. Diese Funktion erfordert eine strenge Regulierung der zellulären Konzentration durch Synthese, Abbau und Export. Zyklisches di-AMP wird hauptsächlich durch die Phosphodiesterase GdpP abgebaut. Der erste Teil dieser Arbeit behandelt die Synthese von zyklischem di-AMP durch DisA und dessen mögliche Regulierungsmechanismen. Der Mechanismus der zyklischen di-AMP-Synthese ist auf atomarer Ebene aufgeklärt, jedoch ist bisher ungeklärt, was der geschwindigkeitsbegrenzende Schritt der Synthesereaktion ist. Eine 2.25 Å Kristallstruktur einer katalytisch hyperaktiven DisA-Mutante aus *Thermotoga maritima* wurde gelöst. In dieser Mutante sind in der zyklischen di-AMP-Synthesedomäne drei Argininreste durch Alanin ersetzt worden. In der Struktur ist erkennbar, dass der Ausgangstunnel einen größeren Durchmesser aufweist, und somit die Wirkung der Mutante erklärt: Die drei Arginreste verlangsamen die Diffusion des Produktes und wirken dadurch geschwindigkeitsregulierend. Des Weiteren wurden die Strukturen von DisA aus drei verschiedenen Spezies gelöst um mit Hilfe der evolutionären Vielfalt Rückschlüsse auf mögliche Regulationsmechanismen zu ziehen. Verglichen mit der 3 Å Struktur von *Thermotoga maritima* zeigen die Kryoelektronenmikroskopiestrukturen des oktameren DisA aus *Bacillus subtilis* bei 8,2 Å und *Mycobacterium tuberculosis* bei 8,6 Å eine erhöhte Flexibilität in der Assoziation eines Monomers. Diese Ergebnisse deuten auf einen Inaktivierungsmechanismus von DisA durch Disassemblierung der Monomere hin. Parallel dazu wurde eine Kryoelektronenmikroskopiestruktur des zytosolischen Teils von GdpP bei 4,7 Å rekonstruiert und ein vollständiges Strukturmodell basierend auf Modellen homologer Domänen erstellt. Die Struktur zeigt eine tetramerische Organisation des Komplexes durch Dimerisierung von je zwei DHH-Domänen sowie zwei GGDEF-Domänen. Die ebenfalls dimerisierten PAS-Domänen sind flexibel mit der nach außen gerichteten, ellipsoiden Ringstruktur verbunden. Ein Vergleich mit veröffentlichten Nukleotid-gebundenen GGDEF-Dimerstrukturen ermöglicht die Postulierung eines Regulationsmechanismus, der auf einer weitreichenden Kommunikation zwischen den DHH-Domänen und den GGDEF-Domänen basiert und abhängig vom Nukleotidzustand der letzteren ist.

In Eukaryoten steht die DNA-Reparatur vor der zusätzlichen Herausforderung der chromosomalen DNA-Architektur. Um dieses Hindernis zu überwinden, spielen Chromatin-Remodeller eine wesentliche Rolle, indem sie dicht gepackte DNA leichter zugänglich machen. Der Remodeller Fun30 ist bei der DNA-Doppelstrangbruchreparatur an der Langstreckenresektion beteiligt. Im Gegensatz dazu ist der modular aufgebaute Multiproteinkomplex INO80 für die Kurzstreckenresektion von Bedeutung. Der zweite Teil dieser Arbeit stellt die Reinigung, biochemische Charakterisierung, sowie die Kristallisation von Fun30 aus dem thermophilen Pilz *Chaetomium thermophilum* vor. Zusätzlich behandelt diese Arbeit die Reinigung und Analyse des DNA-bindenden Nhp10 Modules des INO80

Komplexes. Es konnte gezeigt werden, dass das Nhp10 Protein eine andere molekulare Organisation in seiner Domänenstruktur aufweist, als bisher angenommen. Darüber hinaus wird gezeigt, dass die postulierte Sequenzerkennung durch den INO80-Komplex nicht durch das Nhp10-Modul bestimmt wird.

Summary

Lesions in the genome present a severe threat to every cell. It is essential for the cell to react quickly to such damage and does so by triggering a variety of signalling pathways. Activation of these pathways leads to the recruitment of the repair machinery and complexes that allow access to the tightly packed DNA. Additionally, the loss of genomic integrity requires cell cycle arrest.

In prokaryotes, a key player that scans for genomic lesions is DisA (DNA integrity scanning protein A), a homo-octermeric enzyme that synthesizes cyclic di-AMP. Upon binding to non-standard DNA, synthesis of the second messenger cyclic di-AMP is inhibited leading to sporulation arrest. This function as a second messenger requires tight regulation of cyclic di-AMP levels through synthesis, degradation and export. Degradation of cyclic di-AMP is carried out amongst other by the phosphodiesterase GdpP. The first part of this thesis focuses on the synthesis of cyclic di-AMP by DisA and possible mechanisms of its regulation. The mechanism of cyclic di-AMP synthesis is understood at an atomic level, however the rate-limiting factor of the synthesis reaction remains a matter of speculation. Structural analysis was performed on a catalytically overactive DisA mutant from *Thermotoga maritima* that harbours three arginine residues substituted by alanine in the cyclic di-AMP synthesising domain. The resulting 2.25 Å structure explains the increased catalytic activity of the mutant by revealing an increase in the diameter of the exit tunnel where the arginine residues function as a molecular curtain. Additionally, the comparison of the structures of DisA from three different species helped to shed light on potential regulatory mechanisms making use of evolutionary diversification. Structures obtained with cryo Electron Microscopy (cryo-EM) of the octameric DisA from *Bacillus subtilis* at 8.2 Å and *Mycobacterium tuberculosis* at 8.6 Å showed an increased flexibility in the association of one monomer in comparison to the structure of *Thermotoga maritima* at 3 Å. These results favour an inactivation mechanism of DisA through disassembly of the monomers. In parallel, a cryo-EM structure of the cytosolic part of GdpP was obtained at 4.7 Å in the core protein. Consistent with the observed resolution a complete structural model could be obtained based on models of homologous domains. The structure revealed a tetrameric organization of the complex by dimerization of two DHH domains as well as two GGDEF domains. The dimerized PAS domains are attached with increased flexibility to the outward-facing ellipsoidal ring structure. A comparison of published nucleotide-bound GGDEF dimer structures allows the postulation of a regulation mechanism based on long-range communication between the DHH domains and the GGDEF domains, dependent on the nucleotide state of the later.

In eukaryotes, DNA repair faces the additional challenge of chromosomal architecture. To overcome this obstacle, chromatin remodellers play an essential role by making tightly packed DNA more accessible. The single-chain remodeller Fun30 is known to play a role in the process of long-range resection in DNA double strand break repair while the multi-module containing INO80 complex is involved in short-range resection. The second part of this thesis presents the purification, biochemical characterization and crystal screening of Fun30 from the thermophilic fungus *Chaetomium thermophilum*. Furthermore, the purification and analysis of the submodule of INO80 responsible for DNA binding in *Saccharomyces cerevisiae* presented in this thesis revealed a different domain organization of its major component, the Nhp10 protein than previously considered. In addition, the results reveal that the postulated sequence recognition by the INO80 complex is not determined by the Nhp10 module.

Table Of Contents

Function of second messengers

Nucleotide second messengers in bacteria

Aim of this project

Aim of this project

Purification of DisA

Crystallization of tmaDisA3xR mutant

Structure determination and Refinement

EM-grid preparation

Data acquisition

Data processing

Model building

Cryo-EM-structure of tmaDisA

Purification and cryo-EM structure of bsuDisA

Purification and cryo-EM structure of mtubDisA

Crystal structure of tmaDisA3xR mutant

Comparison of tmaDisA3xRmutant with wild type DisA

Cryo-EM structure of tmaDisA

Purification and cryo-EM structure of bsuDisA

Purification and cryo-EM structure of mtubDisA

Analysis of stability-variations in DisA-complexes among different species

Cryo-EM studies of GdpP

Biological impact of the structure

Aim of this project

Aim of this project

Cloning

Cloning of His-tagged Nhp10, Ies1, Ies3, Ies1, (Ino80p_1-450

Purification of His-tagged Nhp10 in complex with Ies1, Ies3, Ies5 and Ino80p_1-450 from H5 cells

Purification of His-tagged Nhp10, Ies3, Ies5, (Ies1) from H5 cells

Purification of nanobodies from Rosetta cells

Purification of Nhp10, Ies3, Ies5, (Ies1) with nanobody #42 from H5 cells

Purification of HisIes5, Nhp10 from Rosetta cells from two different plasmids

Purification of HisNhp10, Ies3, Ies5 in Rosetta cells from three different plasmids

Purification of HisNhp10, Ies3, Ies5 in Rosetta cells from two plasmids with nanobody #42

DNA-binding detection by Electrophoretic Mobility Shift Assays (EMSA)

Purification of Fun30 constructs

Limited Proteolysis

Stability assays of ctFun30_333-978

DNA-binding detection by Electrophoretic Mobility Shift Assays (EMSA)

Nucleosome-binding Assay

Right-Angle Light-Scattering (RALS) size exclusion chromatography

NADH-coupled ATPase assay

Small-Angle X-ray Scattering (SAXS)

Test purification of SMRCD1a&c

Purification of His-tagged Nhp10 in complex with Ies1, Ies3, Ies5 and Ino80p_1-450 from H5 cells

Purification of His-tagged Nhp10, Ies3, Ies5, (Ies1) from H5 cells

Purification of Nhp10, Ies3, Ies5, (Ies1) with nanobody #42 from H5 cells

Purification of HisIes5 Nhp10 from Rosetta cells from two different plasmids

Purification of HisNhp10, Ies3, Ies5 in Rosetta cells from three different plasmids

Purification of HisNhp10, Ies3, Ies5 in Rosetta cells from 2 plasmids with nanobody #42

DNA Binding Assays of HisNhp10, Ies3, Ies5, nanobody #42

Bioinformatic Analysis of Nhp10

Purification of full-length Fun30 from *Chaetomium thermophilum* (fl ctFun30)

Mass-determination of fl ctFun30 by size exclusion coupled Right-Angle Light Scattering (RALS)

Thermal shift assays

Activity assays of fl ctFun30

Limited Proteolysis of fl ctFun30

Purification of ctFun30_333-978

1.

Cryo-EM single particle analysis of proteins in the cyclic di-AMP pathway

1.1. Introduction

1.1.1. Bacterial nucleotide second messengers

<u>Function of second messengers</u>

Every cell has to communicate with its environment and must react to external stimuli, such as nutritional availability, hyperosmotic or temperature stress. In multicellular organisms, adjacent cells can send signals to their neighbours, for example in neuronal signal transduction or metabolic changes via hormones. However, already unicellular organisms communicate with each other via chemotaxis or during mating [1]. Therefore, external stimuli need to be transmitted through the cellular barrier. As this barrier is built to protect the cell, it is often impermeable to these external signals. For this reason, external stimuli are recognised by specialized receptors that survey the environment and that activate signalling pathways by generating a distinct signal in the cell. In particular, this internal signal is often the activation of a membrane receptor. This internal signal then leads to the activation, release or synthesis of a small molecule that acts as a so-called second messenger, triggering a downstream cellular response through an effector protein. Altogether, an external stimulus is propagated over an internal effect to the second messenger, leading to a response to the initial signal (Fig.1). Examples for second messengers are Calcium, certain lipids such as steroids, inositol-1,4,5-triphosphate, nitric oxide but also derivatives of nucleic acids. [2]
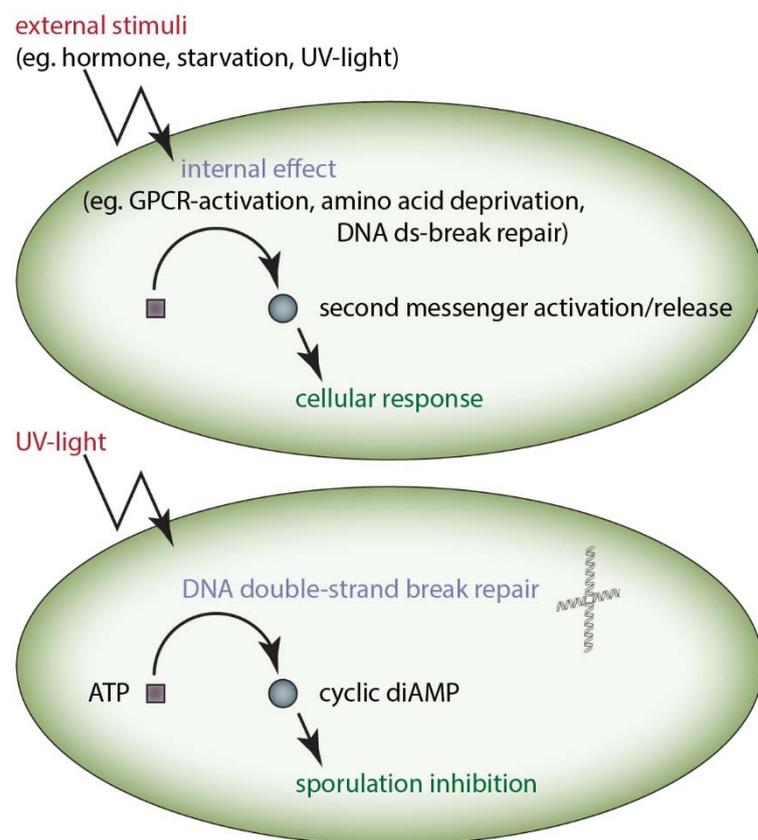
Fig.1: Schematic signalling cascade of second messengers (top) with signalling of UV-light induced DNA double-strand break via cyclic di-AMP as example (bottom)

Nucleotide second messengers in bacteria

Bacteria employ nucleotides as second messengers, the so-called bacterial nucleotide second messengers. The bacterial nucleotide second messengers, known to date, are cyclic adenosine or guanosine monophosphate (cAMP and cGMP), cyclic diadenosine or diguanosine monophosphate (c-di-GMP and c-di-AMP), guanosine tetra- and pentaphosphate ((p)ppGpp) and cyclic guanosine monophosphate – adenosine monophosphate (cGAMP). Some of these nucleotide second messengers also play a role in higher organisms. However, as this work focuses on their function in bacteria, their role in eukaryotes will not be discussed in more detail.

cAMP (Fig.2) is synthesised by the family of adenylyl cyclases and plays an important role in switching bacterial metabolism from glucose to lactose processing under glucose starvation. Adenylyl cyclases are usually inhibited by intracellular glucose. Lack of glucose leads to changes in the posttranslational modification of the glucose phosphotransferase system, activating the adenylyl cyclase and hence increases cAMP level. The cAMP receptor protein (CRP), bound to cAMP in turn positively
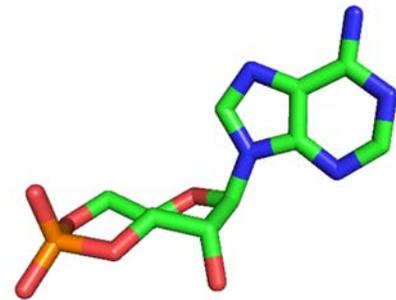
Fig.2: Chemical structure of cAMP; C atoms are depicted in green, blue represents N atoms, red O atoms and orange P atoms

regulates the lac-operon and thereby increases the expression of enzymes metabolising and transporting lactose. CRP is considered to play a key role in managing catabolism by influencing the transport of substrates, glycolysis, the citrate cycle and aerobic respiration. In addition, secreted cAMP is known to alter motility and virulence. cAMP furthermore plays an important role during host infection to enhance pathogenicity. (reviewed in [3])

cGMP (Fig.3) is synthetised by the guanyl cyclase Cya2, as first shown in the cyanobacterium *synechocystis* strain PCC 6803. In *Rhodospirillum centenum* the gene for Cya2 is encoded four genes upstream of a CRP homolog, which was shown to specifically bind cGMP. This indicates a mechanism of signalling for cGMP similar to the signalling pathway of cAMP. cGMP signalling was shown to enhance encystment in some
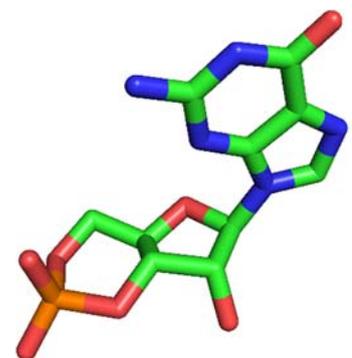
Fig.3: Chemical structure of cGMP

Proteobacteria, a process allowing the cells to survive in unfavourable conditions, like desiccation, by forming dormant cells. [4]

Cyclic di-GMP (Fig.4) is produced by diguanylate cyclases that always contain a dimer of two GGDEF domains. These domains are regulated by allosteric product inhibition. C-di-GMP is hydrolysed into the linear pGpG by either EAL domains or HD-GYP domains. These domains, as well as the GGDEF domain, are termed from the amino acid sequences forming the active centres. A fusion of EAL or HD-GYP domains with GGDEF domains couples synthesis and degradation within one enzyme or the regulation of one function via an inactive second domain. Proteins containing sensor domains fused to the cyclase domains allow production of c-di-GMP as a direct response to different external or internal stimuli such as oxygen and redox conditions but also cell cycle and light. Effector proteins activated upon binding of c-di-GMP are amongst other PilZ domain proteins, such as FleQ, PelD, CdgG and PopA. Some of these proteins harbour degenerated GGDEF domains, leading to an incapability of synthesis of c-di-GMP while retaining binding of the nucleotide second messenger. C-di-GMP effector proteins play roles in cell cycle regulation, regulation of pathogenicity, biofilm formation and motility. By binding to riboswitches, c-di-GMP also influences the expression of downstream genes and thereby shows another means of influencing the cell at many different regulatory levels. C-di-GMP is thus described as a `lifestyle-switch regulator`, whose influence is, however, differing among bacteria dependent on their physiological and ecological environment. (reviewed in [5])

Fig.4: Chemical structure of c-di-GMP

ppGpp (Fig.5) was first discovered in *Escherichia coli* as a signal starvation signal and different stress conditions and is therefore known as a stringent response factor [6, 7]. Both, ppGpp and pppGpp are synthesised by RelA/SpoT homology (RSH)-type synthetases by transferring
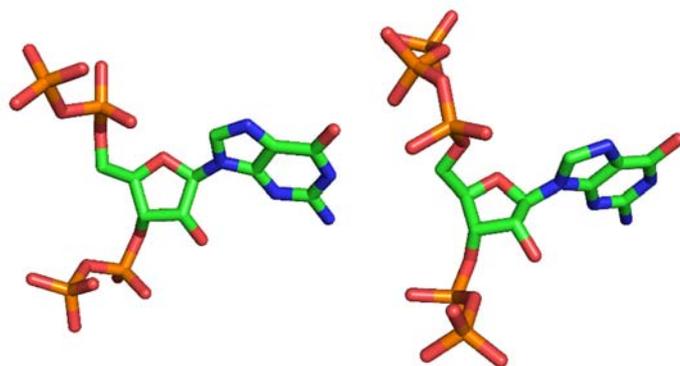

Fig.5: Chemical structure of ppGpp (left) and ppGpp (right)

pyrophosphate of ATP to GDP and GTP. RelA is stimulated upon binding to a stalled ribosome bound by an uncharged tRNA at its A-site as signal of starvation. Some organisms are furthermore able to generate ppGpp by hydrolysing the γ-phosphate of pppGpp [8]. Degradation occurs amongst other by the RSH-type hydrolases by removing the δ, ε-pyrophosphate moiety [9].

(p)ppGpp affects DNA replication by binding to the Primase DnaG. The formation of RNA primers, which function as starting points for the DNA polymerase is inhibited by the binding event [10-12]. Transcription is affected by (p)ppGpp by its interaction with the RNA polymerase, leading to an decrease in rRNA and ribosomal protein synthesis and an increase in transcription from promoters for amino acid synthesis [7, 13]. Furthermore, (p)ppGpp affects GTPases involved in the initiation, elongation and termination of translation, as well as the assembly of the small and large ribosomal subunits [9]. (p)ppGpp also influences the bacterial resistance to acidic conditions by inhibiting glutamate, lysin and arginine decarboxylating enzymes [9]. In conclusion, (p)ppGpp influences replication, transcription, translation and the metabolism in bacteria depending on amino acid availability and starvation [9].

In some other bacteria, such as *Mycobacterium tuberculosis* and *Vibrio cholerea*, ppGpp plays an important role in modulating virulence. In some *Streptomycetes* c-di-GMP is essential for the synthesis of antibiotics. In *Rhizobiaceaes* it plays an essential role in nitrogen fixation. [14-16]

Cyclic di-AMP (Fig.6) was first discovered during structural studies of the <u>D</u>NA <u>i</u>ntegrity <u>s</u>canning protein (DisA) from *Thermotoga maritima*. It is the only nucleotide second messenger essential to most organisms under normal growth conditions [17-19]. It is produced in bacteria and archaea and is essential in all Firmicutes studied so far, however not in all bacteria [18, 20-27]. By now, it is linked

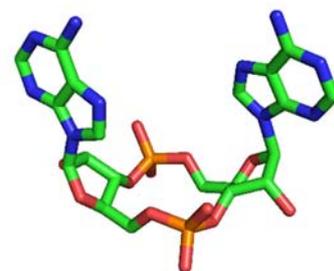Fig.6: Chemical structure of c-di-AMP

to fatty acid synthesis, cell growth, sporulation, biofilm formation, virulence and cell wall homeostasis [28], and most importantly potassium homeostasis [29]. However, it is also essential in some gram-negative or cell wall-less organisms [30-33]. This suggests a role for c-di-AMP as a cofactor for an essential protein or in inhibiting a toxic compound [34].

1.1.2. Synthesis of cyclic di-AMP

C-di-AMP is exclusively synthesised by the family of diadenylatecyclases (DAC) that catalyse the condensation of two ATP into c-di-AMP and release two pyrophosphates as side product. Some distinct enzymes are furthermore able to synthesise c-di-AMP from two molecules of ADP [35].

Different DACs are known, existing either as single c-di-AMP synthesizing enzymes or in parallel in one species. They are responsible for distinct cellular processes. Most pathogenic Fimicutes possess only one enzyme. All DAC enzymes possess a Dimer of DAC-domains, required for the functional activity in c-di-AMP synthesis. Additional domains are responsible for multimerization and regulation. Some DAC enzymes further comprise a transmembrane domain that anchors them to the cell membrane and most likely is responsible for transfering external stimuli to the intracellular domains. The model organism *Bacillus subtilis* expresses three distinct DAC enzymes: DisA, CdaA and CdaS. Whereas CdaA and CdaS are closely related, the sequence similarity to DisA is lower [36]. The recently discovered CdaM represents another class of DAC enzymes expressed in *Mycoplasma pneumonia* and is closely related to CdaA, however no further structural information is known. All DAC domains possess the conserved DGA and RHR motifs, which are crucial for enzymatic activity. [35, 37, 38] An overview of the DAC enzymes is shown in figure 7.

CdaS is only present in specific spore forming bacteria of the Bacillus and Clostridium species and is expressed exclusively during sporulation. It consists of two helices, located N-terminal of the DAC domain. Crystal contacts suggest the formation of a hexameric ring. In this ring, the interface between the monomers would impede an efficient synthesis of c-di-AMP.[36, 39, 40]

CdaA is the most frequent DAC, and in many pathogens only c-di-AMP synthesising enzyme [36]. It is a membrane bound protein due to its N-terminal transmembrane domain. The C-terminal DAC domain is surrounded by two coiled-coiled motifs. The structure of the DAC domain is highly similar to the enzymatic domains of DisA and CdaS and it has been shown, that the enzymatic activity is independent of the transmembrane domain and the coiled-coiled domain [38]. CdaA is known to play a role in cell wall synthesis and potassium homeostasis. Changes in these cellular activities might change the cross-linking of the peptidoglycan of the cell wall [18, 23, 37, 41, 42] and thereby alter cell wall composition and

antibiotic resistance. The precise mechanism remains to date elusive. In addition, CdaA is encoded in the same operon as *glmM* and *cdaR*. *GlmM* encodes the phosphoglucosamine-6-phosphate mutase, which generates a precursor for cell wall metabolism. CdaR interacts with CdaA and stimulates its enzymatic activity [18]. Similar to DisA, CdaA is involved in DNA repair [43].

DisA (<u>D</u>NA <u>i</u>ntegrity <u>s</u>canning protein <u>A</u>) is the third type of DAC in *Bacillus subtilis*. It´s sequence is least conserved to CdaA and CdaS and not present in all bacteria [36]. It is mainly expressed in Actinobacteria and gram-positive spore-forming Firmicutes. Structural studies on DisA from the thermophilic organism *Thermotoga martima* revealed its composition of three parts: An N-terminal DAC domain with conserved Asp-Gly-Ala (DGA) and Arg-His-Arg (RHR) motifs that is required for the enzymatic activity, an C-terminal helix-hairpin-helix (HhH) domain and a helix bundle that separates the other domains from each other. This forms an elongated shape, which allows tetramerization via the long axis. Two tetramers then interact via their DAC domain and form a barrel-shaped octamer. This face-to-face orientation is required for activity, as only DAC-dimers possess catalytic activity [37]. The activity of the enzyme is regulated by the presence of non-standard DNA, like Holliday junctions [37]. This led to the suggestion that DisA is scanning the bacterial genome for DNA lesion [43]. Binding to such non-standard DNA leads to the inactivation of the enzyme and a decrease in c-di-AMP levels, which stalls sporulation [37, 44-46]. The
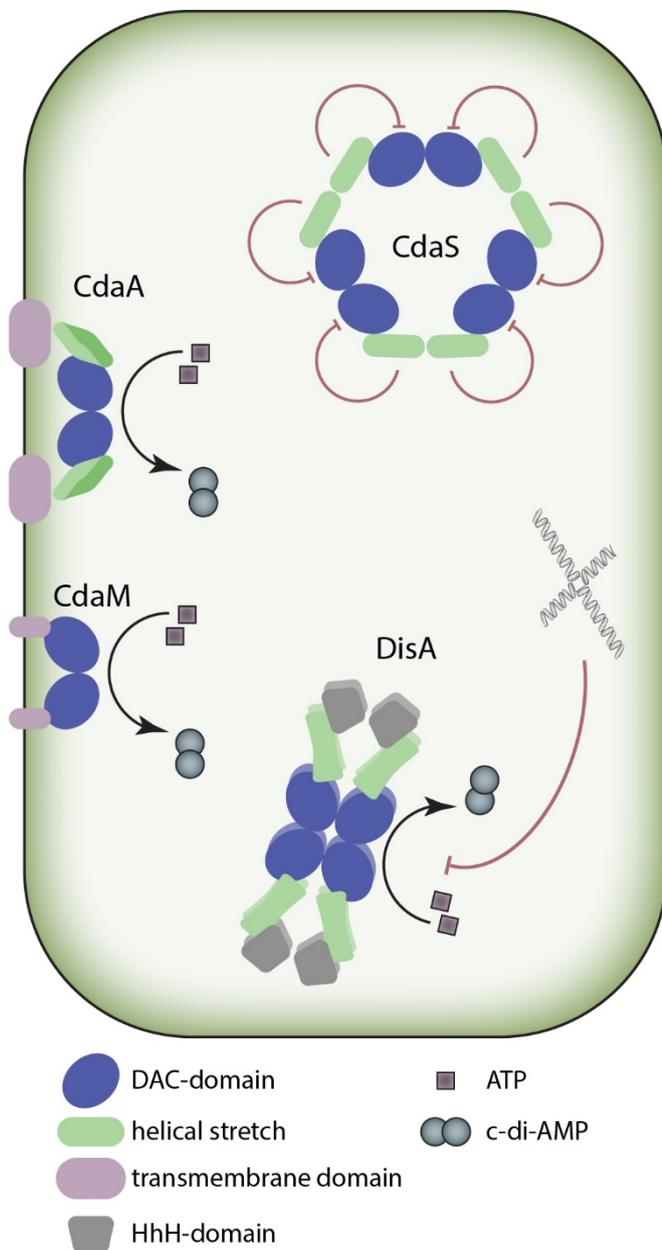

Fig.7: Overview of c-di-AMP synthesis

mechanism of regulation, as well as of the recruitment of repair factors remains yet unknown. One potential regulator is RadA. DisA and RadA share a phylogenetically conserved operon and are reported to interact physically [47, 48]. RadA is known to play a role in DNA repair by processing Holliday junction intermediates [49]. This indicates a potential role of RadA in the regulation mechanism of DisA, as they are both involved in the same pathway. The question of how the recognition of loss of DNA integrity leads to inhibition of sporulation remains unknown so far.

1.1.3. Effector proteins of cyclic di-AMP

The research for c-di-AMP interacting proteins revealed several binding proteins: DarR, KtrA, CpaA, KdpD, PycA, DarR, CbpA, CbpB and several riboswitches. A schematic representation of effector proteins of c-di-AMP is shown in Fig.8.

DarR from *Mycobacterium smegmatis* is a transcription repressor of the TetR family, responsible for repression of its own gene as well as of an entire operon, encoding genes of the fatty acid metabolism and of the *cspA* gene encoding a cold shock protein. Repression of these genes is stimulated by c-di-AMP binding to DarR. DarR mutants have an effect on cell shape and growth and are toxic at high concentrations. However, DarR is only present in selected bacteria. [50]

KtrA from *Staphylococcus aureus* is a subunit of a potassium transporter that imports extracellular potassium into the bacterial cytosol. It forms an octameric ring, which is binding to two transmembrane KtrB proteins and thus represents the cytoplasmic potassium gating component of the complex [51]. KtrA itself consist of two RCK domains of which the N-terminal domain (RCK_N) binds $NAD^+/NADH^+$ and ATP and interacts with KtrB and the C-terminal domain (RCK_C) binds c-di-AMP. Recognition of c-di-AMP by KtrA is assumed to inhibit potassium uptake leading to low potassium concentration upon c-di-AMP accumulation. Another cation transporter, the K^+-H^+-antiporter CpaA, also possesses a RCK_C domain, which shows affinity for c-di-AMP, suggesting a role for the RCK_C domain in regulation of the potassium transport in response to c-di-AMP [36, 52]. A further potassium regulatory system involved in the c-di-AMP response in *Staphylococcus aureus* is the two-component regulatory system containing the sensor kinase KdpD [36], indicating a global role of c-di-AMP in potassium homeostasis.

KimA from *Bacillus subtilis* (formerly known as YdaO protein) represents a second potassium transporter regulated by c-di-AMP. It is a high affinity transporter composed of 11 transmembrane helices with an N-terminal extracellular and a C-terminal intracellular domain. Potassium uptake by KimA, as well as the KtrAB system is inhibited by c-di-AMP reflecting increased potassium concentrations.

PycA from *Listeria monocytogenes* is a pyruvate carboxylase and the only enzyme that generates oxaloacetate in bacteria with an incomplete citric acid cycle. Oxaloacetate is required for the synthesis of the amino acids aspartate and glutamate and is therefore essential for these organisms under growth on media containing sugar as single carbon source [53]. PycA forms a tetramer that binds c-di-AMP in a 4:8 ratio. Allosteric binding at a site located at the interface of two subunits of the carboxyltransferase domain triggers a large conformational change. This inactivates the catalytic activity of the enzyme [54].

DarA from *Bacillus subtilis* (**PstA** in *Staphylococcus aureus* and *Listeria monocytogenes*) [36, 54, 55] is a protein of a widely distributed family in Firmicutes[56] that is very closely related to the group of the P_{II} signal transducers. P_{II} proteins connect the signalling of second messengers such as 2-oxoglutarate and ATP, dependent on their post-translational modifications with their effector proteins. The core architecture of both protein classes is nearly identical [55-59]. Both DarR and P_{II}-like proteins form homotrimers and bind second messengers at the interactions site of two adjacent subunits Binding occurs in a pocket harbouring a consensus sequence of TKLxxxGGFLxxGNTT and a conserved GGA stretch in the C-terminal region. Upon c-di-AMP binding a conformational change takes place, altering the binding properties of downstream effectors [56-58].

CbpA and **CbpA** were identified in screens searching for c-di-AMP binding proteins in *Listeria monocytogenes* [54]. This suggests c-di-AMP to be a regulator of the so far unknown enzymatic activity of CbpA and CbpB. Both proteins contain a CBS-domain (cystathionine beta synthase domain), which binds ligands containing adenosyl groups, such as AMP, ATP, S-Adenosyl methionine or potentially c-di-AMP, and hence regulate adjacent enzymatic domains [60, 61].

Riboswitches are an additional class of effectors with members that are regulated by c-di-AMP. Riboswitches are regulatory sequences within an mRNA that affect the translation of the encoded gene upon binding of a small molecule. A common riboswitch in bacteria, which

is regulated by c-di-AMP, is *ydaO* [62, 63]. *YdaO* is a genetic off switch, which is binding c-di-AMP at picomolar concentrations and thereby inhibits translation [63]. In *Bacillus subtilis*, it was shown to exhibit an overall dimeric structure. Two similar three-way junction elements assembled in opposing orientations generate an internal pseudo-two fold symmetry of the RNA. Two molecules of c-di-AMP can bind at the interface between the dimers in a conformation that substantially differs from the conformation adapted in proteins. The c-di-AMP bound by the riboswitch adapts a more open conformation, in which the adenines are further apart and tilted [64-66]. Ligand specificity is achieved by well-defined interactions of the RNA, allowing a discrimination between c-di-AMP and c-di-GMP and c-GAMP. C-di-AMP stabilizes the structure of the RNA adapted in the riboswitch and thereby causes premature transcription termination [63]. The *ydaO* riboswitch is widespread in bacteria and is involved in the transcription regulation of genes involved in amino acid and potassium transport and in cell wall metabolism. It is mainly located in the 5´untranslated leader region; however, some *ydaO* riboswitches are located at the 3´untranslated region of putative mRNAs. This implies a so far unknown mechanism for regulation by the riboswitch [67]. Interestingly, in some organisms, the *ktrAB* operon, as well as the gene of KimA (*ydaO* gene) is regulated by the *ydaO* riboswitch. This implies a control of the activity of the KtrA and KimA proteins, as well as of their transcribing gene through the *ydaO* riboswitch.
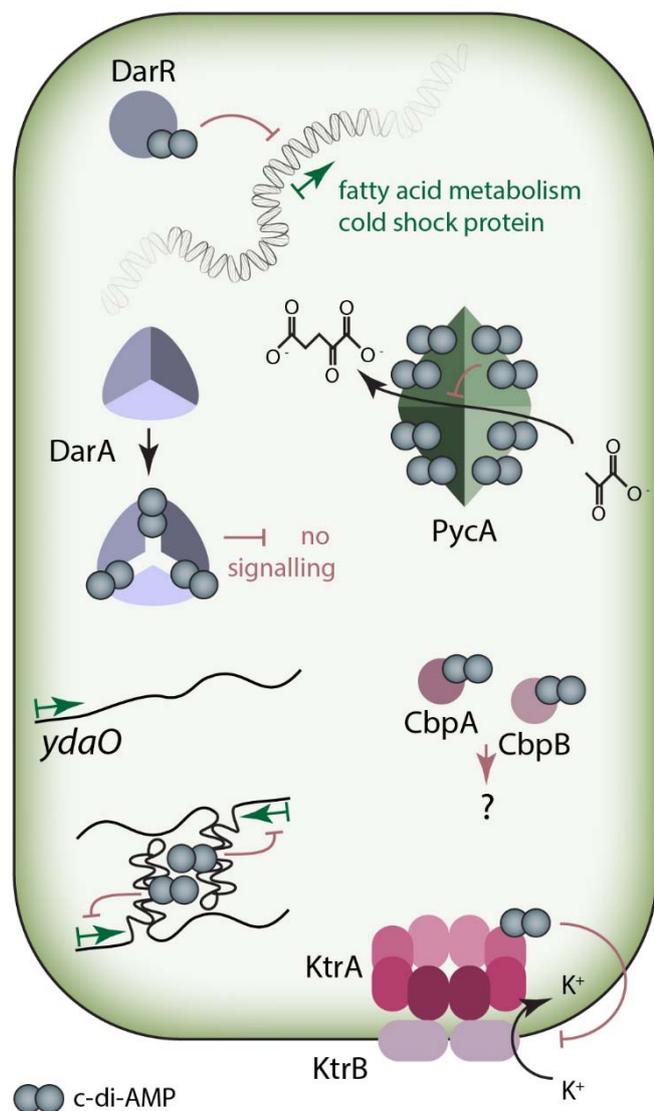

Fig.8: Overview of c-di-AMP effector proteins

Secretion of c-di-AMP is one possible pathway for bacterial cells to control their level of c-di-AMP. C-di-AMP from *Listeria monocytogenes* can be exported via four **multidrug transporters (Mdr)** of the major superfacilitator family [68, 69]. In cells that are deficient of these transporters, increased sensitivity to cell-wall targeting antibiotics can be detected. Although the effect of these antibiotics is associated with extracellular proteins, the addition of extracellular c-di-AMP cannot help to overcome the enhanced sensitivity. This indicates a role of the Mdr transporters in cell wall synthesis [68].

For some bacteria, like *Borrelia burgdorferi*, the knock-out of all known c-di-AMP degrading enzymes was shown to be lethal. This indicates that these bacteria are unable to secret c-di-AMP, and therefore lack the possibility of down-regulating their c-di-AMP level [70].

1.1.4. Degradation of cyclic di-AMP

The enzymes degrading c-di-AMP belong to the family of phosphodiesterases (PDEs). They enzymes bind c-di-AMP as a substrate, hydrolyse a phosphodiester bond and release the linearized compound 5´pApA. Some PDEs are additionally able to degrade the intermediate 5´pApA to two molecules of AMP.

c-di-AMP specific PDEs can be divided into two classes: DHH/DHHA1-domain comprise enzymes, termed by their catalytic triad Asp-His-His and HD-domain containing PDEs, using His-Asp as a catalytic centre [19, 71]. Figure 9 shows an overview of c-di-AMP degradation and export.

DHH/DHHA1-domain containing PDEs are present in many DAC-expressing species, and are very specific in the degradation of c-di-AMP with only weak activity towards c-di- [71]. They are further subdivided into multi- and single domain PDEs. Multi-domain PDEs, GdpP in *Bacillus subtilis*, are anchored to the membrane by two N-terminal transmembrane helices. The transmembrane domain is accompanied by a degenerated PAS-domain, a modified GGDEF domain and a catalytically active DHH/DHHA1-domain [71], which degrades c-di-AMP into 5´pApA [71]. On the other hand, single domain DHH/DHHA1-domain containing PDEs, as can be found in *Streptococcus pneumoniae*, contain only the soluble DHH/DHHA1 domain [17]. Some members of the single domain enzymes are able to degrade 5´pApA into two molecules of AMP. Other single-domain PDEs accomplish degrade ci-di-AMP to 5´pApA as well as 5´pApA to AMP, in a sequential manner. Some single-domain enzymes display a preference

for 5´pApA [17, 27]. In some organisms, like *Borrelia burgdorferi*, single domain PDEs are the only c-di-AMP degrading enzymes [27].

In multi-domain PDEs, the domains fused to the DHH/DHHA1 domain are suggested to have regulatory function. The membrane part of GdpP from *Streptococcus pyogenes* for example, is required for the enzymatic activity *in vivo*, though *in vitro* it seems to be dispensable [72].

The PAS domain is considered to play a role in regulating GdpP activity in the presence of heme [73]. Binding of heme leads to an inactivation of the DHH domain, however a complex of heme and NO can stimulate PDE activity [74]. The residues typically involved in heme-coordination are not conserved in the GdpP-PAS domain. However, the hydrophobic pocket of GdpP PAS was shown to be still able to bind heme [75]. GGDEF domains are originally described as c-di-GMP cyclases. However, in GdpP the catalytic residues for c-di-GMP synthesis are lost. Instead, it can bind and hydrolyse ATP [71]. Besides their function as c-di-GMP cyclases GGDEF domains can also have a regulatory
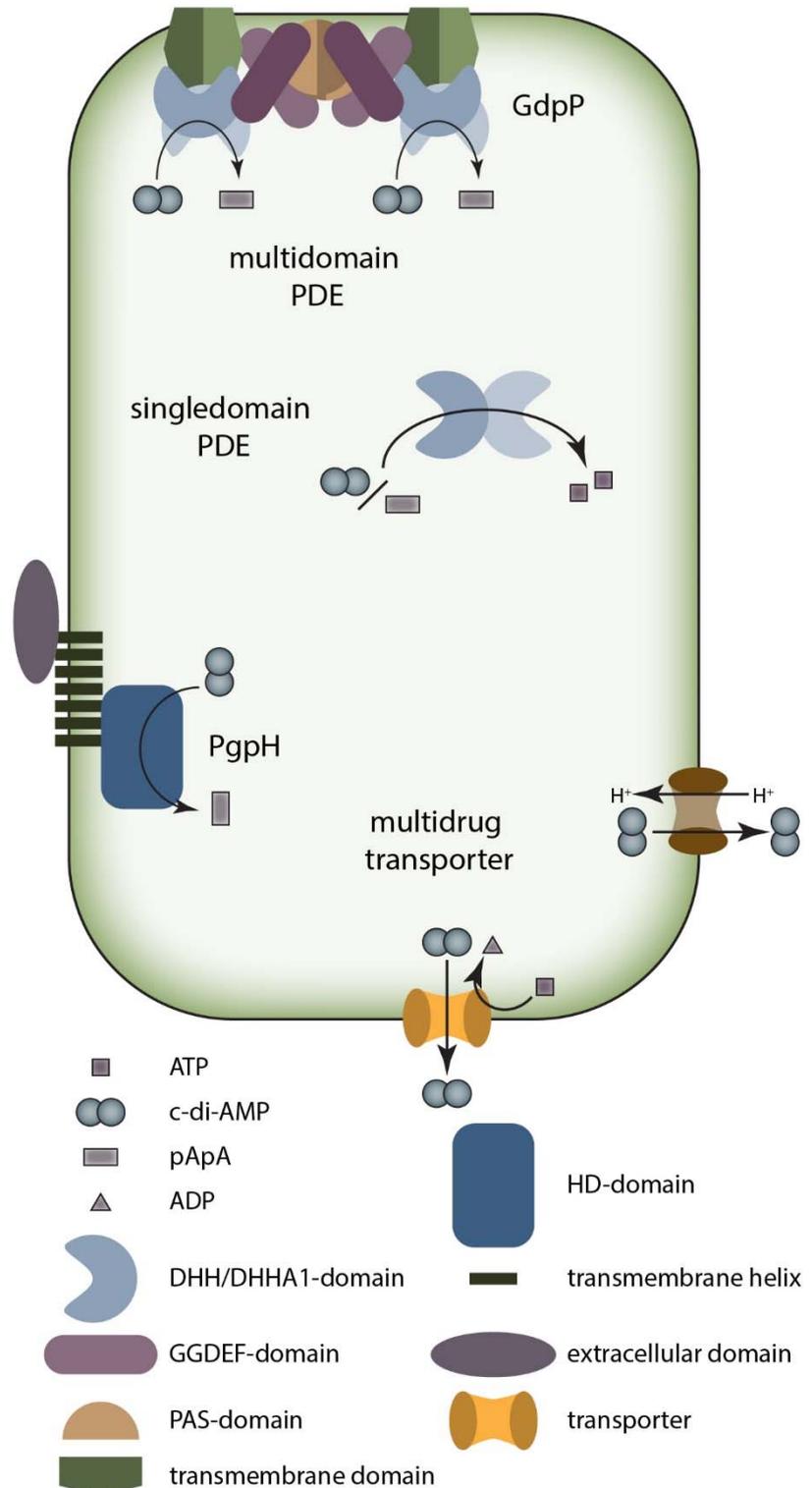

Fig.9: Overview of c-di-AMP export and degradation

function by binding c-di-GMP and allosterically inhibit the enzymatic activity [76, 77]. A similar inhibition mechanism of the GdpP GGDEF domain might thus be conceivable.

HD-domain containing enzymes as alternative family of PDEs are represented by PgpH from *Listeria monocytogenes* [19]. This enzyme consists of an N-terminal extracellular domain, seven transmembrane helices and the C-terminal HD-domain. The HD-domain is known to bind and hydrolyse nucleic acids [34]. It is composed of two lobes with the active site residues and the substrate localized at the interface [77]. Similar to GdpP, PgpH requires manganese for the catalytic activity.

Several organisms possess DHH/DHHA1-domain and HD-domain PDEs. This indicates a specialized function for each class [19]. Unlike GdpP, PgpH harbours an extracellular domain that allows enzymatic regulation via extracellular signals. Both PDE classes are regulated and inhibited by the stringent factor ppGpp [19, 24, 71], providing a link between the c-di-AMP signalling pathway and the stringent response. Increased levels of c-di-AMP trigger stringent response and lead to increased synthesis of ppGpp [24, 76, 78]. This in turn inhibits the degradation of c-di-AMP and functions as a feedforward loop to prevent decrease of the elevated signals of c-di-AMP, keeping the stringent response signal active.

1.1.5. Synthesis of cyclic di-AMP by DisA

C-di-AMP has been fist discovered in the crystal structure of DisA. Following studies have revealed the synthesis mechanism of c-di-AMP by DisA biochemically as well as structurally. Crystal structures of DisA from *Thermotoga maritima* have been solved in the ligand-free state as well as in ligand and product-bound state in the presence of different ATP analogues [37, 79].

The crystal structure shows a dimerization of the DAC domains with the active site in the middle forming an elongated shape. This dimer then forms a tetramer along its long axis. Thereby, the DAC domains form an active centre in the middle of the molecule with the HhH domains facing outward. The pre-reaction state was analysed by crystallization with non-reactive ATP analogues (pdb: 4YVZ, 4YXJ). The role of the conserved DGA (amino acids 75-77) and RHR (amino acids 108-110) motifs could be revealed in coordinating the metal ion and the 2′OH group of the ribose, as well as forming hydrogen bonds to the β- and γ-Phosphate or the O4 of the Ribose, respectively. The structure furthermore showed a polarization of the γ-Phosphate through S127, R128 and R130, facilitating the nucleophilic attack in the α-

Phosphate of the second ATP molecule. The post-reaction state represents DisA bound to c-di-AMP (pdb: 3C21, 4YXM) [79]. Since the first discovery of c-di-AMP occurred through the crystallization of DisA [37], this shows a very tight binding of the product to the enzyme. In the pre-reaction states, most interactions occur via the phosphate groups, consequently leading to a less strong binding of the c-di-AMP molecule, than of the two ATP molecules. This indicates that the bottleneck of c-di-AMP synthesis is the dissociation of the product from the enzyme, rather than the catalytic reaction itself. To exit the octamer, c-di-AMP has to diffuse to the centre of the molecule and then pass through the exit tunnel. This tunnel harbours a diameter of approximately 7.8 Å, tightly fitting c-di-AMP. Interestingly, mutations of the three Arginine residues (128-130) in a loop bordering the exit tunnel to less bulky residues resulted in an increased reaction rate. This strengthens the hypothesis of the exit of the product to be the rate-limiting step in the reaction. Structural analysis of the entire DisA octamer revealed no further structural changes within the molecule, and the adjacent domains remain unaltered during the reaction cycle [79].

DisA is inactivated by non-standard DNA and especially Holliday junctions [37]. This led to the conclusion that DisA scans the integrity of the bacterial genome and stops sporulation, in case of DNA lesions. This allows the cells to repair their genome and prevents the formation of non-viable offspring with a defective. The mechanistic details of c-di-AMP production by DisA are very well studied [79]. The regulation of DisA by DNA lesions, however, remains elusive so far. It is assumed, that binding of the HhH-domain to DNA leads to a conformational change within the DAC domains, which is communicated through the helical bundle [37]. The crystal structures solved so far from *Thermotoga maritima* (tma) DisA capture the HhH-domain in a conformation, which is unable to bind DNA. This can be seen in a direct comparison of the HhH-domain of DisA with other Helix-hairpin-helix motifs from different enzymes such as RuvA (pdb:1c7y; Fig. 10 adapted from [37]). The HhH domains (shown in orange) in the tmaDisA structure adopt a
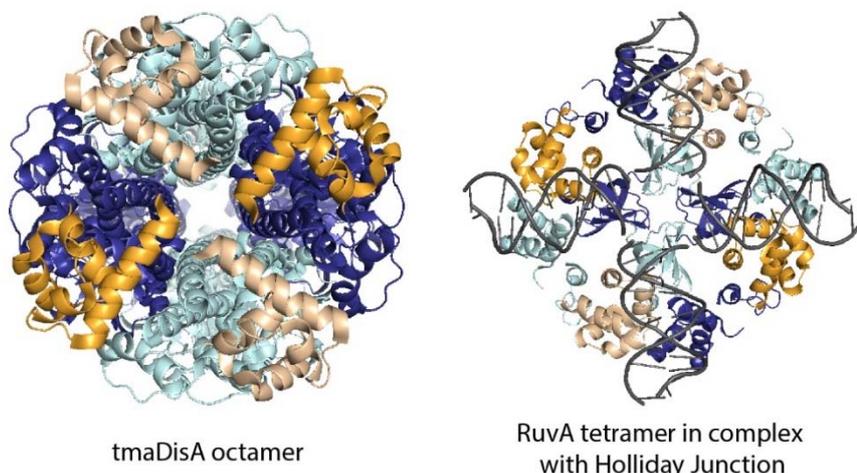

tmaDisA octamer

RuvA tetramer in complex with Holliday Junction

Fig.10: Comparison of tmaDisA and RuvA complexes with DNA

27

conformation perpendicular to the one in RuvA and would therefore collide with the DNA. Crystallization attempts of DisA in the presence of DNA remained unsuccessful, leaving the open question whether binding of DNA to the outer domains of DisA leads to an overall structural change in the molecule. This conformational change in the molecule could then lead to an inactivation of DisA. DisA itself fulfils several functions within the cell by producing c-di-AMP, and bacteria often additionally express the DACs CdaA and CdaS. The inactivation of one molecule DisA by binding to DNA lesions can therefore have only little impact in the overall cellular pool of c-di-AMP. It remains an open question how the recognition of the loss of DNA integrity by DisA is propagated to cause an effect in sporulation progression.

<u>Aim of this project</u>

So far, several mechanisms have been proposed (Fig. 11): One possibility would be that binding to DNA leads to disassembly and dissociation of the octameric DisA complex. The Holliday junction is then free to be bound by another intact DisA, leading to a more significant effect in cellular c-di-AMP level by the disassembly of an increased amount of octamers. Alternatively, the DisA-DNA complex could be recognized by an effector protein, which then prevents sporulation of the cell. A potential interaction partner is the hexameric protein RadA. RadA is transcribed from the same operon as DisA and a physical interaction between both proteins has been reported in previous studies [47, 48]. In addition, RadA is involved in DNA
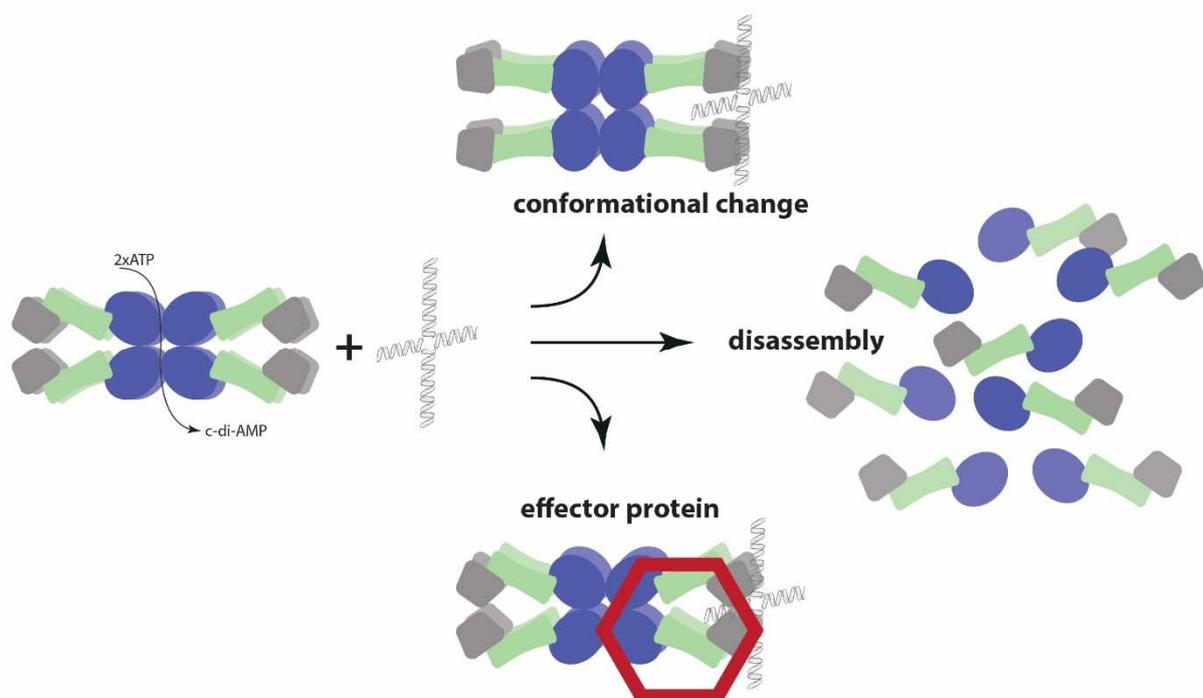

Fig.11: Potential regulatory mechanisms for DisA inactivation through Holliday Junction

repair as it processes Holliday junction intermediates. [49] To gain further insights into DisA signalling, a non-crystallographic structure of DisA was considered to be of help. Such a structure could help to proof, that the conformation of the HhH-domain in the crystal structure of tmaDisA is not artificial, i.e. stabilized through the contacts of crystal packing. The first aim of this project was therefore, to obtain a high-resolution structure of tmaDisA using electron microscopy, hopefully leading to an unbiased solution-like structure.

Crystallization attempts of DisA homologues from species other than *Thermotoga maritima* were so far unsuccessful. Especially non-thermophilic species seem to have an increased flexibility in their respective DisA octamers. This led to the idea that a structural analysis of non-thermophilic DisA species might help to gain insight into further possibilities for regulation of DisA. *Bacillus subtilis* was chosen as one non-thermophilic organisms, as it is one of the major structural model organisms for bacterial nucleotide second messengers. A pathogenic bacterial strain was considered to be of interest, as it might help to gain insight into the role of c-di-AMP in pathogenic organisms. The second aim of this project was hence to gain structural insights into DisA from *Bacillus subtilis* (bsu) and *Mycobacterium tuberculosis* (mtub) using electron microscopy. This would allow a direct comparison of the structure from tmaDisA with bsu and mtubDisA by applying the same method.

Another open question concerning in the field of DisA is the speed regulation of c-di-AMP synthesis and the rate-limiting step. A mutant of DisA with the three arginine residues in the loop region mutated to alanine was shown to increase the activity of the DAC domain, leading to the hypothesis, that the arginine residues reduce the diameter of the exit tunnel, and hence influence the dissociation rate of the product from the active site. The aim of this study was to gain insight into the influence of the bulky residues on the activity of the enzyme by analysing the diameter of the exit pore in the crystal structure.

1.1.6. Degradation of cyclic di-AMP by GdpP

Up to date, the c-di-AMP degrading enzymes have been structurally characterized only on the level of individual domains ([80], pdb:4zmm, [81]). The role of the adjacent domains, their interplay with the DHH/DHHA1 domain and their means of regulation is poorly understood. So far, insights into the mechanism of degradation as well as of substrate specificity of c-di-AMP through DHH/DHHA1 domains are gained using x-ray crystallography and mutagenesis studies [81, 82]). These structural studies reveal a substrate specificity of the molecules

through the formation of three distinct binding positions. Whereas both enzymes, degrading c-di-AMP to AMP and in enzymes degrading 5´pApA to AMP bind the first adenine in the same site, the binding of the second adenine is unique to the specificity of the enzyme. Mutational studies of the DHH/DHHA1 domain of GdpP from *Staphylococcus aureus* revealed, that degradation of c-di-AMP to 5´pApA could be altered to allow full degradation to AMP by enlarging the cavity required for 5´pApA binding and degradation [82]. However, the question remains, why the substrate specificity of the multi- and single-domains PDEs correlates with the localization of the PDEs within the cell, as well as the role of the additional domains remains to be elucidated. Structural information on GdpP type PDEs might help to shed light on this open question of the degradation of c-di-AMP. A comparison between the DHH/DHHA1 domains of the soluble, single-domain PDE with the catalytic domains of multi-domain GdpP in the context of the adjacent domains might help to find differences causing substrate specificity. Capturing multi-domain GdpP in pre- and post-reactive state would help to gain insight into the regulatory function of the neighbouring domains. Addition of binding partners to the adjacent domains, such as nucleotides, heme or nitric oxid could help understanding the regulation of GdpP through cellular signals.

Aim of this project

The aim of this project was therefore, to obtain a structure of the soluble part of GdpP of *Streptococcus pneumoniae* using cryo-electron microscopy. A structure of GdpP in its apo state or bound to the substrate could function as a framework for further analysis.

1.2. Methods

1.2.1. DisA-constructs from different species

In the following buffers used in the experiments are abbreviated with T for Tris, N for sodium chloride, G for Glycerol and I for Imidazole. Values in brackets state the pH and indexed numbers give the concentration in mM.

Purification of DisA

DisA constructs were purified as published in [37]. All DisA constructs were expressed in BL21 (DE3) Rosetta cells. Cells were grown at 37°C to an OD 600 of around 0.6. Following induction with 0.2 mM Imidazole temperature was reduced to 18°C for overnight expression. Cells were resuspended in $T_{50}(8)N_{300}I_{10}$ and lysed by sonication of twice 8 min. The lysate was loaded twice onto Ni-NTA beads and stringendly washed with lysis buffer, and in the case of mtubDisA with a high-salt ATP-buffer ($T_{25}(8)N_{1000}G_{10}I_{10}ATP_2$). The Imidazole concentration was stepwise increased to 30, 60 and 300 mM. The elution was detected by SDS-Page analysis, and sample-containing fraction were pooled and submitted to Dialysis in $T_{50}(8)N_{300}DTT_2$. Sample was concentrated if necessary before injection into size exclusion chromatography. Therefore a S200 column was used equilibrated with $T_{50}(8)N_{200}DTT_2$. Sample quality was checked by SDS-page analysis.

Crystallization of tmaDisA 3xR mutant

Crystals were grown in hanging-drop vapour-diffusion geometry at 20 °C over a reservoir of 300μL containing 0.2 M sodium Acetate, 30% MPD, 0.1 M Tris at pH8. After several days of incubation suitable crystals were obtained in drops of 1μL reservoir and 1μL protein at 6 mg/mL were soaked in cryo-protectant of 6μL reservoir supplemented with 1μL MPD for 40 sec before flash-cooling in liquid nitrogen. Diffraction data were collected to 2.2 Å at P13 at PETRA3 in Hamburg.

Structure determination and Refinement

The obtained data was indexed using XDS and scaled with XSCALE [83, 84]. For Molecular Replacement Phaser [85] within the CCP4 program suite [86] was used with ligand-free wildtype DisA as search model. Following autorefinement in phenix refine [87], manual model building and subsequent refinement steps were performed in Coot [88] in cyclic rounds.

1.2.2. EM studies of proteins in the cyclic di-AMP pathway

<u>EM-grid preparation</u>

For DisA samples were diluted in size exclusion chromatography buffer to a concentration of 0.4 mg/mL (tmaDisA) or 0.8 mg/mL (bsu and mtubDisA). 4.5 µL of samples were applied to glow discharged R2/1 Cu-200 mesh holey carbon grids (Quantifoil) and immediately blotted with a blot-force and time of 8 and 8 at a humidity of 95%. Cross-linked sample of wtGdpP and the first attempt of the inactive mutant was vitrified at a concentration of 0.1 mg/mL and 0.2 mg/mL respectively, in a volume of 4.5µL with a blotforce of 8 and blotting time of 8s. The inactive mutant of GdpP was diluted to a concentration of 10 mg/mL with 0.05% NP40 and 4 µL were applied to the grid. Blotforce was 4 for 3.5 seconds. Plunge-freezing occurred into liquid ethane cooled by liquid nitrogen using a Vitrobot Mark IV (FEI).

<u>Data acquisition</u>

Technical details of data acquisition are displayed in table 1.

Table 2: data collection details

Dataset	tmaDisA	mtubDisA	bsuDisA	GdpP	Gdpp+NP40
Microscope	Titan	Arctica	Arctica	Halo	Titan
Energy filter	GIF quantum energy filter (20 e^-V)	no	no	no	GIF quantum energy filter (20 e^-V)
mode	integrated	integrated	integrated	integrated	integrated
Software	SerialEM	EPU	EPU	EPU	SerialEM
camera	K2	Falcon III	Falcon III	Falcon III	K2
Pixel size	0.82 Å	1.612 Å	1.612 Å	1.43 Å	1.35 Å
Dose	54 e/ $Å^2$	49.35 e/ $Å^2$	50.5 e/ $Å^2$	50.9 e/ $Å^2$	e/ $Å^2$
Dose/frame	2.16 e/ $Å^2$	1.23 e/ $Å^2$	1.26 e/ $Å^2$	1.27 e/ $Å^2$	1.27 e/ $Å^2$
exposure	5 s	2.01 s	2.01 s	2.01s	5 s
frames	25	40	40	40	25
defocus	-1.4 to -3.5	-1.5 to -4	-1.5 to -4	-2.5 to -4	-1.4 to -3.5
Numer of micrographs	8539	923	1362	269	3968

Data processing

Micrographs were gain normalized (for datasets collected at Titan), aligned and dose-weighted using MotionCor2 [89] before importation into relion 2.1 [90]. Estimation of defocus was performed using ctffind4 [91] and Gautomatch (https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/) was used for automated particle picking. Particles were selected by several rounds of reference-free 2D classification in the relion 2.1 software to remove noisy or uncentered particles as well as contaminations. For DisA datasets, the crystal structure of tmaDisA was used at 60 Å as a template for 3D classification. For GdpP an initial model was calculated in the relion software using the first 300 micrographs. After various rounds of 3D classification, the best model was subjected to auto-refinement and subsequent Post-processing using a mask generated from the 3D model or the autorefinement. In case of tmaDisA movie processing was performed.

Model building

During the course of this thesis, two EM-structures were build. For tmaDisA, the original crystal structure was used as starting model. In the case of GdpP two homology models were modelled, one for the DHH domain using phyre2 and one for the GGDEF domain based on a model calculated with swiss-model [92] using RbdA from *Pseudomonas aeruginosa* as homologue. The two models were fitted into their respective densities and united using UCSF Chimera [93]. This procedure was performed for chain A and chain B, which were then duplicated and used for chain C and D respectively. The model and the EM density obtained from the post-processing step were imported into the Phenix software package [87] and subjected to real-space refinement using global minimization and B-factor refinement. The model was furthermore improved by manual building in Coot [88].

1.3. Results

1.3.1. Crystallization of tmaDisA3xR mutant

To analyse the impact of the three Arginine residues R128-130 in DisA, crystal screens were set up of the corresponding mutant (termed tmaDisA3xR).

Fig. 12: Crystals of tmaDisA3xR

Crystallization of tmaDisA3xR mutant was successful with the growth of long and thick rod-shaped crystals (Fig.12). The diffraction pattern obtained at the synchrotron allowed data collection and structure determination. The resolution limit was chosen by I/σ 2 cut-off. Data collection and refinement statistics are shown in table 2.

Table 2: data collection and refinement statistics
(highest resolution shell is shown in parentheses)

| Data Collection			Refinement	
Beamline	EMBL P13		Resolution	38.5-2.5
Wavelength (Å)	0.9762		No. of Reflections	47551
Spacegroup	P42(1)2		R_{work}/R_{free}	17.1 / 22.1
cell dimensions			Molecules/ASU	2
a, b, c (Å)	108.9,108.9,166.7		**No. of atoms**	
a, b, g (°)	90,90,90		Protein	5572
Resolution	50-2.25 (2.31-2.25)		nucleotides/ligands	44 / 12
CC$_{1/2}$	99.9 (62.7)		water	207
Rmeas (%)	12.4 (169)		**RMSD**	
I/sI	17.2 (2.05)		Bond lengths (Å)	0.007
Completeness (%)	98.5 (98)		Bond angles (°)	0.880
Redundancy	14.5 (14.1)		**Ramachandran (%)**	
			favoured	97.84 %
			allowed	2.16 %
			outliers	0.00 %
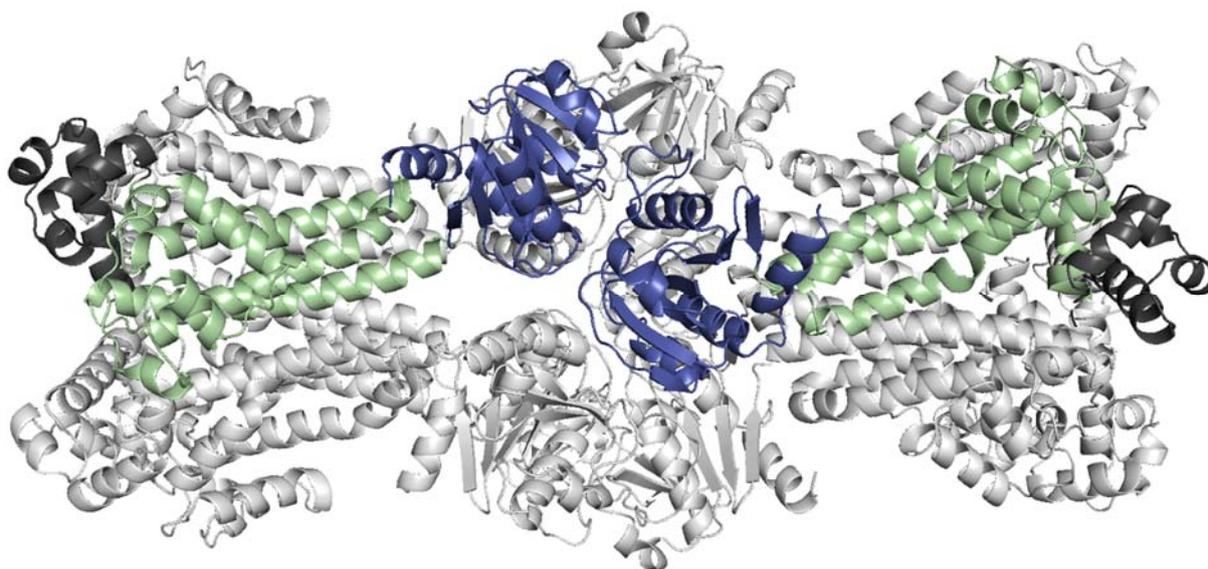
Fig.13: Crystals structure of tmaDisA3xR mutant

The structure revealed the formation of an octamer, as expected for DisA (Fig.13). Furthermore, the composition of inward facing DAC domain and outward facing HhH domain separated by the helical bundle remained similar to the wild type. The DAC domain adapted the shape of the globular α/β-domain composed of one β-sheet sandwiched between the five α helices, as known for c-di-AMP synthesising DAC domains. In direct comparison to the wild type tmaDisA structure, no differences besides the three mutated amino acids are visible at first glance. This is also represented in the root mean square deviation of 0.21 between the wild type structure and the mutant, calculated using PyMOL (Schrödinger LLC). This supports the finding that the only difference between the structures reside in the three mutated arginine residues. In the obtained crystal structure, c-di-AMP is very well visible as ligand, indicating an active enzyme. (Fig. 14)
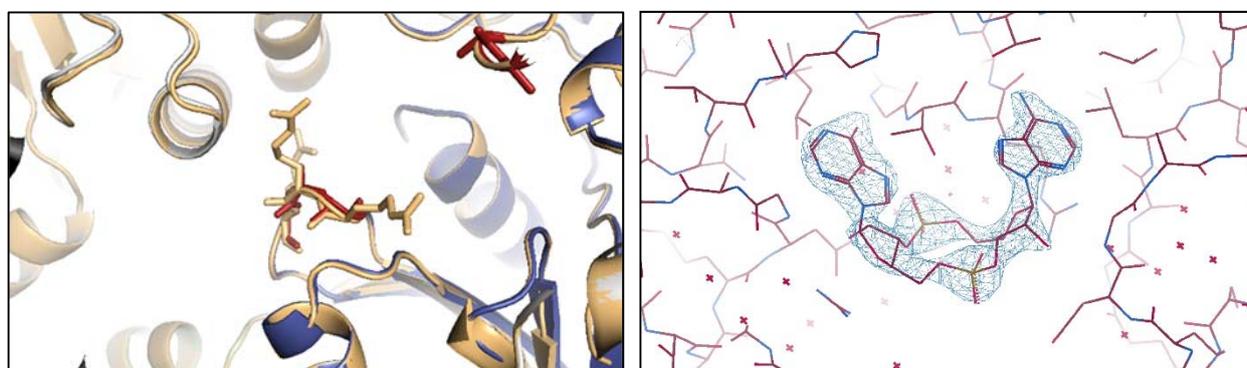

Fig. 14: left: Close up on the three mutated arginine residues. The orange structure shows wild-type DisA, the blue structure represents the mutant with the three Alanine residues shown in red. Right: Polder-map of the tmaDisA3xR structure, in red shown the crystal structure and in blue the electron density clearly showing occupancy for c-di-AMP

The obtained structure was analysed using the MOLE 2.5 online toolkit [94] to compare the channel formation in the mutant tmaDisA3xR structure with the wild type structur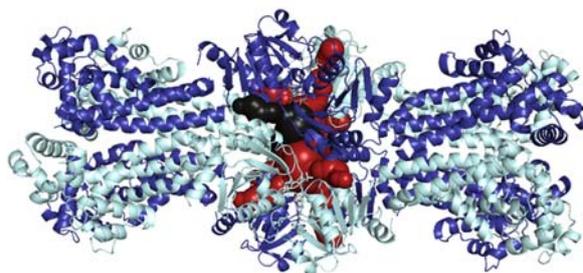e. The overall channel formation is displayed in Fig. 15. The exit channels for the synthesised c-di-AMP are situated between the DAC domain

Fig. 15: Channels in tmaDisA3xR mutant

and the adjacent helical bundle. C-di-Amp has to diffuse to the centre of the molecule before it can exit the octamer between any of the eight channels. A comparison with the channel formation in wild type DisA, as displayed in Fig. 16, shows a significant increase in channel diameter at a similar probe radius.
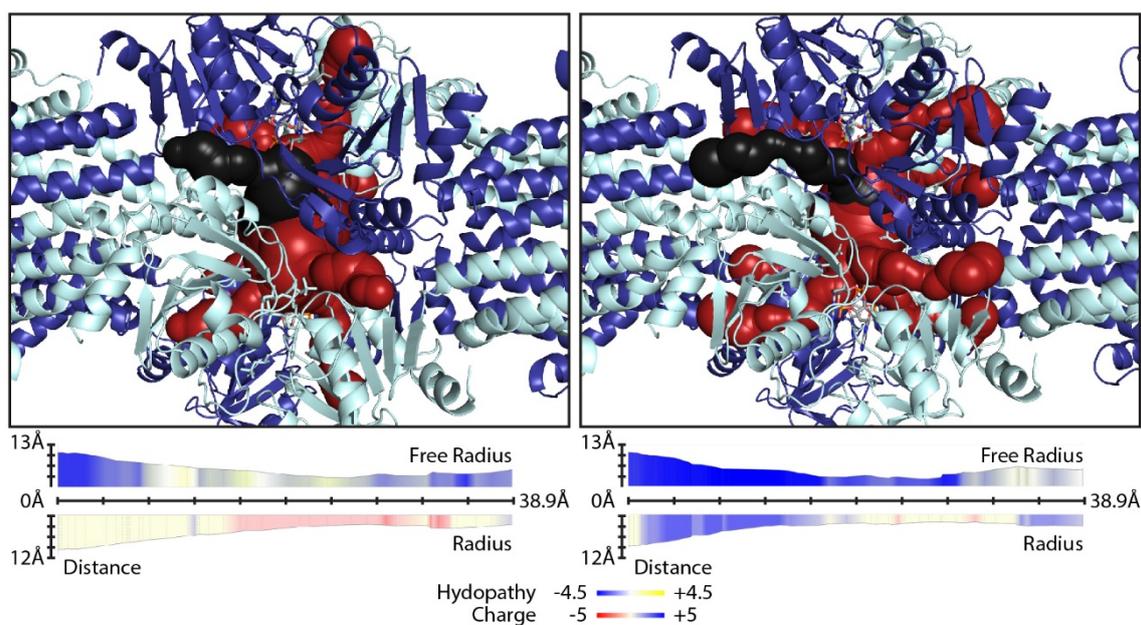

Fig. 16: Structural (top) and graphical (bottom) representation of channel formation in tmaDisA3xR mutant (left) in comparison with wild type DisA (right)

1.3.2. EM structures of DisA from different species

<u>Cryo-EM structure of tmaDisA</u>

tmaDisA is known to bind DNA via its HhH-domains at the C-terminus [37]. However, comparison of its crystal structure with structures of homologues in complex with DNA – such as RuvA (1c7y) reveals a potential clash in the HhH-domains of DisA, i.e. DNA-binding is impossible in the conformation observed in the crystal structure. One possible explanation for the wild type structure could be an artificial conformation stabilized through crystal packing. To exclude this, a solution structure sample of tmaDisA was vitrified on EM-grids and a dataset recorded on a Titan Krios using a K2 Summit Camera in post-GIF recording mode using a nominal pixel size at the specimen level of 0.82 Å (Fig. 17 top left). The obtained 8539 micrographs were gain normalized, dose weighted and frame-aligned using MotionCor2. After defocus estimation with CtfFind4, the first 1000 micrographs were used for particle picking in
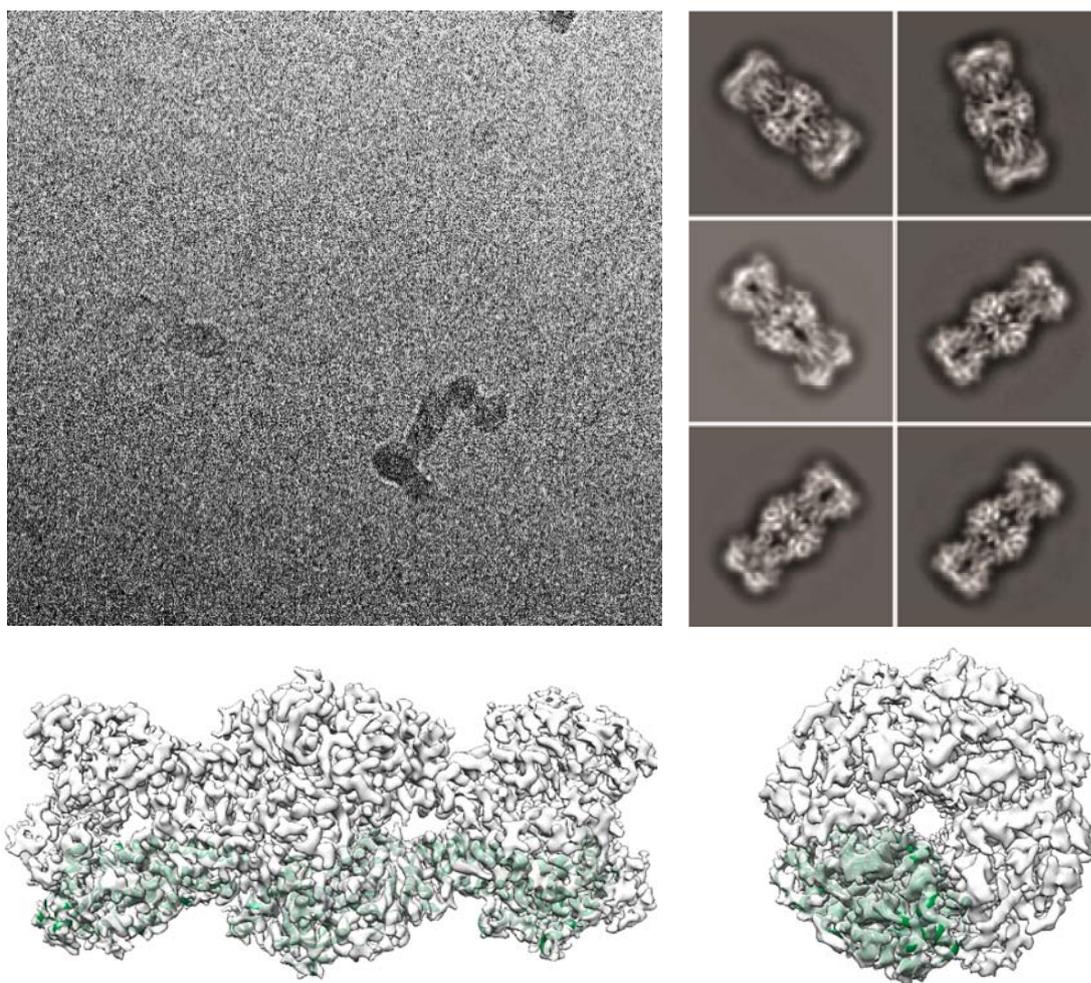

Fig. 17: top left: Selected EM micrograph of the tmaDisA grid; top right: Selected 2D classed of tmaDisA; bottom: Final EM structure of tmaDisA in side and top view with a fitted dimer of the crystal structure

Gautomatch. Selected 2D classes obtained after extraction with a binning factor of 2.34 and 2D classification were used for template based picking in relion, to enhance accuracy and particle centering of the picking with Gautomatch on the full dataset. Extraction of the second round of picking yielded around 1.3 million particle candidates, which were subjected to two rounds of 2D classification. The remaining 360 thousand particles were used to calculate a 3D model by 3D classification in one class, applying C4 symmetry and using the crystal structure of the octamer as template. This template was then utilized as a reference for classification of the 2D selection into 5 classes. The best classes (Fig. 17 top right) were selected and reextracted as unbinned particles. 3D Classification into 1 class with the crystal structure as reference was again used for the creation of a reference as well as a mask. Both, 3D structure and mask were then used for 3D refinement of the extracted, unbinned particles. The refined particles were then subjected to movie refinement and the last frames 16-25 were excluded from further data processing. After motion track alignment and per-frame B-factor estimation with the previous refinement as reference, 3D refinement was performed, followed by the creation of a new mask. This mask allowed solvent flattening in the next round of Refinement. This refinement was used in a final post-processing step with a B-factor of -74 $Å^2$, yielding a resolution of 3 Å (Fig. 18 left). This electron density at atomic resolution was then used for building the structure using coot, utilizing the wild type crystal structure (pdb:3c1y) as template (Fig. 18 bottom). This allowed easy comparison of the outer HhH-domains. An overview of the processing procedure is shown in Fig 19.
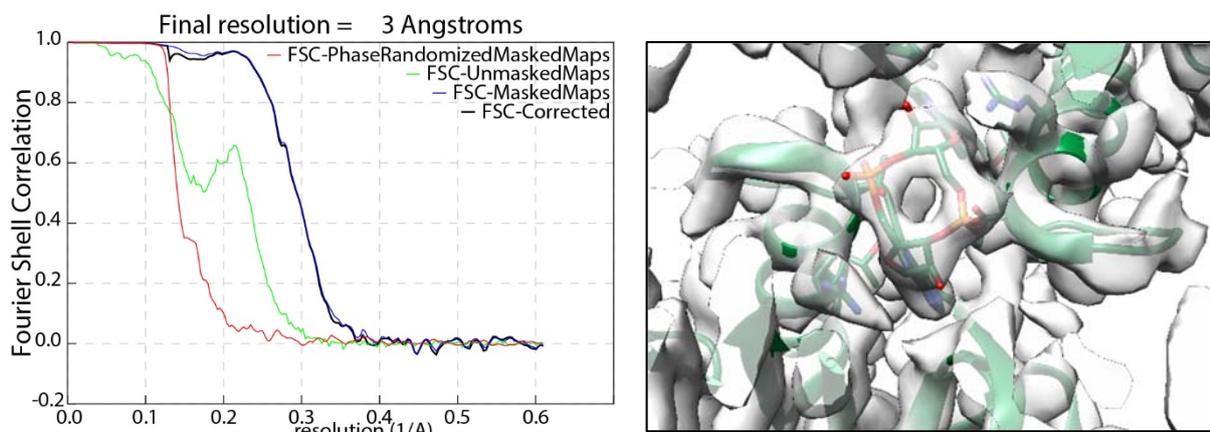

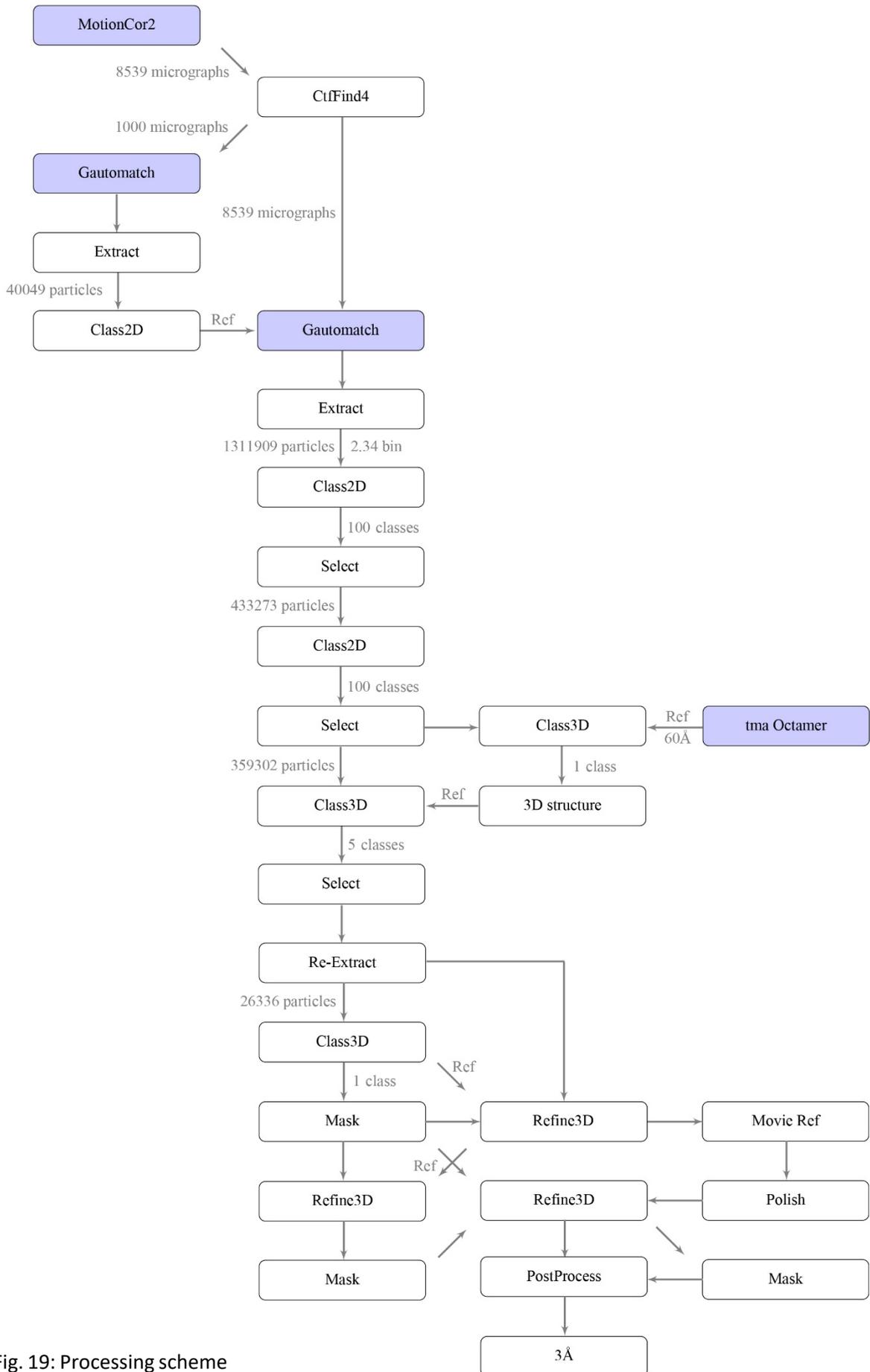Fig. 18: left: FSC-blot of tmaDisA dataset; right: Active centre of tmaDisA with c-di-AMP build in the EM density

Fig. 19: Processing scheme of the tmaDisA dataset

As expected for a density at 3 Å, side chains can be unambigously fitted in their densities. C-di-AMP can be identified and build in the DAC domain (Fig. 18 right). The most interesting domain however is the HhH domain. Here the map shows a clear density for the helix in question, confirming the conformation shown in the crystal structure (Fig. 20).
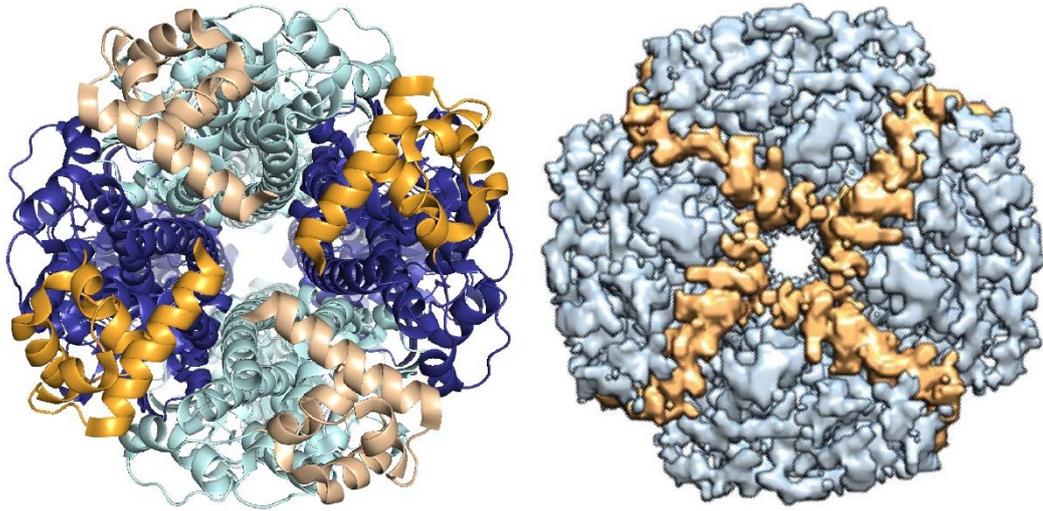

Fig. 20: left: Crystal structure of tmaDisA with monomers coloured in dark and light blue alternatingly, HhH domains are coloured in orange and light orange; right: EM structure of tmDisA with the long helix of the HhH-domain coloured in orange

Purification and cryo-EM structure of bsuDisA

Thermotoga maritima is a thermophilic organism, growing at 95°C. The structural rigidity of tmaDisA might therefore be accounted for by its structural studies at comparably unphysiological conditions. This is also shown by the inability of analysing the melting temperature of tmaDisA with standard unfolding analysis, such as differential scanning fluorimetry (DSF). The temperature analysed in DSF typically ranges from 4°C to 95°C. In the case of tmaDisA no unfolding occurs within the measuring range. This indicates unusual behaviour of the protein for *in-vitro* experiments, especially structure determination, as structural analysis at 95°C is not possible. To overcome this potential bias of a thermophilic species, other mesophilic organisms were chosen for structural analysed. *Bacillus subtilis* was used as a model organism for EM studies on bsuDisA, as previous crystallization attempts of other species than tmaDisA were unsuccessful, including *Bacillus subtilis*. After Ni-affinity purification, the SEC profile of a Superdex 200 10/300 showed one clear peak at the elution volume corresponding to an expected hydrodynamic radius for the DisA-octamer complex (Fig. 21). The sample was subsequently analysed for structure determination using cryo-EM.
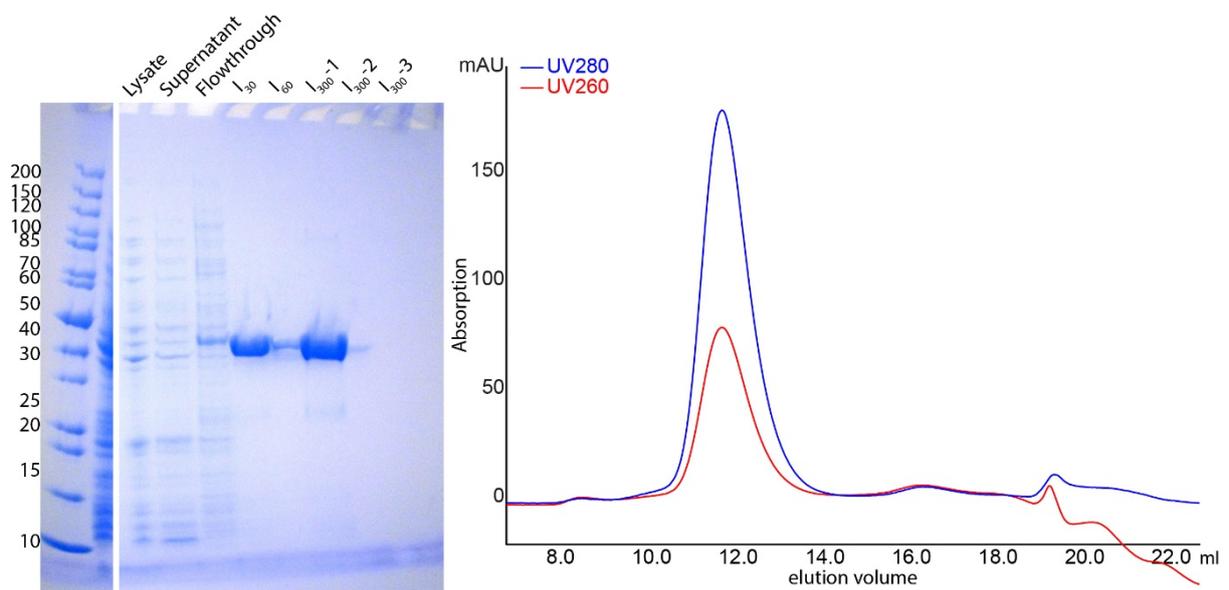

Fig.21: Purification of bsuDisA with SDS PAGE of an affinity purification and a chromatogram of size exclusion chromatography
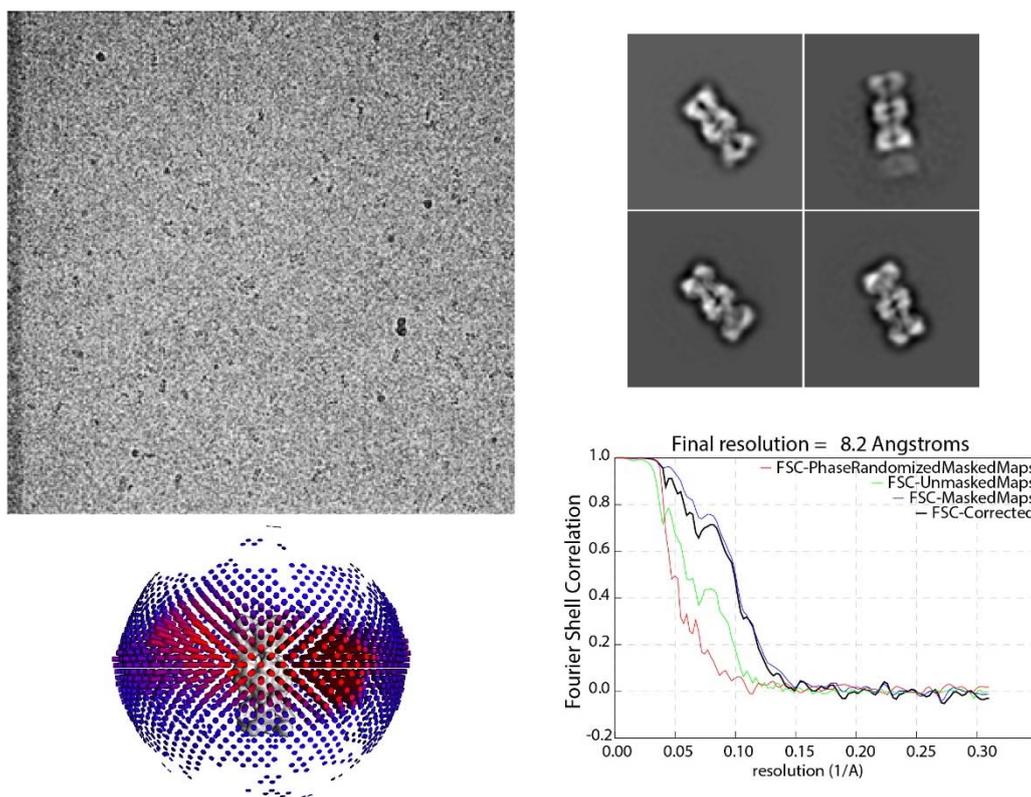
Fig. 22: top left: Micrograph of bsuDisA grid, top right: selected 2D classes, bottom left: Angular distribution, bottom right: FSC-blot

Data acquisition yielded 1.300 micrographs (Fig 22 top left) with around initial 421,000 particle candidates. By two rounds of 2D classification (Fig. 22 top right), the number of particles was reduced to 34,000 particles. The tmaDisA structure was used as template for one initial 3D class, which was subsequently used as reference in 3D classification yielding around 15,000 particles. After refinement, a resolution of 8.2 Å was achieved with a B-factor of -780 Å^2. The structure is shown in Fig. 23. A flowchart of the processing procedure is shown in Fig. 24.

Fig. 23: EM structure of bsuDisA

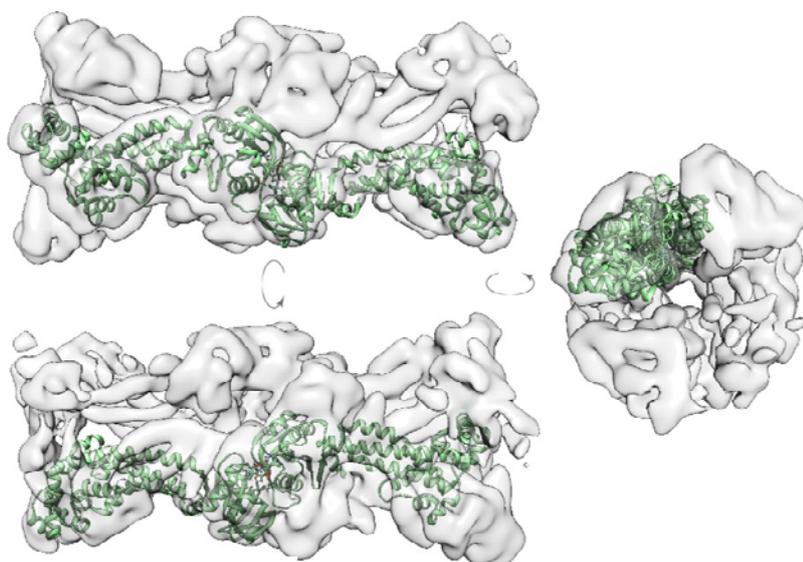The dataset of bsuDisA revealed a barrel shaped structure – as expected for a DisA complex, with the overall composition similar to the crystal structure of tmaDisA. In the structure of bsuDisA a loss in symmetry can be detected. This is due to an increased flexibility in one of the monomers. The HhH domain possesses a better visibility than the inward facing DAC domain for the flexible monomer. This variation in association of the monomer cannot be detected in the 2D classification. To circumvent a bias in selection of chosen classes, the obtained structure was confirmed by two independent processing procedures of the dataset. Both reconstructions led to the same model indicating that the obtained model does not only represent a minor conformation in a heterogenic sample. Furthermore, the construction of a processing artefact through preferred orientation can be excluded by a sufficient angular distribution around the long axis of the barrel.
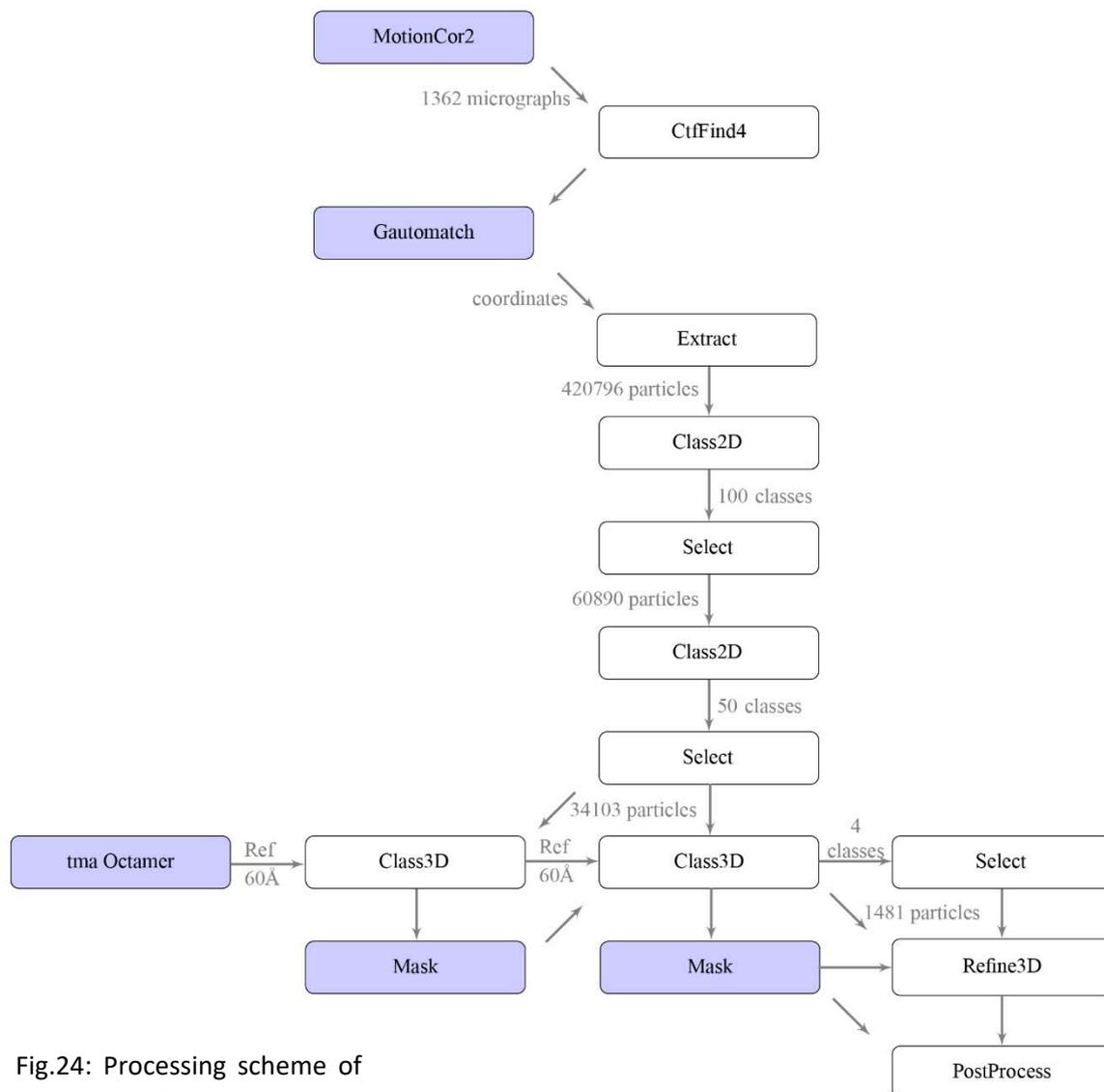

Fig.24: Processing scheme of the bsuDisA EM dataset

43

Purification and cryo-EM structure of mtubDisA

Another organism, which was chosen for analysis, was the pathogenic mycobacterium tuberculosis. Therefore, mtubDisA was cloned into a pET28 vector between the NdeI and SacI restriction sites, yielding a plasmid containing an N-terminal His-tagged mtubDisA sequence. The protein was expressed inBL21 (DE3) Rosetta cells and purified by Ni-NTA affinity purification, yielding sufficient amounts of clean protein. Subsequent exclusion chromatography on a Superdex200 16/600 column revealed two major species by the formation of a shoulder in the chromatogram, indicating instability of the protein or some degree of unfolding, showing decreased rigidity or stability in comparison to bsuDisA. The main peak fractions corresponding to DisA elution were pooled and stored at 8°C (Fig. 25).
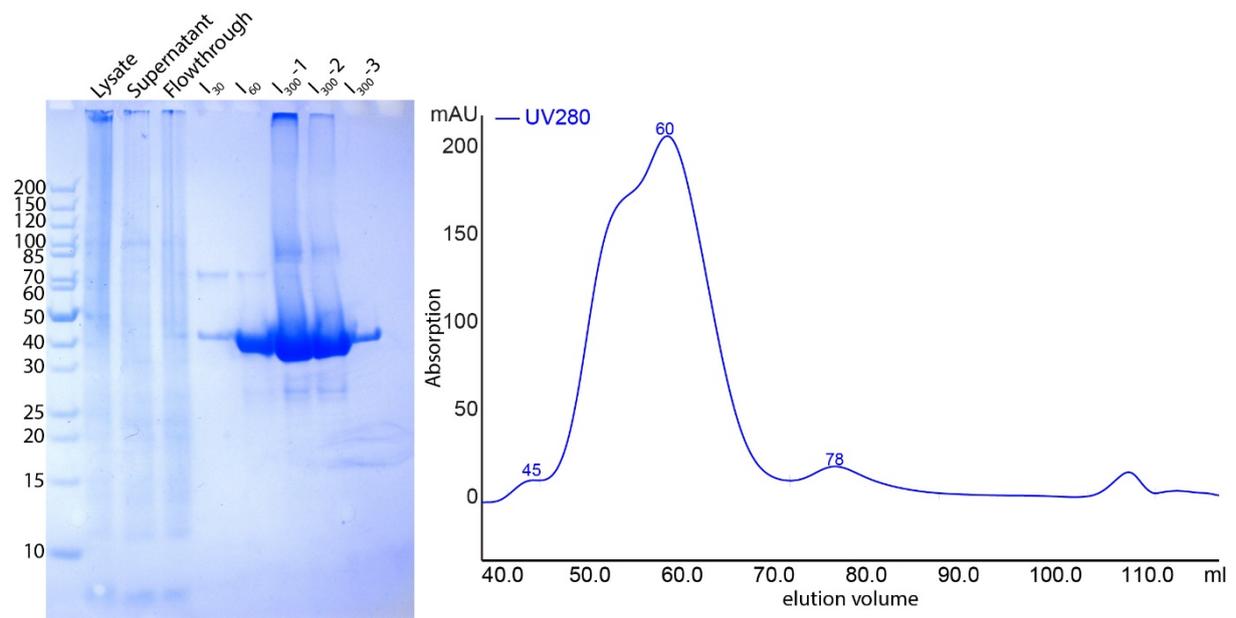
Fig. 25: Purification of mtubDisA with SDS PAGE of affinity purification and chromatogram of size exclusion chromatography

For quality control, directly after elution, as well as 10 days after storage the sample was analysed by recording a melting curve based on differential scanning fluorimetry, using a Tycho NT6 (NanoTemeper Technologies). The direct comparison revealed an increase in the amount of folded protein upon storage at 8°C for 10 days. This is indicated by the reduction of the ration of 350nm/330nm Absorption at the initial starting measuring points in the unfolding profile. This shows an improvement of sample quality over time, which could

Fig. 26: top row: Profile of melting curve analysis in the Tycho NT6; middle row: Micrographs of the samples on day 1 and on day 10; bottom row: 2D classes of the two samples

furthermore be seen in the EM-micrographs, as less particles in the 2D classification appear to have lost a subunit. However, this difference in sample quality might also be accounted for by slight differences in grid preparation, such as variations in humidity of the blotting-paper or application time. (Fig. 26)
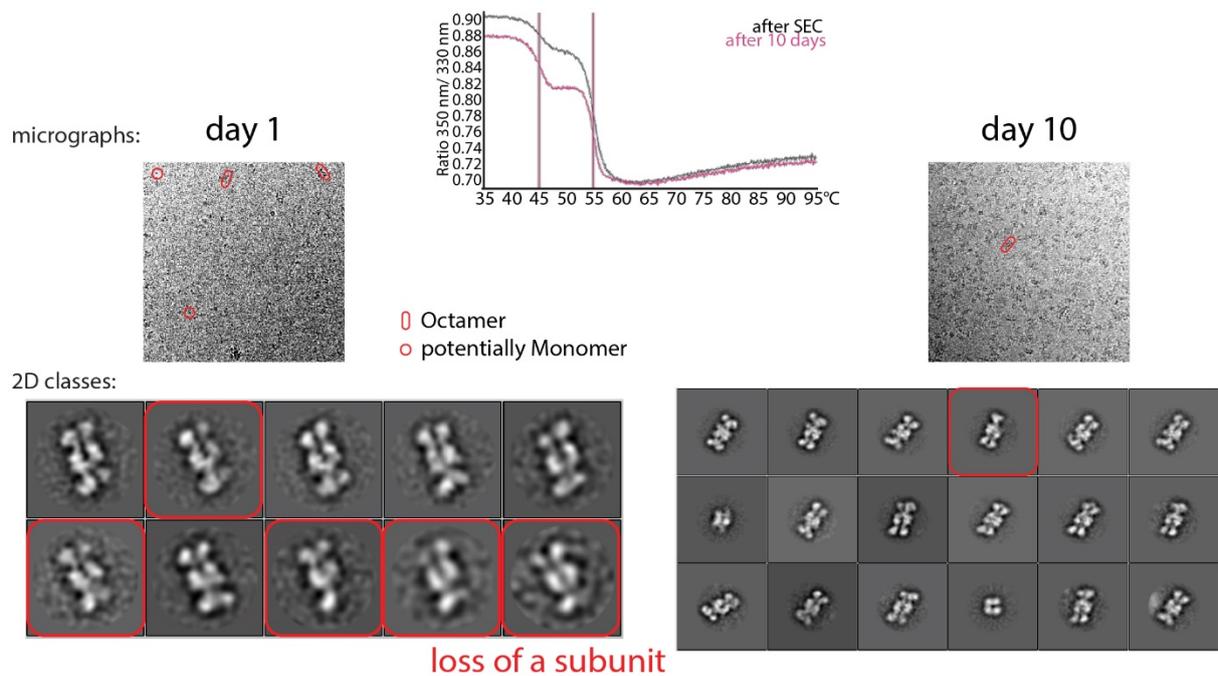
During data acquisition, 928 micrographs were recorded. Around 320,000 particle candidates were picked initially using Gautomatch. After 2 rounds of reference free 2D classification in relion 2.1 54,000 particles remained in the dataset, which were furthermore reduced to 13,000 particles after selecting the particles of the best resolved 3D class For initial 3D classification, the crystal structure octamer of tmaDisA was filtered at 60 Å and used as initial model. To remove model bias, 3D classification was repeated using the best class of the first round as model and mask. The final round of refinement yielded a resolution of 8.6 Å (Fig. 27). The map was further sharpened with a B-factor of -650 Å^2 using the post-processing routine in Relion. An overview of the processing procedure is given in Fig 29.
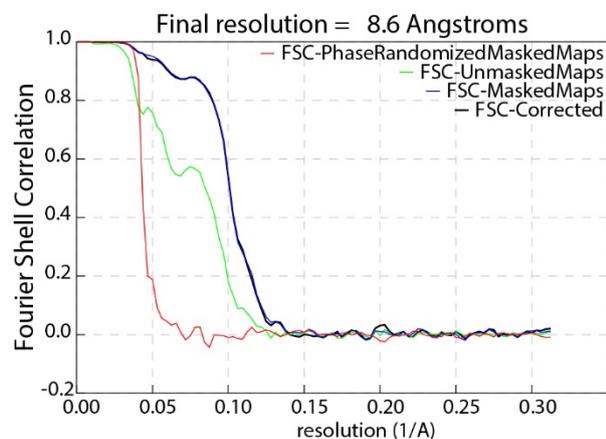

Fig.27: FSC blot of mtubDisA dataset

The structure resembles a sheared barrel. In the middle of the barrel along the long axis a cleft is visible both in top and side view (Fig. 28). This cleft is already present in all 3D classes showing a barrel shaped DisA-like structure, and can be even recognized in the 2D classification. This allows excluding selective processing as a reason for the cleft in the 3D structure. All eight monomers are clearly present, although the visibility of the connections between the outer HhH domain and the inward facing DAC domain of two monomers is highly dependent of the chosen threshold. Local resolution was calculated using localres from the relion2.1 software package. This shows a clear decrease in resolution from 7.6 Å in the core of the complex extending to 10.8 Å in the outermost regions (Fig. 28 right). This indicates a higher amount of structural flexibility. Additional evidence comes from the rather high B factor of -650 Å^2. Typically, B factors are in the range of 450 for this range of resolution. However, the most striking feature is the loss of symmetry in comparison to the highly C4 symmetrical tmaDisA crystal structure. This already becomes apparent even in the 2D classification.
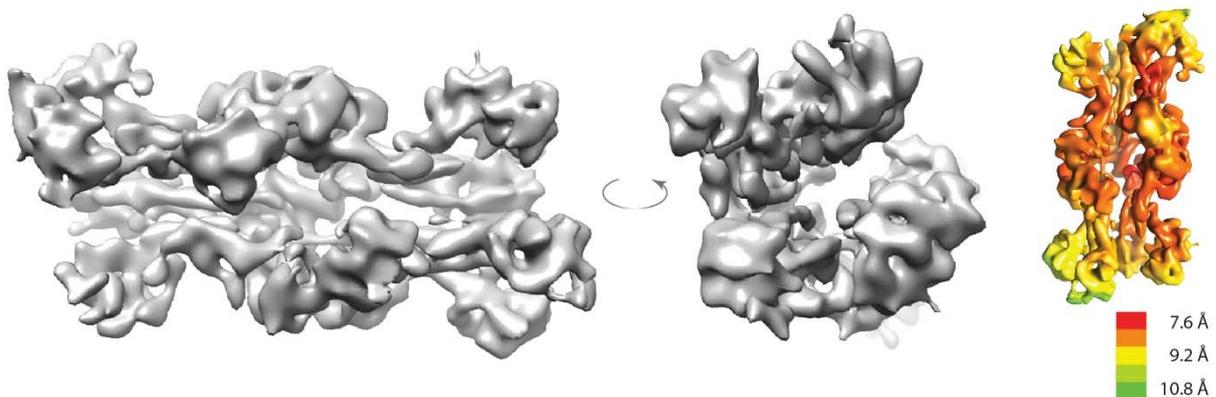

7.6 Å
9.2 Å
10.8 Å

Fig. 28: left: EM structure of mtubDisA in side and top view; right: Local resolution of the
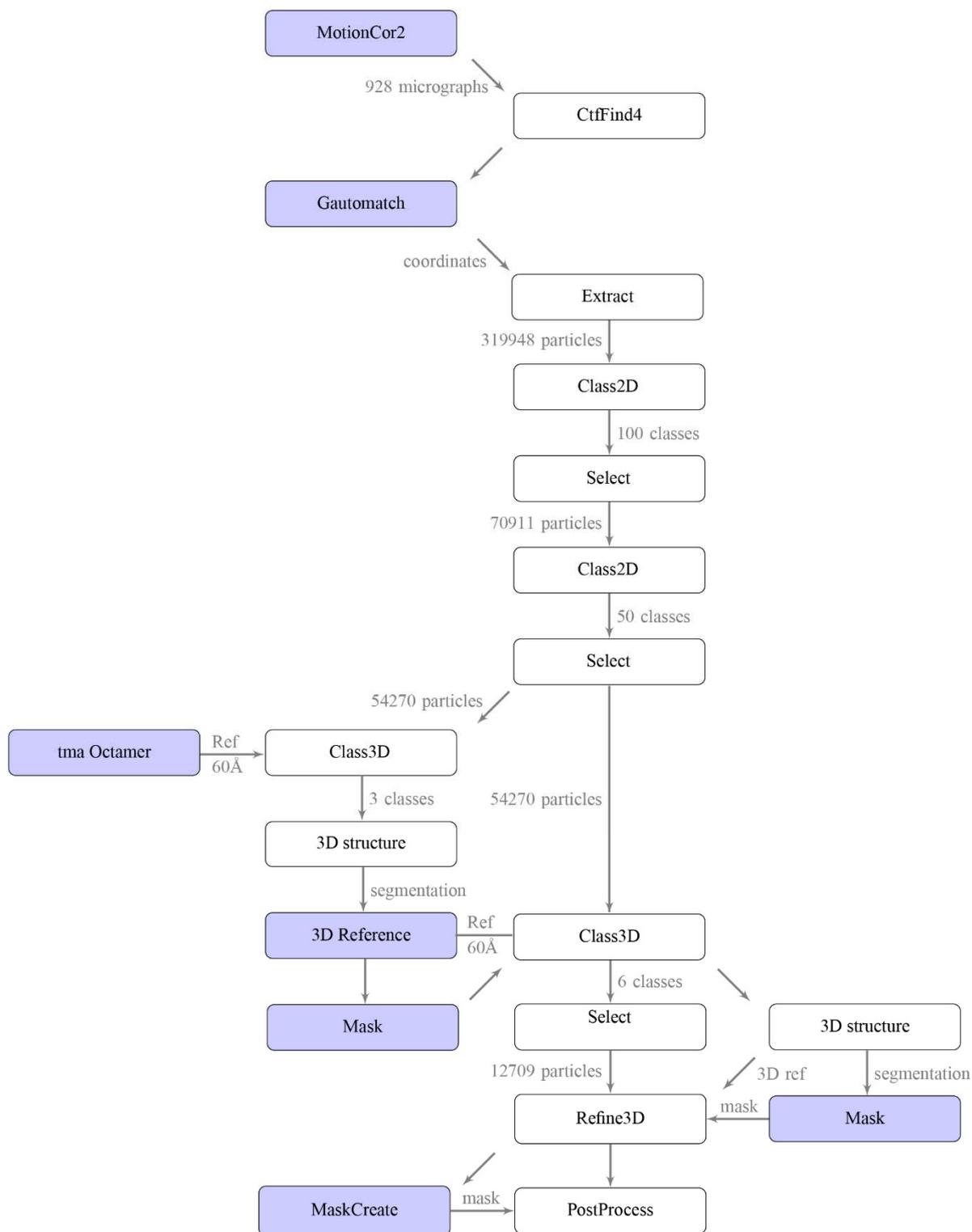
Fig. 29: Processing scheme of the mtubDisA EM dataset

1.3.3. Model of GdpP based on a near-atomic cryo-EM structure

For the phosphodiesterase GdpP, crystal structures of the individual domains, DHH-, GGDEF- and Pas-domain, are solved so far from orthologous proteins or homologs. Crystallization attempts of the full-length protein were unsuccessful. Therefore, the overall organization of the protein remains elusive, including its oligomeric state. In our lab, structural and functional studies of a soluble GdpP construct lacking the transmembrane domain are performed by Adrian Bandera. Size exclusion chromatography coupled right-angle light-scattering of GdpPΔTM, lacking the first 62 residues, yielded a molecular weight of around 270 kDa. This made this project applicable for cryo-electron microscopy. These structural studies were performed together with Adrian Bandera. Screening of initial grids of the wild type GdpPΔTM vitrified in its size exclusion chromatography buffer revealed a disassembly of the protein during grid preparation. Therefore, the sample was stabilized using BS3 as a cross-linking agent. Screening of this sample showed again a disassembly of the complex (Fig.30 left).
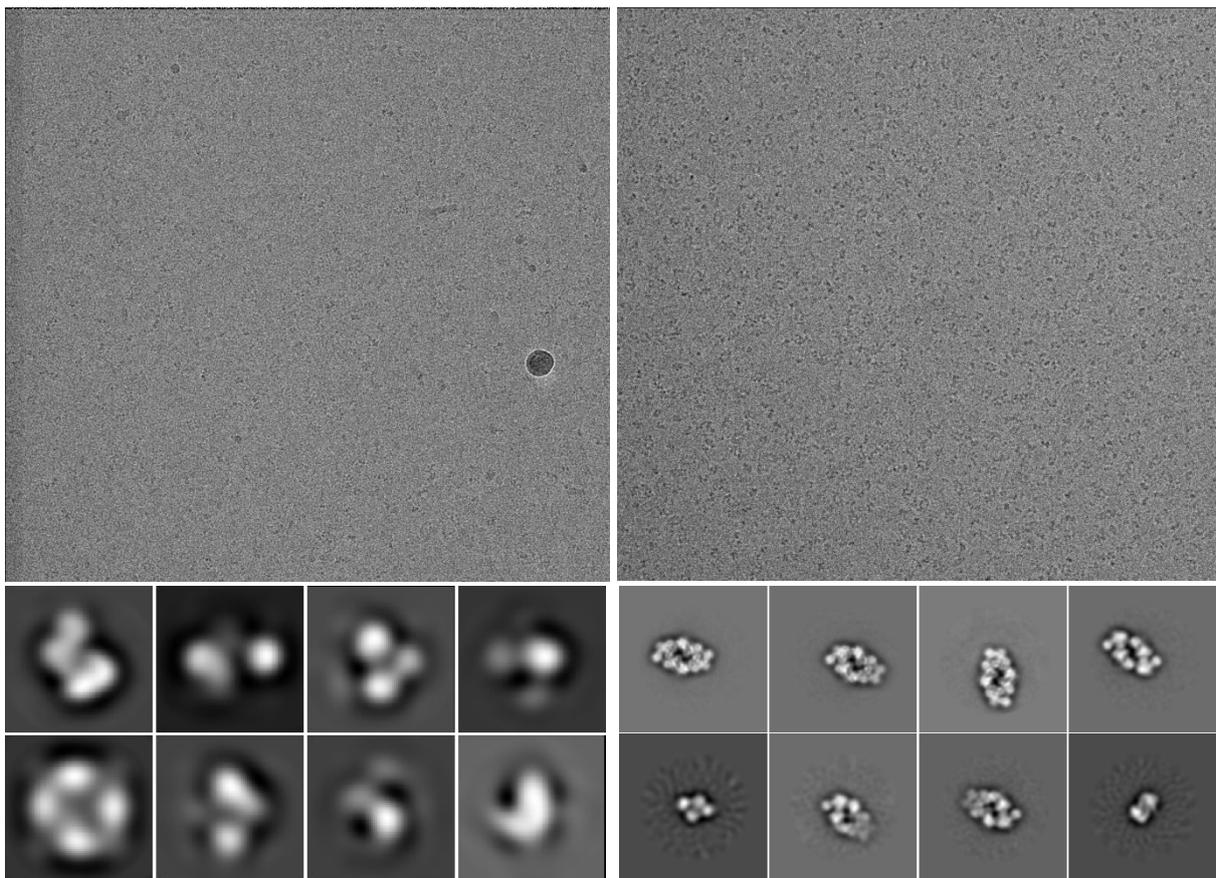
Fig.30: Selected micrograph and 2D classes of cross-linked wtGdpP (left) and of cross-linked inactive mutant of GdpP (right)

As GdpP has enzymatic function as a phosphodiesterase, the stability of the protein complex might be dependent on the nucleotide state. Therefore, an inactive mutant bearing two mutations in the catalytic domain (D418N and D498N) was considered for structural studies. The mutant was purified and cross-linked in the presence of c-di-AMP and thereafter vitrified. Initial screening showed a stable complex on the micrograph at around the expected size (Fig. 30 right). A small dataset containing 269 micrographs was recorded for initial processing (overview of processing shown in Fig. 31). Around 190 thousand particle candidates were picked using reference free picking in Gautomatch. After 2D classification an initial model was generated, the obtained density segmented for noise removal and then used as reference for

Fig.31: Processing scheme of the cross-linked inactive mutant of GdpP
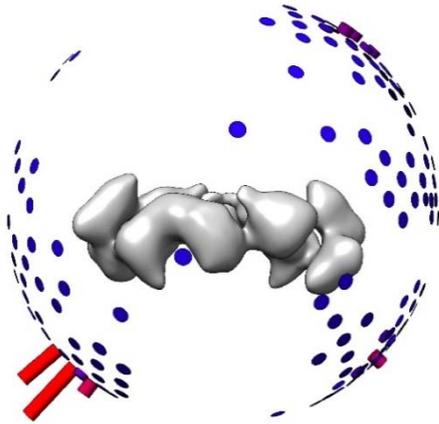
Fig.32: Angular distribution of cross-linked, inactive mutant of GdpP

3D classification. The best classed obtained in the classification were selected and this process iteratively repeated several times. As the 3D classes showed poor angular distribution, particle picking was repeated using the 2D classes as reference. The best three classes of the first round of processing were used as a 3D model for 3D classification. After further classification, autorefinement was run but the obtained density still showed poor angular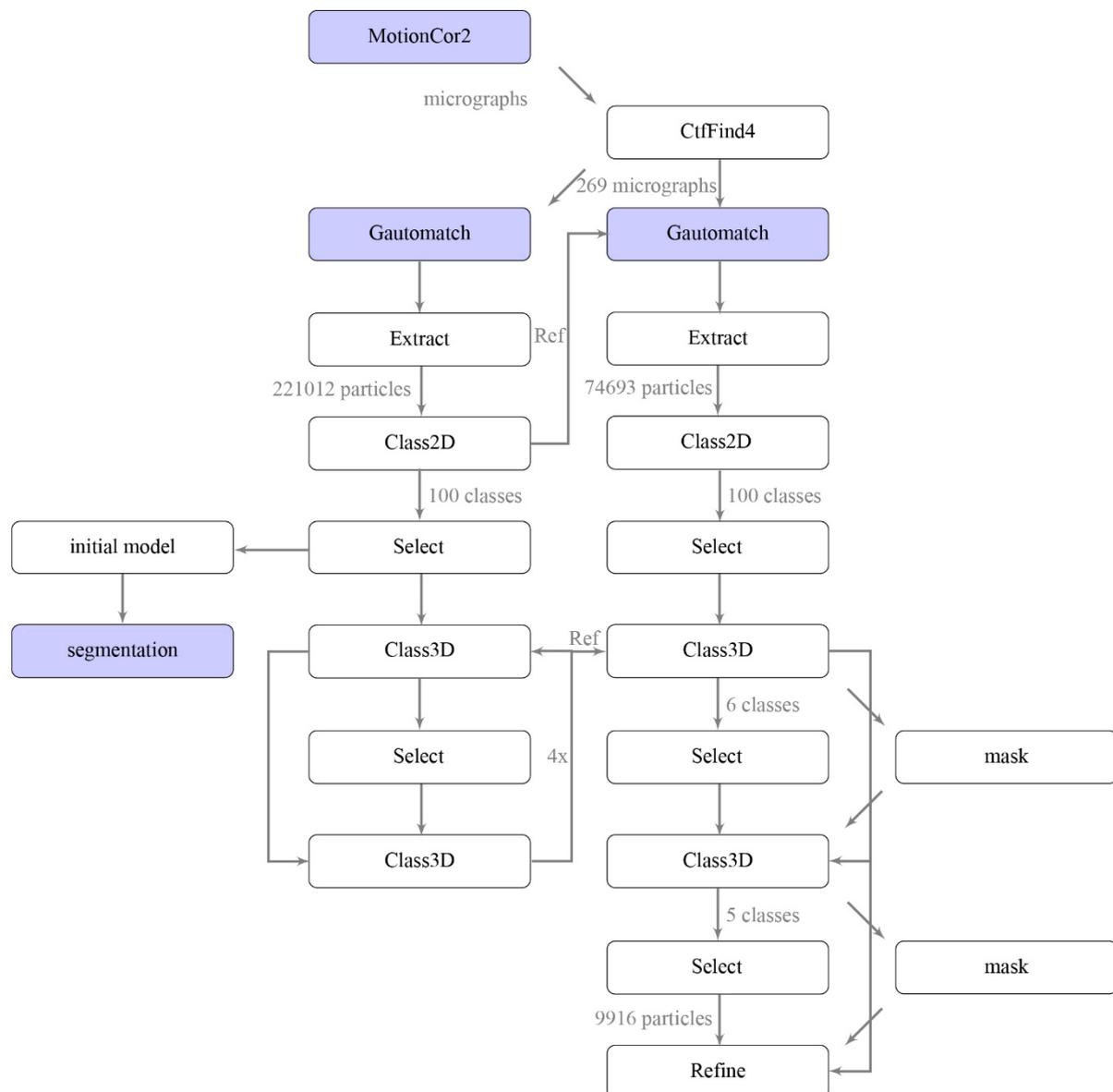 distribution indicating a severe preferred orientation of the sample on the grid (Fig.32). To overcome the obstacle of preferred orientation of the sample on the EM grid, the effect of detergent addition on the orientation bias was tested. To this end, the sample was mixed with 0.05% NP40 before vitrification. As initial processing of a small dataset showed decent angular distribution, a larger dataset was recorded. Therefore, 3968 micrographs were recorded on Titan. A subset of the first 300 micrographs was selected for reference-free particle picking using Gautomatch. 88,000 particle candidates were obtained and subjected to 2D classification. The best classes showing different orientations were selected for template-based picking in Gautomatch. After a further round of 2D classification and selection the particles of the initial 300 micrographs were used for generating an initial model and a mask hereof. To check for orientational distribution 3D classification was performed on this subset, showing a proper angular representation. (Fig.33)
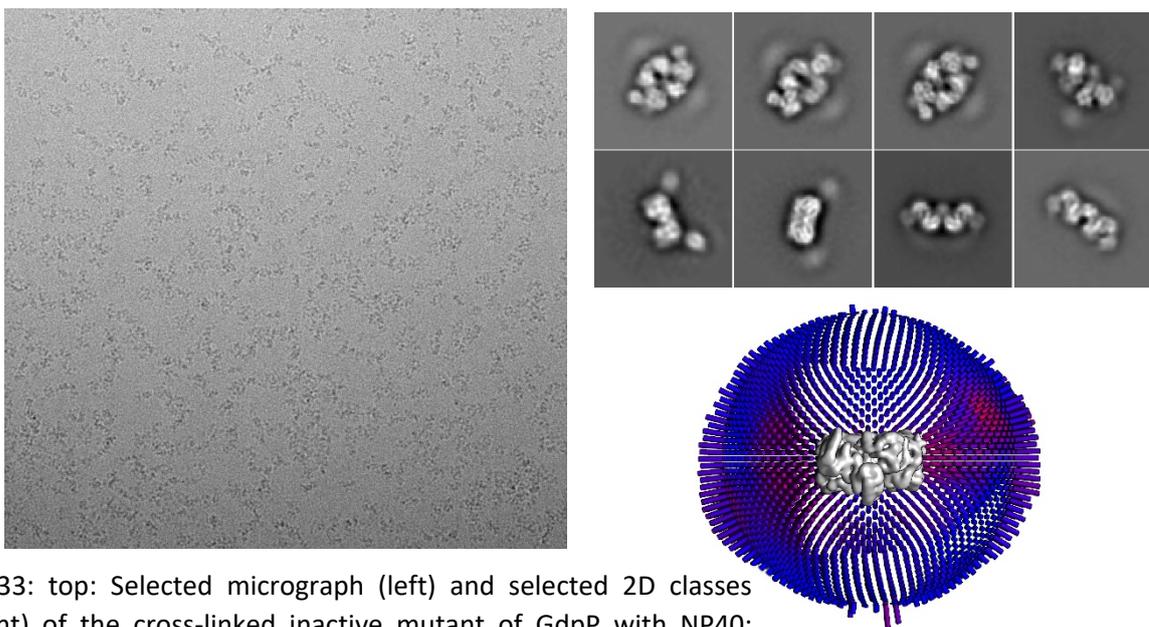
Fig.33: top: Selected micrograph (left) and selected 2D classes (right) of the cross-linked inactive mutant of GdpP with NP40; bottom right: Angular distribution of the respective dataset

50
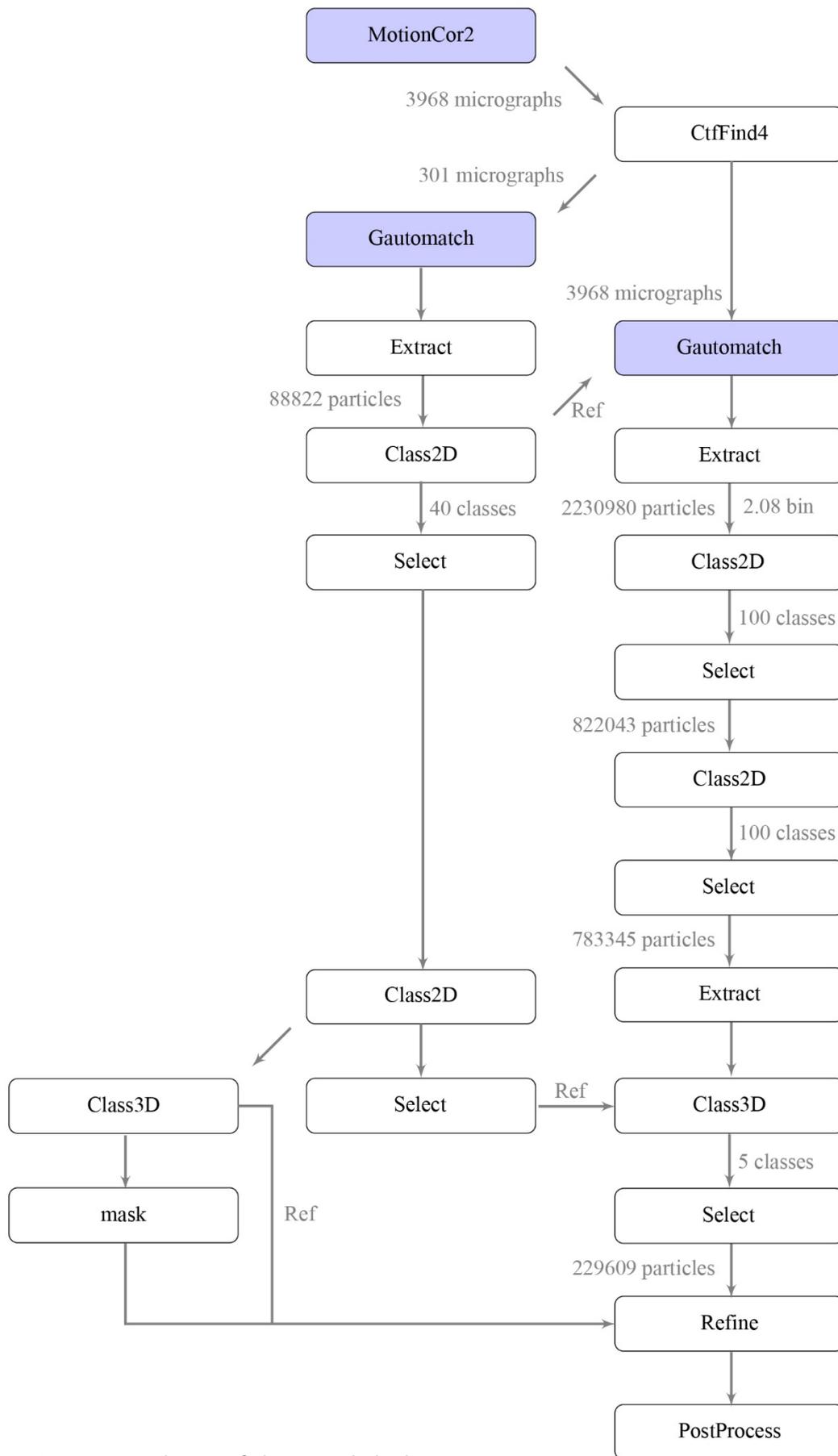
Fig.34: Processing scheme of the cross-linked
inactive mutant of GdpP with NP40

Template based picking of the full dataset yielded 2.2 million particle candidates, which were extracted with a binningfactor of 2.08. 2D classification into 100 classes and subsequent selection of the best classes was performed in two rounds, reducing the particle number to 0.8 million particles. These were reextracted as unbinned particles before subjection to 3D classification using the initial model as reference and masked template. The best resolved out of the five 3D classes were selected and used for Refinement and post-processing using the first 3D classification as reference. The obtained dataset shows a resolution of 4.7 Å with a B-factor of -352 Å^2. The entire dataset was processed using C2 symmetry, and processing using C1 symmetry showed no improvement or major changes in the obtained map. An overview of the processing scheme is given in Fig. 34.

Crystal structures of the individual domains are available from homologues, and therefore, a complete model could be build. Phyre2 [95] was used to generate a model of the DHH domain from GdP from *staphylococcus aureus* (pdb:5xsp) and swissmodel [92, 96-99] was used for the GGDEF domain building a model based on RbdA from pseudomonas aeruginosa (PDB 4zmm). Both models were fitted individually as rigid bodies into the density of both monomers
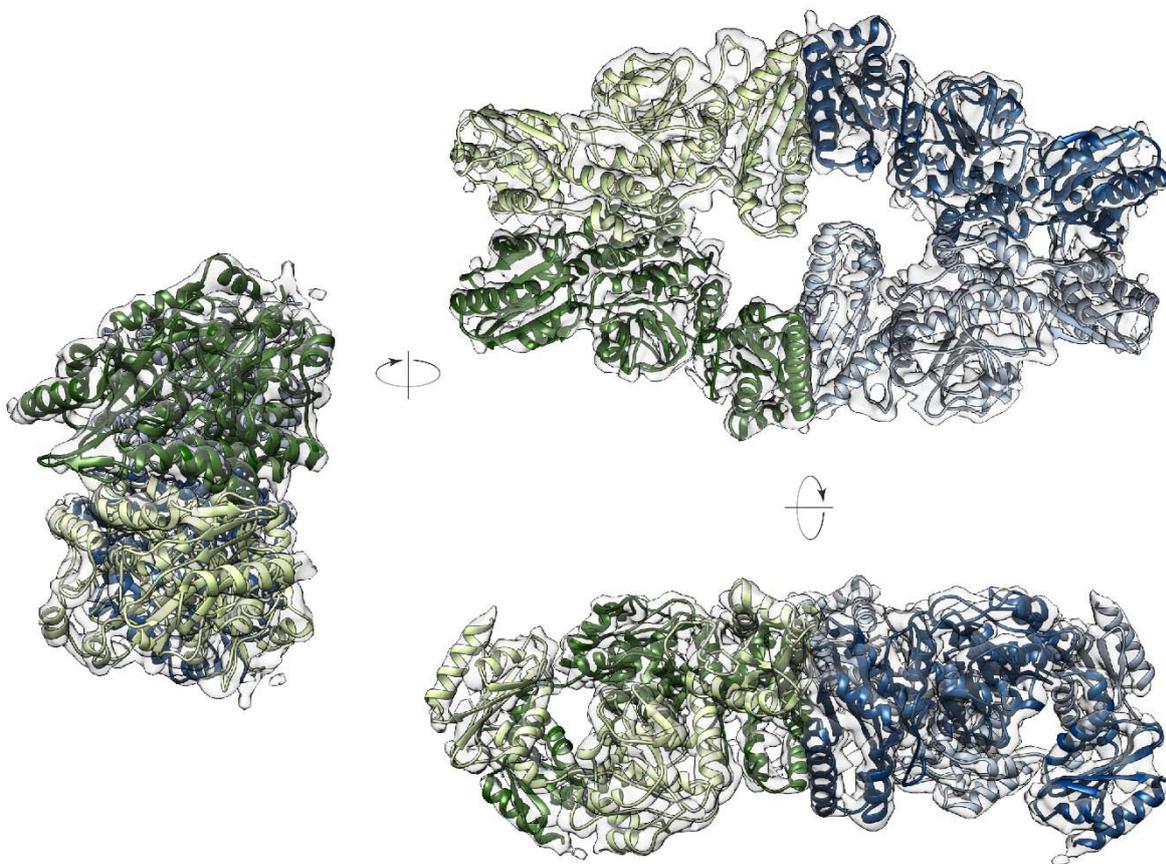
Fig.35: Structure of GdpP in top, side and front view

in the asymmetric unit using UCSF Chimera. Subsequently, they were combined and duplicated into the other half of the tetramer. The obtained model was real space refined into the EM-density map using in the Phenix software package. The obtained model was manually refined and rebuilt using Coot. A complete structural of GdpP was obtained ranging from amino acid 170 in the GGDEF domain to amino acid 650 at the C-terminus of the DHH domain, spanning the entire enzymatic activity carrying domains. The obtained structure is shown in Fig. 35.

The EM-density revealed an elongated, ring-like tetramer with dimerization via the GGDEF domain along the elongated axis and a DHH-dimerization at the pointed end. The structural rigidity decreases towards the top of the DHH domain as seen by loop regions possessing less defined density. The DHH domain is present in the open, nucleotide-free state, and also for GGDEF no nucleotide can be found. For the PAS domain, no density can be assigned indicating high structural flexibility.

1.4. Discussion

1.4.1. Crystallization of tmaDisA3xR mutant

Crystal structure of tmaDisA3xR mutant

Previous work in the lab showed an increased efficiency in synthesising c-di-AMP in a DisA mutant harbouring three mutations in the arginine stretch 128-130 to alanine (DisA3xR) [79]. This amino acid stretch resides in the DAC domain of the enzyme and is in the active center involved in coordinating the γ-Phosphate of ATP. A mutation to alanine would therefore be considered to reduce the catalytic efficiency, rather than increasing it. The current hypothesis is a widening of the exit channel as reason for the enhanced productivity of the DisA3xR mutant. To test this hypothesis a crystal structure of the tmaDisA3xR mutant was solved. The structure at 2.2 Å showed no major differences to the wild type structure, except for the three mutated arginine residues. The overall barrel-shaped structure remained intact, as well as the DAC-domain fold in the catalytically active, central domain. The presence of c-di-AMP in the crystal structure proofs that despite the mutations the enzyme has not lost its catalytic activity.

Comparison of tmaDisA3xR mutant with wild type DisA

A superposition of the structures from tmaDisA3xR and the wild type revealed no major changes, as can be seen in the RMSD of 0.21. This small value indicates almost identical structures. Previous work in the lab showed the tmaDisA3xR mutant to possess slightly increased activity. To test if this enhanced enzymatic activity might be caused by an increase in the exit tunnel, the MOLE
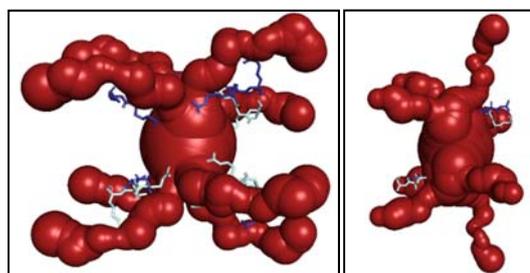
Fig.36: Channel formation in DisA with channels shown in red, arginine residues in light and dark blue; top: Wild type DisA, bottom: tmaDisA3xR mutant

2.5 online toolkit was applied to find and compare the channel formation in both structures (Fig.36). In this analysis, a clear difference becomes apparent. An increase of the channel diameter upon mutation of the arginine residues strengthens the hypothesis of the elongated residues to function as a molecular curtain decelerating the diffusion rate of c-di-AMP and hence slowing down the catalytic rate. These results constitute now the first experimental evidence for the former hypothesis, that the clearance of the active site constitutes the rate-limiting step for c-di-AMP synthesis by DisA.

1.4.2. EM structures of DisA from different species

<u>Cryo-EM structure of tmaDisA</u>

DisA is a protein with enzymatic activity which needs to be regulated by external signals, like e.g. DNA binding. Previous work in the lab [37] could show a inactivation of DisA upon binding to non-standard DNA, like Holliday junctions. Comparison of the structures with homologous complexes – such as RuvA – however, reveal a clash in the DNA-binding pocket. This led to the conclusion, that either a conformational change upon DNA binding occurs, or that the rigidity of the crystal structure might be stabilized in an artificial conformation due to crystal-packing. Previous crystallization attempts of a DNA-bound complex were unsuccessful, as well as structural analysis using EM in the course of this study remained without success. This might be either due to a disassembly of the complex or by an inability of the DNA-bound complex to attach to the EM-grid during vitrification. To analyse if the HhH-domains are really in an inactive position in vitro or if the crystal packing has influenced their orientation cryo-EM studies of the complex in the DNA free state were performed. TmaDisA attached to the grids

during the vitrification process and with a dataset of around 8,500 micrographs a structure could be calculated at a resolution of 3.0 Å showing C4 symmetry. The structure could be build using coot, with the original crystal structure as starting model. However, the variations between the crystal structure and the EM structure are only minor and in the range of side chain variations. The HhH-domain at the outer most region of DisA was nicely resolved, confirming the conformation obtained in the crystal structure (Fig.37).
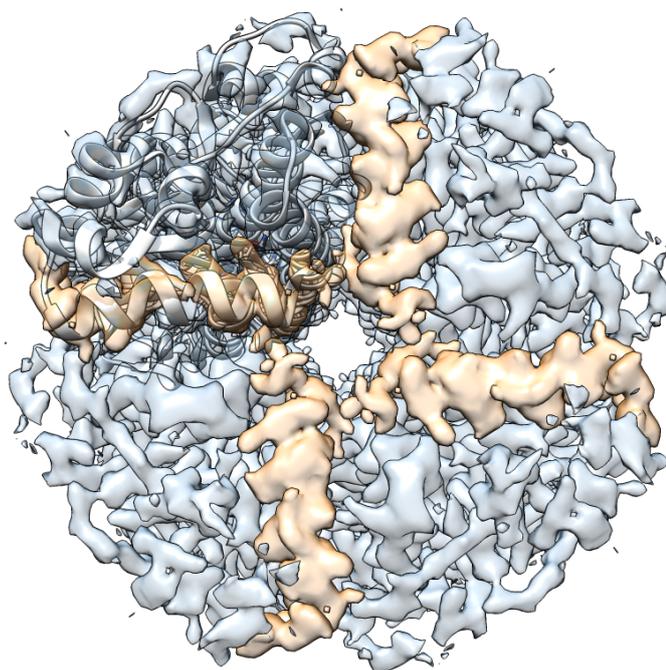

Fig.37: Top view of segmented EM density of tmaDisA with fitted crystal structure of the tmaDisA Dimer (white), the orange density represents the long helix of the HhH domain

Purification and cryo-EM structure of bsuDisA

As EM studies of tmaDisA could not answer the question of the regulation of DisA, DisA homologues from different species were additionally considered for structural analysis. In previous studies, DisA from *bacillus subtilis* was used for crystallization attempts, and, though no crystals were obtained, the purification yielded decent amounts of pure protein. As one reason for seeing no structural variability in tmaDisA studies might be its analysis at comparably unphysiological conditions for a thermophilic species, unsuccessful crystallization attempts of another species might indicate higher structural flexibility and hence increase the chance of capturing a different confirmation applying a different, in solution structural method, such as cryo-EM. Therefore, the purification was repeated based on the previously established protocol in the lab. Vitrification of the sample and application to the grid was achieved and with a dataset of around 1,300 micrographs a structure of 8 Å was obtained. The overall architecture of the obtained structure resembles tmaDisA, as bsuDisA also possesses

a barrel shaped structure. However, a loss of the C4 symmetry can be detected. This is mainly due to the increased flexibility of one monomer. The density of the last monomer is only visible at high threshold. This, however, indicates not a complete absence of the monomer, but an increased flexibility or partial occupancy. Furthermore, a fit of the tmaDisA asymmetric unit, represented in green, and of the HhH-domain, represented in orange, into the density indicates the same conformation of the HhH-domain, as far as this assumption is possible at the obtained resolution (Fig.38).
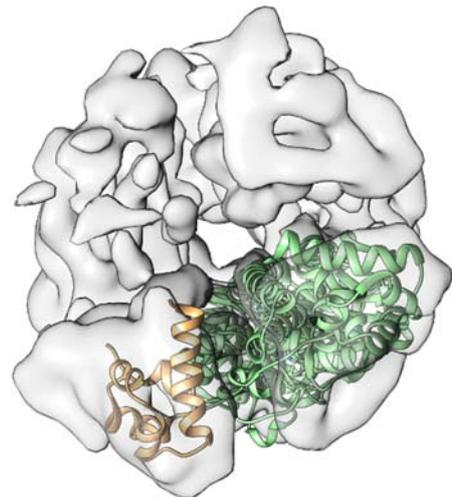
Fig.38: Top view of EM density of bsuDisA with fitted crystal structure of the tmaDisA Dimer (green); the HhH-domain is shown in yellow

Purification and cryo-EM structure of mtubDisA

To allow a better comparison between tmaDisA and DisA from non-thermophilic organisms, *Mycobacterium tuberculosis* was chosen for further structural analysis of DisA as second sample. This should allow seeing if the loss of C4 symmetry obtained in bsuDisA was specific to this species or a general feature of non-thermophilic DisA structures. For mtubDisA, already the purification showed to be less homogeneous, than for the other DisA species. The size exclusion profile showed the formation of a shoulder, indicating structural inhomogeneity.

Differential scanning fluorimetry revealed an enhanced folding after longer incubation periods at 8°C. This could also be seen in initial EM studies, as in the recorded micrographs of the sample, which was incubated for 10 days in the fridge, more intact particles were visible. This was furthermore shown in the 2D classification, were significantly less 2D classes showed a complex with one subunit lost (shown in Fig.26 in the results section). However, the loss of the C4 symmetry in comparison to tmaDisA becomes apparent already in the 2D classes. The classes show particles which seem twisted and a cleft formation on one side becomes visible (Fig.39). Although the obtained 3D structure
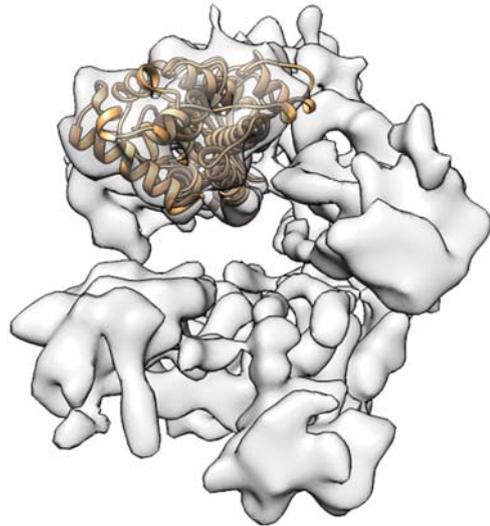
Fig.39: Top view of EM density of mtubDisA with green fitted crystal structure of the tmaDisA Dimer and in yellow fitted HhH-domain

at 8.6 Å shows a tmaDisA-like organization, it confirms the impression obtained from the 2D classes of a sheared barrel and shows a prominent cleft formation along the long axis, which is even visible in the top view. Although, the visibility of the helices connecting the DAC and the HhH domain is dependent on the threshold, all subunits are equally visible and present, unlike in bsuDisA. As in the bsuDisA-structure, the mtubDisA structure shows the same conformation for a superimposed HhH-domain as in the tmaDisA crystal structure. However, slight differences between the HhH-domain densities of the individual monomers are present. This might indicate different flexibilities of the dimers along the cleft. In Fig.39, the first monomer to the right of the cleft shows most density with a decrease of density in clockwise direction. This might indicate an opening mechanism of the cleft and represent an intermediate state between the tmaDisA structure and the bsuDisA structure.

Analysis of stability-variations in DisA-complexes among different species

All three structures of DisA from *Thermotoga maritima*, *Bacillus subtilis* and *Mycobycterium tuberculosis* possess the same overall organisation of DisA and its composition as barrel-shaped octamer. In addition, the orientation of the domains of DisA remains similar in different species. In all three species, the DAC domains oligomerize in the centre of the molecule with the HhH-domain pointing to the outside, while both are separated by three helices in a bundle.

One aim of this work was to show a conformational change in the outer HhH-domain, to allow DNA binding as seen in homologous structures. Unlike this expectation, all three structures proof the conformation of the HhH-domain in the crystal structure to be the in solution state of DisA and neither of them shows any hint of a rotation of the helix, which would be necessary to allow DNA binding. However, this does not exclude a change of conformation upon DNA binding.
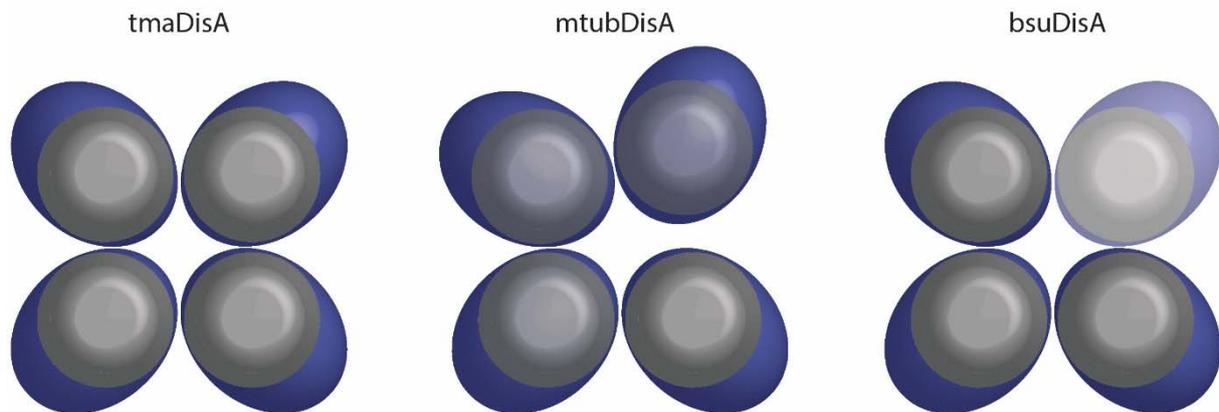
Fig.40: Schematic representation of the occupancy and flexibility of the Dimers in the three different structures of DisA from *Thermotoga maritima* (left), *Mycobacterium tuberculosis* (middle) and *Bacillus subtilis* (right)

In a direct comparison of the three structures, an increase in flexibility in the non-thermophilic organisms becomes apparent. In comparison to tmaDisA, mtubDisA shows the cleft formation and a slight decrease in occupancy of the HhH-domain, whereas in the bsuDisA structure one entire monomer shows decreased occupancy whilst the structural integrity of the barrel as such remains intact (Fig.40). This might indicate a disassembly mechanism as regulation for DisA. In this case the mtubDisA structure would show an intermediate state, where the opening of the barrel-ring and the increase of flexibility of the outer domain would facilitate the loss of one first subunit, as can then be seen in the bsuDisA structure (Fig.41).
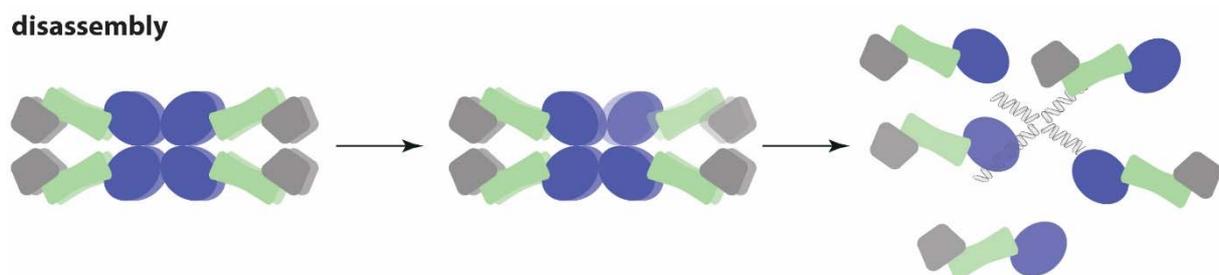
disassembly

Fig.41: Schematic representation of a possible disassembly mechanism based on the insight gained from the three EM structures

However, the role of a so far unknown effector protein cannot be excluded. In such a case, the binding of an effector protein could stabilize the octameric barrel and hence reduce the accessibility of the active site (Fig.42). However, such a regulation pathway would be aggravated by the loss of one subunit as seen by bsuDisA, as first a re-association of the entire complex would be necessary. Therefore, a disassembly mechanism seems more likely.
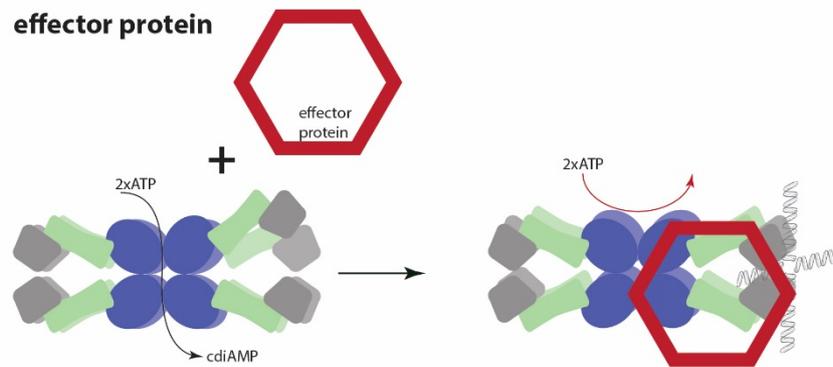
Fig.42: Schematic representation of a possible inactivation of DisA through the interaction with an effector protein, as suggested by the data from the three EM structures

The third suggested mechanism, so far, was that the Holliday junction leads to a conformational change in the HhH domain, leading by a long-range communication to an inactivation of DisA. This mechanism however becomes unlikely, as an increased flexibility and decreased occupancy rather leads to a destabilization of the complex, than to a conformational change.

1.4.3. Structure of GdpP

Cryo-EM studies of GdpP

For protein purification and solubility reasons, a construct without the two transmembrane helices (GdpPΔTM) was used throughout the entire study. The single-particle analysis of GdpPΔTM encountered various problems. Sample directly vitrified after size-exclusion chromatography showed no particles visible in the micrographs, presumably due to complex disassembly. Upon cross-linking of the sample, shorter fragments were visible; however, no full complex assembly was detectable at the microscope. Only upon cross-linking of an inactive mutant harbouring two point mutations in the DHH domain (D418N, D498N) a stable complex could be detected in the micrographs. However, upon processing of the data, severe preferred orientation of the complex in the ice was visible in the angular distribution shown for the 3D classes. To overcome this obstacle the use of a detergent was tested, and NP40 yielded good results in improving the angular distribution of the particles and allowing further data processing and the calculation of a 4.7 Å structure (Fig.43). Despite extensive processing trials, the overall resolution could not be improved. One reason for this might be enhanced flexibility of the sample, which would explain the necessity of cross-linking. Due to flexibility of the side chains the overall occupancy amino acid residues decreases resulting in a loss of resolution in
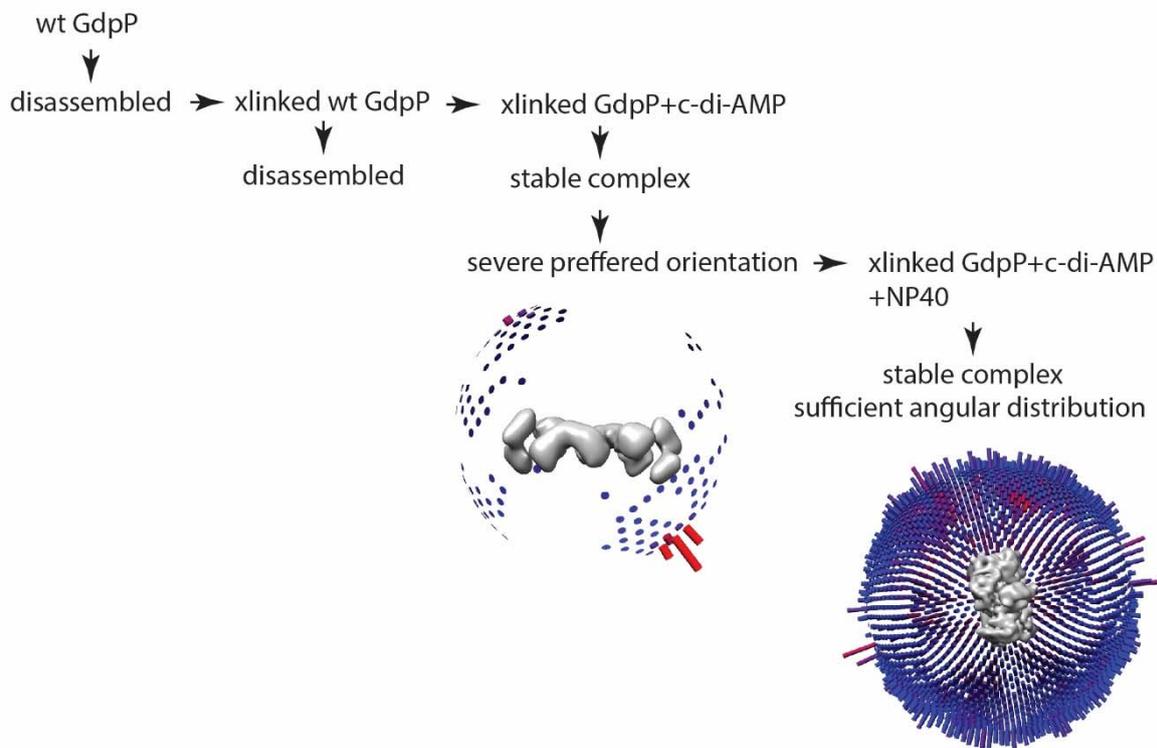
Fig.43: Optimization workflow for GdpP

the electron density. Another cause for limiting the resolution of the obtained density is a clustering of the particles in thick ice. This dampens the signal and decreases the signal to noise ratio. Therefore, information of the particle electron density is lost and the resolution decreases.

Biological impact of the structure

The EM structure presented in this work represents GdpP with an inactive mutant in the DHH domain and the GGDEF domain. It is the first structure of a c-di-AMP-specific GdpP-type phosphodiesterase in the context of its neighbouring domains. GdpPΔ clearly shows a tetrameric state for oligomerization state in solution, as indicated by size-exclusion chromatography. Tetramerization occurs via the dimerization of the DHH domains, as well as of a dimerization of the GGDEF domains, therefore creating a rather flat and elongated ring-like structure. The N-termini of the GGDEF domains are outward facing, leaving space for the PAS domain, which is lacking in the structure due to high flexibility. (Fig44)
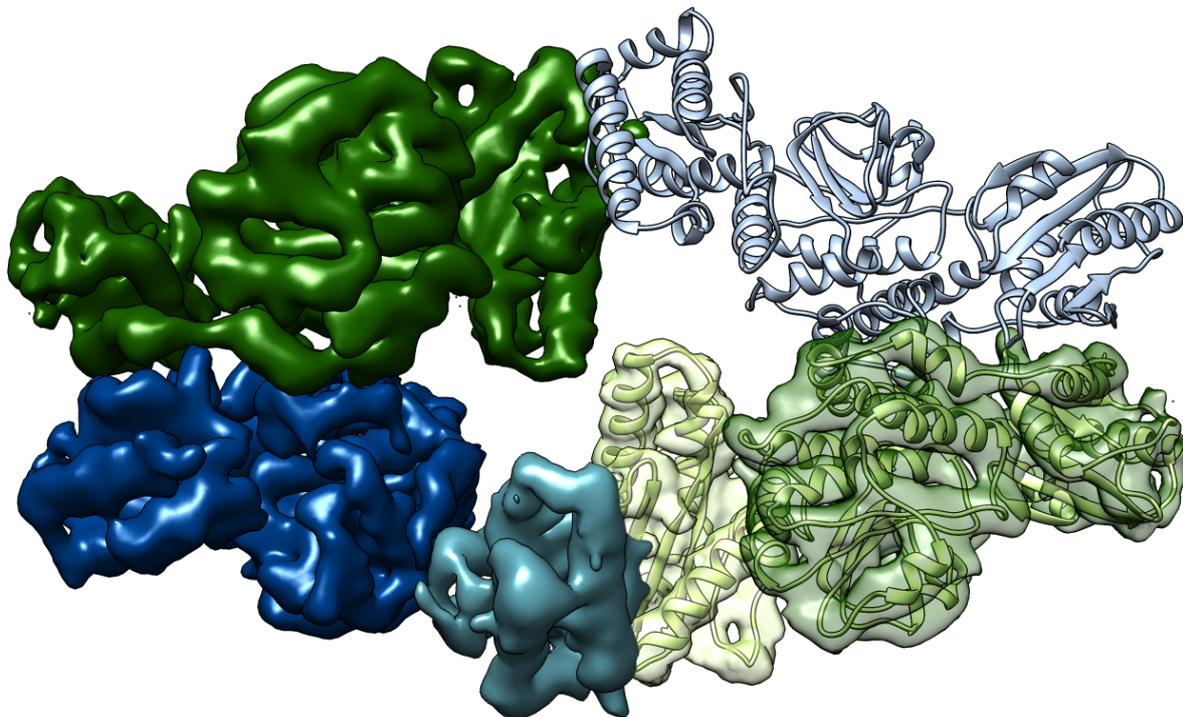
Fig.44: Tetramer of GdpP based on the EM structure; left: EM density, in green depicted one monomer, dark blue shows the DHH domain and light blue the GGDEF domain; right: light blue and light green show the structural model of GdpP modelled according to the EM density, the surface model shows the EM density of DHH and GGDEF domain in dark and light green, respectively

The PAS domain, although present in the construct, is not resolved in the structure, due to its high flexibility. In some of the 2D classes, density for the PAS domain is visible, as well as in the final EM-density at very high threshold (Fig 45 left). This high flexibility might be accounted

for by the absence of the transmembrane part of GdpP. Furthermore, a high flexibility of the PAS domain might be required to allow regulation of the activity of GdpP by conformational changes triggered through structural changes in the PAS domain. This might be a mechanism for regulation of GdpP through extracellular signals, such as proposed for RbdA in [100]. Through the direct link of the PAS domain to the GGDEF domain, a forwarding of a signal to the GGDEF domain and hence triggering conformational changes in the DHH domain might be a possible mechanism for regulation of GdpP.

The superposition of the GGDEF domains of GdpP and PA0861 (Fig. 45 right) reveals a huge difference between the location of the respective PAS domains. PA0861 possesses a PAS-GGDEF-EAL domain architecture with the EAL domain hydrolysing c-di-GMP under the regulation of PAS- and GGDEF domain. In the PA0861 structure, the PAS domain would clash with the position of the second monomer. However, the connection between the GGDEF domain and the PAS domain is only composed of one helix, which might harbour a higher flexibility, which was stabilized in the crystal structure through crystal contacts. Therefore, a transition between these two states, either during regulation or during crystallization is possible. Furthermore, the orientation of the EAL domain in regard to the GGDEF domain varies between the two structures. In the GGDEF-domain superposition of the two structures, the EAL domain protrudes from the GdpP ring and is in closer proximity to the GGDEF and hence forms a larger interface.
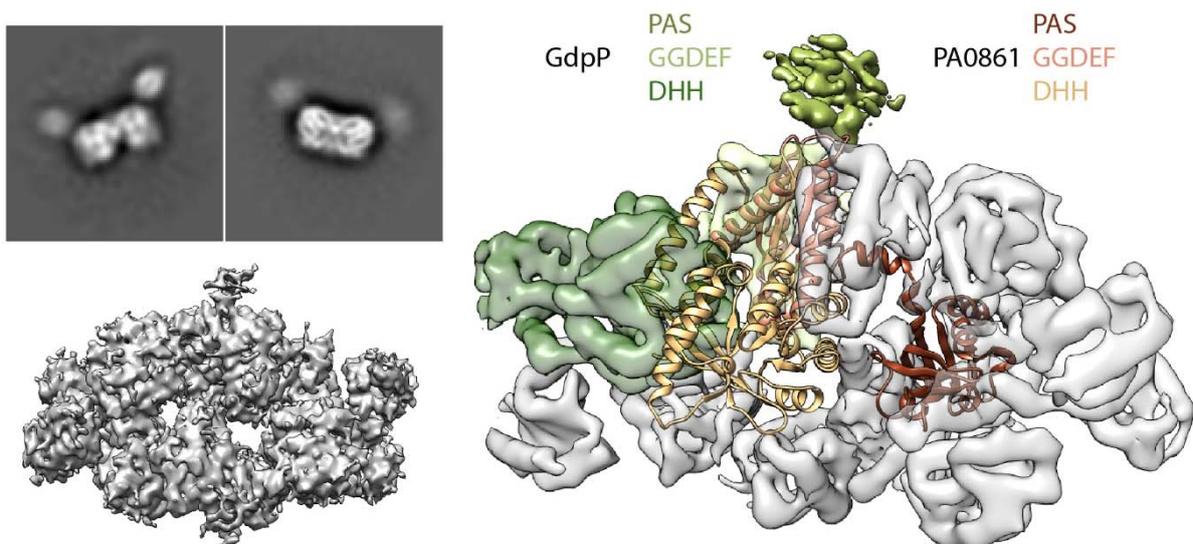
Fig.45: left: Selected 2D classes and EM structure of GdpP at higher threshold, showing density for the PAS domains; right: EM density of GdpP in superposition with the structure of PA0861 fitted on the GGDEF domains, revealing a distinction in the location of the PAS domain between GdpP and PA0861, shown in intense green and red respectively

The obtained density and model clearly show the DHH domain to be in the inactive, nucleotide free state, as the molecule was captured in the open conformation of the two lobes. A comparison to the crystal structure of PDE from *Thermotoga maritima* (pdb: 5o25) in the ligand free state shows a high overall resemblance of the structure (Fig.46). However, a slight increase in the opening of the DHH domain can be seen by an increased angle between the two lobes. Therefore, GdpP was clearly
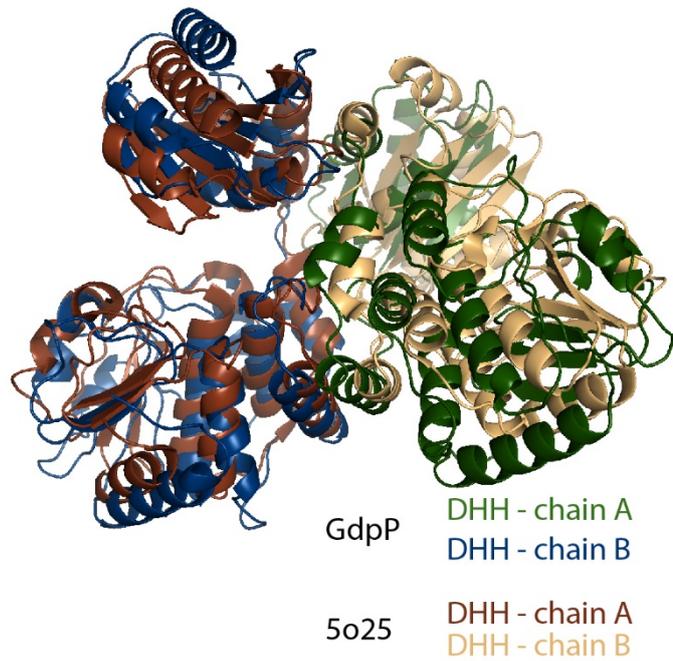
Fig.46: Superposition of the DHH domains of GdpP and PDE1 from *Thermotoga maritima* shown in blue with green, and red with yellow, respectively

cross-linked and vitrified in the inactive state of the DHH domain, despite the presence of c-di-AMP in the crosslinking conditions. This indicates that through other regulatory means the DHH domain is kept in its inactive state in the structure. One possible regulatory mechanism for the activity of the DHH domain of GdpP could be a conformational change of the GGDEF domain upon nucleotide binding. Therefore, a comparison of the dimerization of the GGDEF domain in the EM-structure with crystal structures of GGDEF-dimers in complex with nucleotide are used for

Fig.47: Superposition of the GGDEF domains of GdpP and two other GGDEF domains in nucleotide bound state. The domains are depicted in blue with green for GdpP and yellow and red for *Pseudominas aeruginosa* and *Thermotoga maritima*, respectively

analysis (Fig. 47). For GGDEF-domain dimerization different crystal structures are known. As a comparison here, a GGDEF-domain from *Thermotoga maritima* (pdb: 4urs) is used, depicted in red, as well as a the GGDEF-domain of Dcsbis from *Pseudomonas aeruginosa* (pdb: 4zmm), shown in yellow. Both structures show the dimerization of the c-di-GMP bound state. However, big differences in the dimerization are visible. Both structures adapt a V-shaped form. Even though both structures show the dimerization at the bottom end of the beta-sheets the second monomer faces to different sites. This can be seen in the position of the rotation axis. In the structure from *Thermotoga maritima*, the rotation occurs along the shorter axis. In the GGDEF dimer of *Pseudomonas aeruginosa* Dcsbis, the monomers are flipped around the long axis. However, in both structures the c-di-GMP molecule sits at the pointed end of the V, roughly at the same position. In the EM structure of GdpP the

Fig.48: EM structure of GdpP in superposition with GGDEF domains of *Thermotoga maritima* and *Pseudomonas aeruginosa*, shown in red and yellow respectively. GdpP GGDEF is shown in blue. left: View of full tetramer with superposed GGDEF domains; right: Close up on the GGDEF structures; bottom panel shows a superposition of GdpP monomer with with the second dimerized GGDEF domain in superposition with *Thermotoga maritima* GGDEF dimer

dimerization of the GGDEF domains occurs via the short axis, comparable to the *Thermotoga maritima* structure. However, the two monomers are in closer proximity than in the *Thermotoga maritima* structure. A superposition of the c-di-GMP structures with the structure from the EM density clearly shows the differences in the orientations of the dimer (Fig.48). In both superpositions the EM map clearly shows no density for c-di-AMP, confirming the apo state. In the superposition of the *Psuedomonas aeruginosa* structure, the second monomer lies entirely outside of the density. For *Thermotoga maritima*, a superposition of the crystal structure dimer with the EM density the rough orientation of the second monomer shows certain overlaps with the map, especially from the top view. However, from the side, an increased opening of the V-shaped structure can be seen in the *Thermotoga maritima* crystal structure in comparison to the EM-structure of GdpP. Therefore, a topological change in the quarternary structure is a likely regulatory mechanism for the activity of the DHH domain. The two different dimerization states of the same nucleotide state in *pseudomonas aeruginosa* and *Thermotoga maritima* allow the postulation of two different models for a topological change of the GGDEF domain in GdpP upon binding of c-di-AMP. One possibility would be that the GGDEF dimer adopts the conformation as seen in the *Pseudomonas aeruginosa* structure. Such a large conformational change in the middle domain of a protein complex, however, seems rather unlikely. Furthermore, a larger conformational change might sterically prevent the dimerization of the DHH domain, which is most likely required for its functional activity. Therefore, topological changes, which might affect the oligomerization, are rather unlikely;
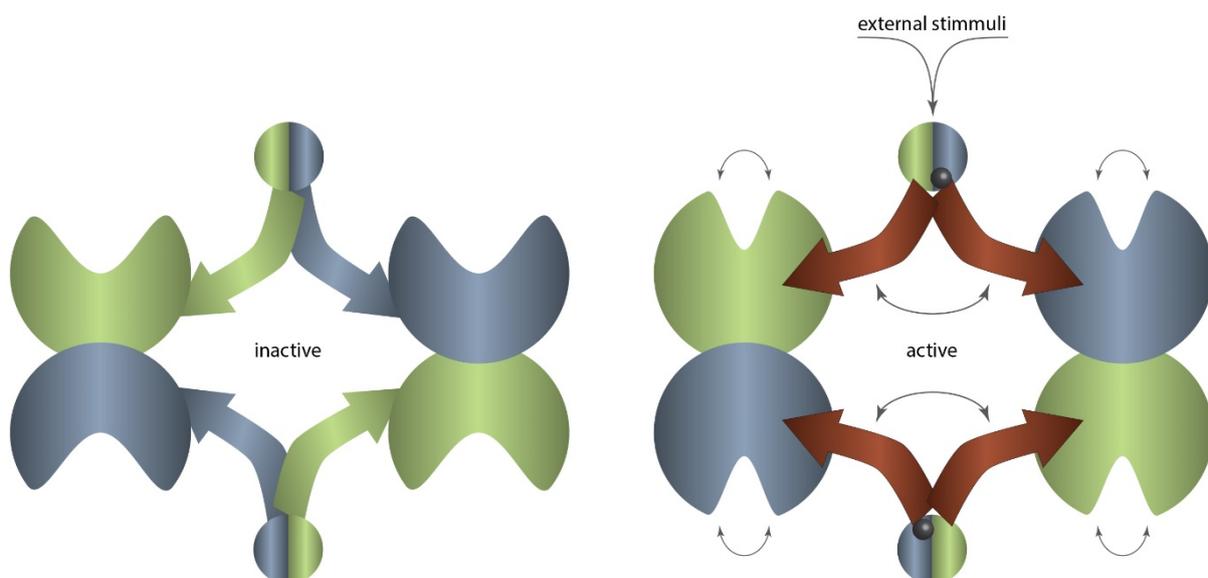
Fig.49: Suggested model for the regulation of GdpP through a conformational change of the GGDEF Dimer

however, they cannot be excluded. A further opening of the GGDEF-dimer, as seen in the *Thermotoga maritima*, would be more likely. This smaller conformational change would be less hindered through the oligomeric state or the dimerization of the PAS-domain. This might lead to a long-range communication with the DHH domain and allow regulation of its activity (Fig.49).

In order to further characterize the regulation of GdpP additional analysis of the oligomeric state dependent on the bound nucleotide is proposed. Therefore, protein biochemical and biophysical analysis, such as analytical ultracentrifugation, static light scattering and native page might be applied. Crystallization attempts of the GGDEF-dimer in the presence of c-di-AMP is suggested, as a structure of the dimer in the nucleotide bound state might allow to draw conclusions on the conformational changes of the entire complex during regulation. EM structures of a second nucleotide state would equally be useful to gain further insights into the regulatory mechanisms of GdpP.

2.

Characterization of the DNA-binding of chromatin remodellers involved in DNA end resection

2.1. Introduction

2.1.1. DNA double strand break

A break in the double strand of the DNA (DSB) highly endangers the genome integrity and stability and therefore the survival and health of the cell and organism is threatened. To restore the genetic information several repair pathways have evolved. Two distinct pathways have to be distinguished: ligation based repair mechanisms, like non-homologous end joining, and recombination based DSB repair by homologous recombination (HR). In HR an undamaged homologous DNA strand, such as a sister chromatid, is used as template, ensuring error-free repair of the genome [101].

2.1.2. Homologous Recombination

The first step in HR is the removal of a short segment of the 5´ strand, revealing a 3´- single stranded DNA overhang. This process, performed by the MRN-Sae2 complex (MRE11Rad50NBS1) and DNA2, EXO1, is known as short and long range DNA end resection, respectively, and is essential for homology search and the invasion of the single strand DNA (ssDNA) into the double strand of the homologous DNA fragment. The formation of protein filaments around the ssDNA is used to protect the otherwise vulnerable ssDNA fragment and remove potential secondary structures. These filaments are first formed by the heterotrimeric replication protein A (RPA) before they are exchanged for Rad51 by recombination mediators Rad52 and Rad55-57 in yeast or BRCA2 and Rad51 paralogs in human. Rad51 forms a nucleoprotein filament on the ssDNA, called the presynaptic filament, by binding in a helical fashion. The presynaptic filament aligns with its homologous DNA sequence on the sisterchromatid, leading to the D-loop formation. After D-loop expansion the second 3´ssDNA overhang on the opposite site of the DSB is captured forming a double Holliday Junction (dHJ). Resolvases help to form either cross-over or non-crossover products in cleaving the dHJ (Fig.50) [102].

During the process of DNA resection, the obstacle of DNA packing into chromatin needs to be overcome. In this process, chromatin remodellers play the essential role in ejecting, sliding or exchanging nucleosomes or histone variants. Additionally nucleosome-associated proteins, as
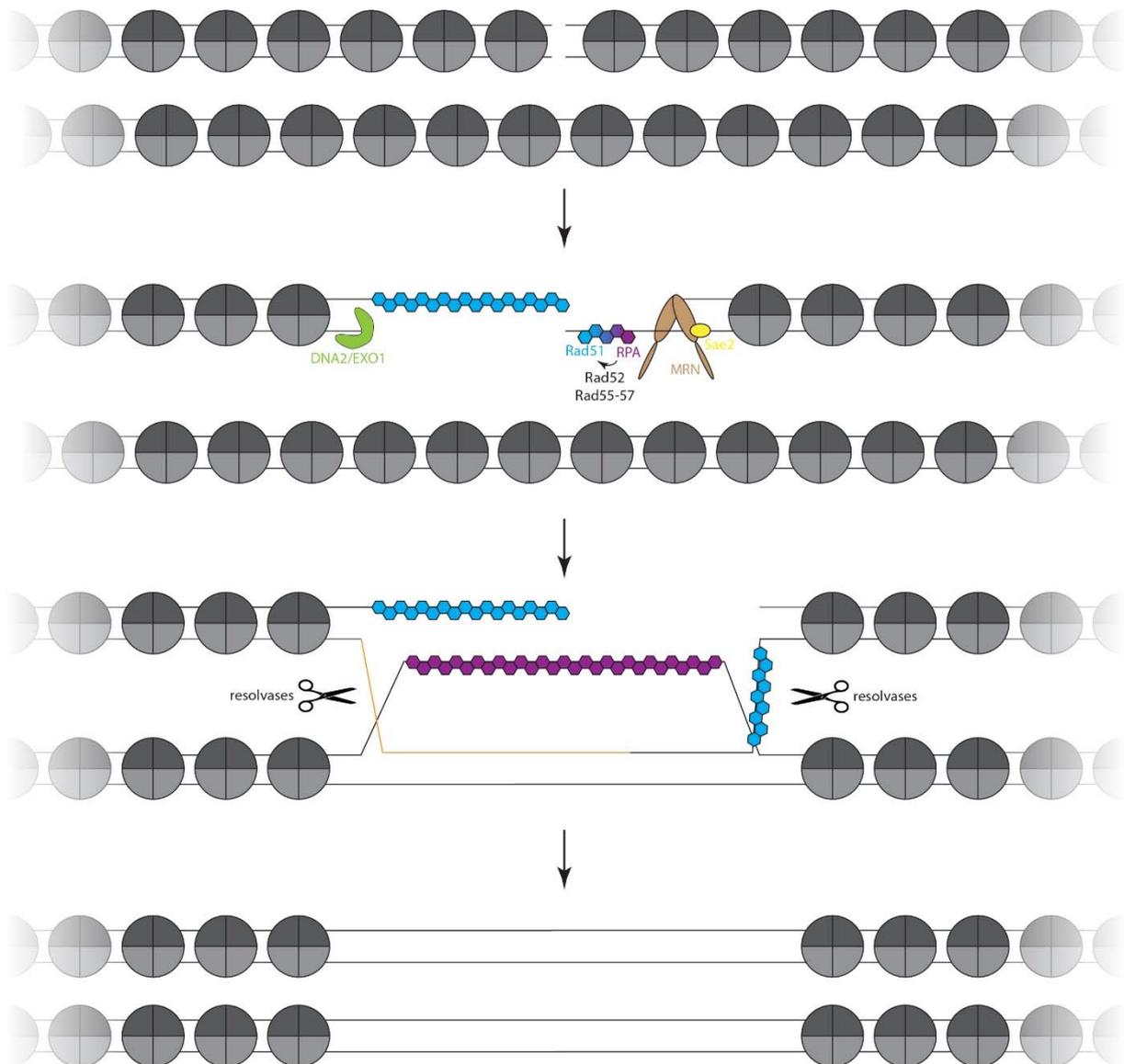
Fig.50: schematic representation of HR, described and based on section 2.1.2.

for example Rad9, are able to block resection [103-105]. Recruitment of Rad9 to chromatin is a hallmark of DNA double strand break signalling. It works as a checkpoint adaptor and is known to bind S139 Phosphorylated H2A.X Histones in the vicinity of double strand breaks to prevent the degradation of the exposed DNA [106].

2.1.3 Fun30 in long range resection

The ATP dependent single subunit chromatin remodeller Fun30 (Function unknown now 30, SMARCAD1 in humans) has initially been reported to localize to DSBs [107-109]. Its depletion causes similar effects as Exo1/DNA2 mutants, indicating a function of Fun30 in the long range resection process in DSB repair. Further evidence for its role in this pathway is provided by

effects on initial DNA end resection in-vivo [108-110]. For efficient activity of Fun30, its phosphorylation by the cyclin dependent kinase (CDK) is required [111]. This modification leads to the interaction with the phosphoprotein-binding scaffold protein Dpb11 and DNA damage sensor 9-1-1, localizing the complex to the site of the DSB [107]. Increased Rad9 occupancy at the site of DSB in the absence of Fun30 shows the remodeller to antagonize the function of Rad9 as roadblock for resection [108]. Therefore, a current model for the influence of the Fun30 chromatin remodeller on DSB repair long range resection is shown in Fig.51.

In this model the phosphorylation of Fun30 by CDK1 leads to an interaction with Dpb11 and a localization of Fun30 to DSBs. A complex of Fun30, Dpb11 and 9-1-1 subsequently leads to a removal of Rad9 bound histones allowing access to DNA to the long range resection machinery.
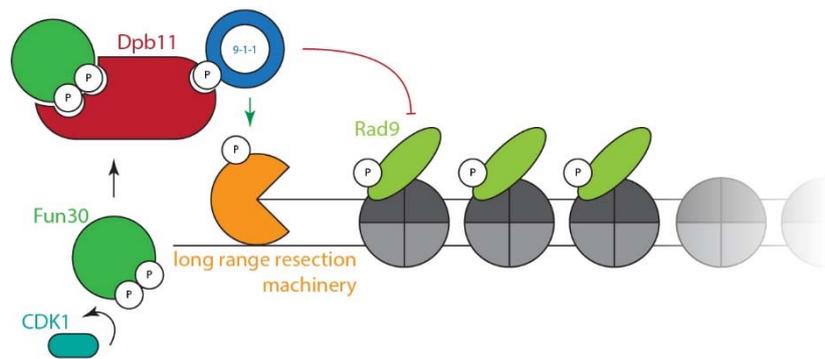
Fig.51: model of Fun30 function and recruitment in long range resection

Recent findings furthermore suggest a role of Fun30 in promoting RNA polymerase II (RNAPII) transcription. In the yeast *Saccharomyces pombe*, Fun30 induces nucleosome disassembly at transcription sites to facilitate RNAPII propagation through the chromatin environment together with the histone chaperone FACT [112].

Fun30 is conserved amongst eukaryotes, and has homologues in *Saccharomyces cerevisiae* (Fun30), *Saccharomyces pombe* (Fft-proteins), mouse (Etl-proteins) and human (SMARCAD1 with various isoforms). It possesses a Snf2-ATPase that follows a long N-terminal linker. For this linker no domain structures can be detected by bioinformatics. In some organisms, however not all, a CUE-sequence-motif has been be detected and is shown to be important for the activity of Fun30 [113]. In-vitro characterization of Fun30 from Saccharomyces cerevisiae revealed DNA- and chromatin-stimulated ATPase activity, as well as dimerization of Fun30. Furthermore, binding to nucleosomes, as well as histone-exchange was shown [114].

<u>Aim of this project</u>

One aim of this thesis is to gain insights into the mechanism of activation of Fun30 as well as the relevance of the N-terminus on the dimerization. Therefore, the homologue of Fun30 in *Chaetomium thermophilum* (ctFun30) was identified by bioinformatics means in order to facilitate purification and structural analysis by utilizing the enzyme of a thermophilic organism. Functional assays, such as DNA- and nucleosome-binding assays, as well as ATPase and remodelling assays are considered to be helpful in analysing the functional activity of ctFun30. Crystal screens of various constructs and different DNA length are applied, as a structure of ctFun30 might help to shed light on its mechanism of function and regulation.

2.1.4. Ino80 in short range resection

Another chromatin remodeller known to play a role in DSB repair by HR is INO80 (<u>ino</u>sitol requiring <u>80</u>), as INO80 deficiency causes strong recombination defects [115, 116]. A reduction of the
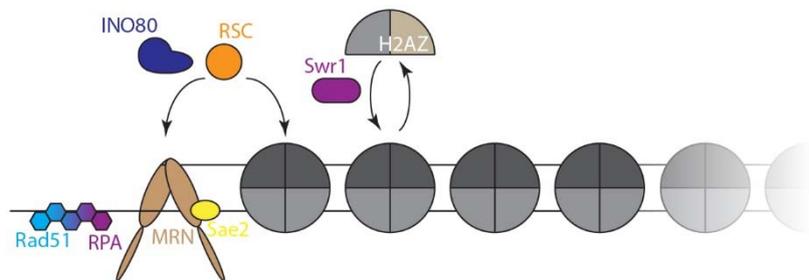
Fig.52: model of INO80 function in short range resection

number of Rad51 foci in INO80 deficient cells is considered to have an impact on HR [117]. Absence of Ino80 leads to a reduction of ssDNA by reduced DNA end resection. Furthermore an increased amount of RPA can be found on ssDNA, as the cells are incapable of exchanging RPA with Rad51 [118]. This effect is mainly considered to be caused by insufficient exchange of the histone variant H2A.Z in INO80 deficient cells [118]. This indicates, that INO80 is the antagonist of the chromatin remodeller SWR1 responsible for H2A.Z incorporation at the sites of DSB [119-121] (Fig.52) .

2.1.5. Other functions of INO80

Despite other chromatin remodellers playing a role in promoter architecture, INO80 is the only remodeller responsible for the positioning of the +1 nucleosome next to the nucleosome free region at the transcription start site in genome wide *in-vitro* assays in the absence of other factors [122]. Ino80 is known to have an impact in several cellular functions by altering

expression levels several important genes that encode for example proteins in phospholipid biosynthesis and phosphate metababolism [123, 124]. INO80 possesses a further role as transcriptional co-regulator by repressing the expression of non-coding transcripts [125, 126]. Additionally, INO80 is involved in assisting DNA replication recovery at halted replication forks by facilitating the release of stalled RNAPII from chromatin [127].

2.1.6. Structure and subunit composition of Ino80

INO80 is a multiprotein complex composed of a set of 9 conserved subunit, as well as of a varying number of subunits specific for the organism (Fig.53; adapted from [128]). The major subunit of the INO80 is the Ino80-ATPase. In INO80, as well as in all members of the INO80 family, the core ATPase is formed by a superfamily 2 (SF2) DEAD/H-box helicase [129]. The Snf2 ATPase consists of a RecA-like fold in each of the two lobes (DEXX and HELICc) separated by an insertion domain. Biochemical data, as well as crosslinking and structural analysis revealed a modular organization of yeast and human INO80, allowing a clustering of subunits into distinct submodules [128, 130-132]. Recent cryo-EM structures of the INO80 complex bound to a nucleosome has revealed a molecular basis for the remodelling reaction [133, 134]. The Rvb1/Rvb2 AAA+ ATPase heterohexamer acts as an assembly scaffold for the complex, stabilizing the Ino80 insertion domain. The RecA motor domains bind to the DNA entry side of the nucleosome. The nucleosome is fixed on the opposite side by the Arp5-Ies6 module, DNA
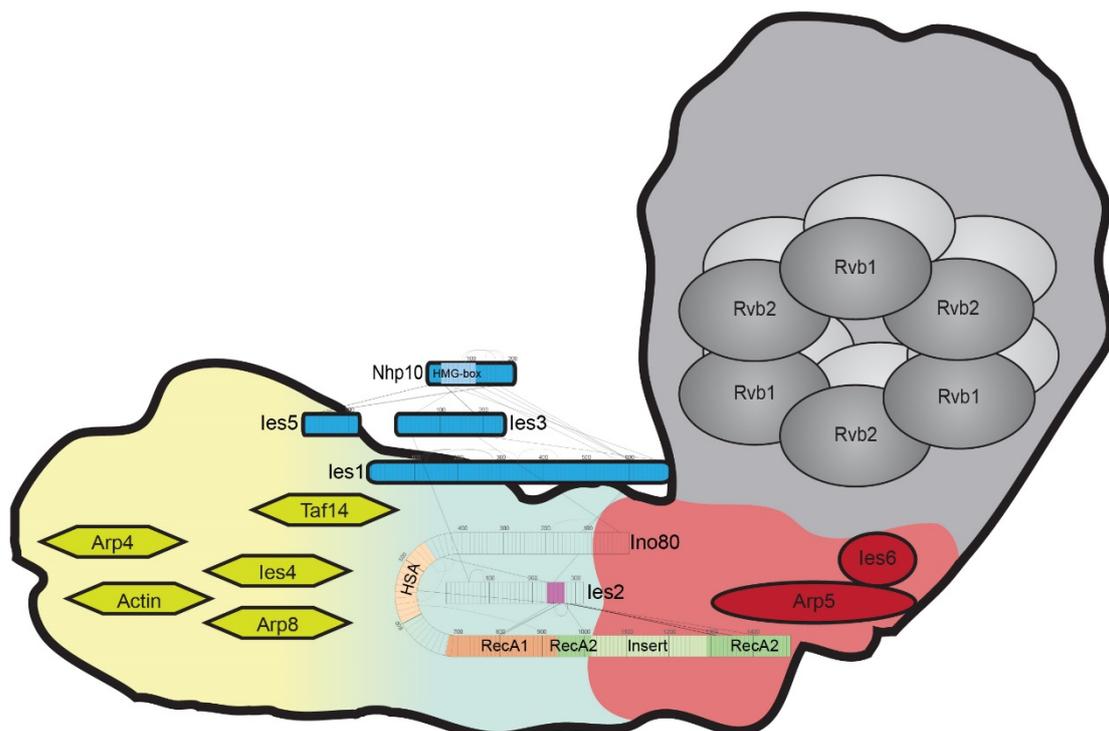

Fig.53: Subunit composition of INO80 with crosslinking data of the Nhp10 module

pumping of the ATPase against the Arp5-Ies6 grip leads ultimately to a larger nucleosome translocation step. This mechanism differs drastically from the one observed by the single-chain remodeller, e.g. Chd1 that bind to the super-helical location (SHL) 2 on the nucleosomal DNA [135, 136]. In addition to its catalytic core activity, the Ino80 ATPase comprises a scaffolding function for the additional modules. Directly, N-terminal of the ATPase region is a segmented HSA (helicase-SANT-associated) domain, that binds to the ARP-module of the INO80 complex, composed of Actin, Arp4, Arp8, the TATA-binding protein-associated factor Taf14 and Ies4 [128, 130, 137-139]. All three actin-fold proteins use their barbed ends to bind to three segments of the HSA helical elements. The ARP module thereby structures the HSA domain in a way, that the solvent exposed side of the HSA helices provide a binding platform for extra-nucleosomal DNA. This binding interface has been shown to be required for DNA sliding and +1 nucleosome positioning [140].The N-terminus of the Ino80 ATPase is bound by the species specific subunits, which differ between organisms. In *Saccharomyces cerevisiae* the species specific submodule is termed Nhp10-module named after the central protein Nhp10 (non-histone protein 10). Additional proteins of the Nhp10 module are Ies1, Ies3 and Ies5. The Nhp10 module is not directly required for INO80 sliding activity [128, 141], however it is required for nucleosome binding [128] and causes a switch-like response of Ino80 to varying length of flanking DNA [141]. Therefore, the Nhp10 module is crucial for recruitment of INO8 to DSB and might potentially play a role in localizing INO80 to the +1 nucleosome and the nucleosome free region [142-144]. Between the species no significant similarities can be found within the components of the species specific subunits. It is therefore assumed to play an important role in adapting to the variations in chromatin structures and regulation among the different organisms.

2.1.7. Nhp10 and HMG boxes

The Nhp10 protein as the central component of the species specific Nhp10 module is composed of two HMG (high mobility group) boxes, Box A and Box B followed by an acidic tail (Fig.54). Both domains are considered to bind to DNA, whereas

Fig.54: domain architecture of Nhp10

the acidic tail reduces the affinity for DNA [145]. The HMG box B possesses the sequence of a classical HMG box domain, but the HMG Box A bears only little similarity to other HMG box

domains. In general, the L-shaped from of HMG boxes binds the DNA via the minor groove by a partial intercalation of hydrophobic residues. A preference for distorted DNA is reported [146]. In other proteins harbouring two HMG box domains Box B is reported to contribute primarily to DNA binding affinity whereas BoxA is responsible for bending of the DNA. For Nhp10 both HMG boxes display a preference for distorted DNA, which might contribute to the recruitment of INO80 to stalled replication forks or other non-canonical DNA structures. In Addition, binding to both blunt and sticky DNA ends is reported for the HMG Box A, thereby leading to a protection of the DNA form exonuclease digestion. That leads to the hypothesis of Nhp10 playing a role in DSB repair beyond INO80 recruitment [145, 146].

Aim of this project

No structural information is known about the Nhp10-Ies1-Ies3-Ies5 subcomplex. The aim of this project is to structurally characterize this complex using crystallography. Furthermore, DNA binding studies with the purified complex are part of the project, to gain in-depth knowledge of the recruitment of the Ino80 complex to the DNA, as well as to help crystallization.

2.2. Methods

2.2.1. The Nhp10 submodule of the INO80 complex

In the following buffers used in the experiments are abbreviated with T for Tris, H for Hepes, P for Phosphate buffer, N for sodium chloride, G for Glycerol, I for Imidazole, ß for ß-mercaptoethanol, Mg for Magnesium, S sucrose, E for EDTA and db for desthiobiotin. Values in brackets state the pH and indexed numbers give the concentration in mM.

<u>Cloning</u>

Standard Cloning protocols were used. Vector maps of the plasmids used in this thesis are listed in the Appendix if not stated differently.

<u>Cloning of His-tagged Nhp10, Ies1, Ies3, Ies1, (Ino80p_1-450)</u>

A pFBDM vector containing His-tagged Nhp10, Ies1 and Ies3 was present in the lab and furthermore modified by fusion with a vector containing Ies1 and Ino80p_1-450. This vector was cloned using the following primers:

 Ies1-fwd   GCGCTCGAGATGGGGAAGAGAGTATACGA

 Ies1-rev   CGCGCTAGCTCACTCCAGCTTAAACATGGCGGTT

 Ino80_1-fwd   GCG GTCGAC ATGTCACTGGCAGTTCTACT

 Ino80_450-rv   CGCGCGGCCGCCAACCCGTGTCTAGTGTTGA

<u>Purification of His-tagged Nhp10 in complex with Ies1, Ies3, Ies5 and Ino80p_1-450 from H5 cells</u>

H5 cells were infected with 2mL/L virus based on the MultiBac system containing His-tagged Nhp10, Ies1, Ies3, Ies5 and Ino80p_1-450 (a fragment of the Ino80 ATPase containing the first 450 amino acids) at a density of 1 Million cells per mL. The infected cells were shaken at 27 °C for 72 h. Cells are harvested at at 4,000 rpm for 10 min and then were resuspended in $P_{25}N_{1000}G_{10}ß_3Mg_4$ supplemented with Protease Inhibitor. The cells were lysed by 8 cycles of sonication for 40 sec followed by 20 sec cooling. To remove cell debris, the lysate was differentially centrifuged s at 15.000 rpm in an SS34 rotor before loading onto a NiHisTrap crude FF column (GE Healthcare). The column was washed intensively with lysis buffer and eluted by applying a gradient with increasing imidazole concentration (from I30 to I300).

Fractions were selected based on several criteria: protein concentration and purity by SDS-page analysis and the absorption ratio 260nm/280nm of less than 1 than exclude nucleic acid contaminations. The chosen fractions were pooled and dialysed overnight against $H_{25}N_{200}G_{10}Mg_4ß_3$ to remove imidazole and exchange the phosphate buffer. After dialysis the sample was loaded onto a Heparin column. The Heparin column was eluted in a salt gradient from 200 mM to 1000Mm NaCl. DNA-free fractions were pooled and concentrated for size exclusion chromatography. Typically, size exclusion chromatography on a superpose 6 10/300 in $H_{25}N_{200}G_3Mg_2ß_1$ yields two peaks, one containing the full 5 component complex and the other one lacking the Ino80p_1-450 construct.

Purification of His-tagged Nhp10, Ies3, Ies5, (Ies1) from H5 cells

H5 cells were infected and cultivated in a similar manner as described for the 5 component complex, with a virus lacking the Ino80p_1-450 fragment. The major difference is that cell lysis was done by addition of 1% NP40 keeping the nuclear envelope intact. This method allows separation of DNA-bound complexes in the nucleus from DNA-free complex in the cytoplasm and thus reduces nucleic acid contamination in the protein preparation. The obtained cell nuclei were pelleted at 6000 g for 30 min. Additionally, the supernatant was supplemented with benzonase to remove any DNA from damaged nuclei, and centrifuged again at 18000rpm for 30 min in an SW40 rotor. After loading the lysate twice on a NiHisTrap crude FF the column was extensively washed using a chaperone wash buffer comprising ATP to remove chaperones and impurities. Bound proteins were eluted in an Imidazole gradient from 60 mM to 250 Mm sodium chloride. Fractions containing protein free of DNA contaminations were pooled and injected into size exclusion chromatography on a Superose 6 10/300. The major peak fractions were analysed by SDS-page showing a pure complex containing HisNhp10, Ies3 and Ies5. Ies1 however seemed to be lost during purification in the absence of the Ino80p_1-450 fragment.

Purification of nanobodies from Rosetta cells

E.coli cells expressing the nanobody #42 into the periplasm were grown in LB media at 37°C to an OD_{600} of around 0.7 and subsequently induced with 0.5 mM IPTG. After shaking at 22 °C for 2.5h cells were harvested and shock frozen. The pellet was resuspended in $T_{50}S_{500}E_{0.5}$ and incubated with lysozyme for 30 min. After centrifugation at 13000 rpm for 30 min, the lysate was loaded onto strep-beads. After washing with lysis buffer the column-bound beads were used for further purification of epitope containing complexes.

Purification of Nhp10, Ies3, Ies5, (Ies1) with nanobody #42 from H5 cells

H5 cells were infected with a virus containing 10 different subunits of the INO80 complex (exact), including Nhp10, Ies1, Ies3 and Ies5. The cells were harvested and cytoplasmically lysed in $P_{25}N_{500}G_{10}$ as described in the purification of the 3 component complex. For affinity purification, strep beads bound to nanobody #42 were used as affinity-matrix. Bound proteins, were eluted together with the nanobody 42 with $P_{25}N_{500}G_{10}db_{2.5}$. The eluted protein was concentrated and injected into Superose 6 10/300 size exclusion chromatography column in a buffer containing $T_{25}N_{300}G_3DTT_2$. The sample eluted in one major peak containing Nhp10, Ies3, Ies5, nanobody #42 and a substoichiometric fraction of Ies1.

Purification of HisIes5, Nhp10 from Rosetta cells from two different plasmids

Two plasmids, cloned by Sebastian Fenn, containing HisIes5 and Nhp10 respectively were cotransfected into Rosetta cells and grown at 37°C to an OD_{600} of approximately 0.7. After induction with 0.5 mM IPTG cells were incubated at 18°C on. After Harvesting the cells were resuspended in $T_{25}N_{500}G_{10}I_{30}$ supplemented with Protease Inhibitor. The cells were sonicated three times for 5 min before pelleting at 15000 rpm for 30 min. Benzonase was added and the lysate loaded on NiHisTrap crude FF column, extensively washed and eluted in a Imidazole gradient of 30 mM to 250 mM. Despite the addition of protease inhibitor, Ies5 was heavily degraded.

Purification of HisNhp10, Ies3, Ies5 in Rosetta cells from three different plasmids

HisNhp10, Ies3 and Ies5 were cotransformed from 3 distinct plasmids into Rosetta cells. The cells were grown at 37°C and induced with 0.5 mM IPTG at an OD_{600} of 0.8. After overnight incubation at 18°C cells were harvested and lysed in $T_{25}N_{200}G_{10}I_{30}$ by sonication in the presence of protease inhibitors. After centrifugation at 15000 rpm for 30 min, the supernatant was loaded onto Ni-NTA gravity beads, washed extensively with lysis buffer and eluted in $T_{25}N_{300}G_{10}I_{200}$. In the elution fractions only Nhp10 was prominently present.

Purification of HisNhp10, Ies3, Ies5 in Rosetta cells from two plasmids with nanobody #42

As purification from two plasmids showed decent yield but either low stability or DNA contamination, stabilization by addition of nanobody #42 was tested. Therefore, Rosetta cells containing a plasmid with HisNhp10 and one plasmid containing Ies3 and Ies5 were grown to an OD_{600} of around 0.8. After induction with 0.5 mM IPTG cells were incubated over night at

18°C. Lysis occurred in $T_{25}N_{500}G_{10}$ in the presence of protease inhibitors by sonication. After centrifugation at 15000 rpm for 30 min benzonase was added and the lysate loaded onto strep-beads bound by nanobody #42. After an ATP-wash, the complex was eluted by addition of 2.5 mM desthiobiotin. Elution was concentrated and injected into S200 10/300 in T25N300G3. The size exclusion profile showed one peak in the void volume containing DNA-contaminated sample and one major peak of DNA-free complex.

DNA-binding detection by Electrophoretic Mobility Shift Assays (EMSA)

For testing of DNA binding an EMSA was used. Therefore, 50 nM DNA was mixed with increasing concentrations of the protein in a total volume of 20µL. The samples were incubated for 20 min and then loaded on a 1% Agarose gel. After 1h at 100V the gel was analysed by fluorescence detection.

2.2.2. Fun30 chromatin remodeller

Purification of Fun30 constructs

After extensive testing the best purification strategy for Fun30 was found to be suitable for all constructs and organisms. Therefore, His-tagged Fun30 was transformed into Rosetta cells, which were grown at 37°C to an OD_{600} of 0.4. After induction with 0.5 mM IPTG the temperature was reduced to 18°C for overnight incubation. After Harvesting, the cells were resuspended in $T_{25}(8)N_{1000}G_{10}I_{30}$, sonified twice for 8 min and spun down at 13.000 rpm for 30 min. After addition of high-salt active benzonase, gravity Ni-NTA beads were used for affinity purification. Stepwise elution with Imidazole concentrations of 30 and 60 mM were employed to remove contaminations and DNA-bound sample. Elution of the protein of interest occurred at 200 mM Imidazole. In overnight Dialysis against 500 mM salt, the tag was removed by addition of Precision Protease. Dependent on the purity of the sample a Heparin column was used in case of remaining nucleotide contaminations or smaller degradation fragments. However, this step was shown to be very complicated, as Fun30 constructs are mainly unstable in low salt conditions required for binding to Heparin columns. Other columns proofed to be inefficient to remove contaminations and keep the protein soluble. Protein containing fractions were pooled, concentrated and injected into an S200 size exclusion column.

Limited Proteolysis

Purified protein was used for limited proteolysis at a concentration of 1 mg/mL. Proteases used for the assay were Trypsin, Chymotrypsin and Subtilisin at concentrations of 0.1 mg/mL, 0.01 mg/mL and 0.001 mg/mL. The samples were incubated for 30 min at room temperature and analysed by SDS-page. Promising bands were blotted on nitrocellulose membrane by Western Blotting, stained with Ponceau red and the excised bands were send for N-terminal sequencing.

Stability assays of ctFun30_333-978

To improve the buffer conditions used in the purification thermal shift assays assays were performed with the Rubic Screen from Molecular Dimensions. Therefore, 20µM of protein were used. For detection of hydrophobicity SYPO-orange was used.

DNA-binding detection by Electrophoretic Mobility Shift Assays (EMSA)

To test for DNA-binding, 50 nM DNA was mixed with increasing concentrations of the protein of interest in a total volume of 20µL. After 20 min incubation, the samples were loaded on a 1% Agarose gel. After 1h at 100V the gel was analysed by fluorescence detection in case of fluorescently labelled DNA, or by Gel-red detection.

Nucleosome-binding Assay

For testing the binding of Fun30 constructs to nucleosomes, 50 nM nucleosomes were used in a total volume of 10µL. Varying concentrations of constructs of interest were added to the nucleosomes and incubated for 30 min. After addition of 2µL ficol, the samples were loaded on a 3-12% native page gel and run for 3h at 130V. Cy3 and Cy5 labels at the nucleosomes or the nucleosomal DNA were detected using a bioimager.

Right-Angle Light-Scattering size exclusion chromatography

For determining the accurate size of a protein-peak in gel-filtration, around 2 mg/mL of sample of interest was loaded onto a S200 10/300 column connected to the Viscotek 270 Dual Detector. Results were analysed using the Omnisec software in comparison to BSA-standard.

NADH-coupled ATPase assay

ATPase activity of Fun30 was tested following the protocol presented in [147].

Small-Angle X-ray Scattering (SAXS)

SAXS was performed after a size exclusion chromatography run on a S200 10/300 in a buffer of $T_{25}(8)N_{500}G_3Mg_2$. Therefor, 75µL were injected at a concentration of 10 to 20 mg/mL with and without ATPγS.

Test purification of SMRCD1a&c

Gene expression of SMRCD1a&c in E.coli was conducted in a similar manner as ctFun30. Expression in H5 cells occurred with 1mL of virus per 1L of H5 cells at 1 million cells/mL for 72 h. Cells were lysed in $T_{25}(8)N_{1000}G_{10}Mg_2I_{10}$ by sonicating twice for 8 min and spun down at 15.000 rpm for 30 min. The supernatant was treated with benzonase and loaded on Ni-NTA beads. The beads were washed with 30 and 60 mM Imidazole prior to elution with 200 mM Imidazole.

2.3. Results

2.3.1. The Nhp10 submodule of INO80

<u>Purification of His-tagged Nhp10 in complex with Ies1, Ies3, Ies5 and Ino80p_1-450 from H5 cells</u>

Experiments on components of the Nhp10 submodule previously performed in this group were based on expression from insect cells. Therefore, purification of the His-tagged Nhp10 in complex with Ies1, Ies3, Ies5 and Ino80p_1-450 was initially based on H5 insect cell expression. Analysis of the purification by SDS-page and
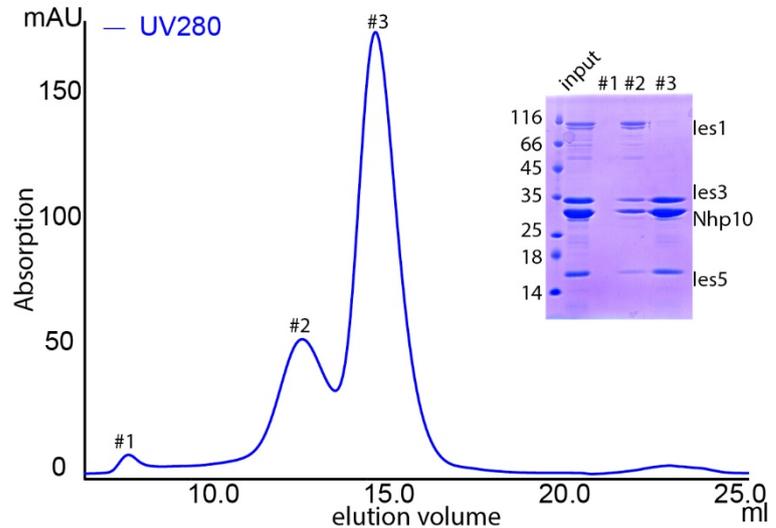
Fig.55: Purification of His-Nhp10, Ies1, Ies3, Ies5, Ino80_1-450 with size exclusion chromatography and SDS-PAGE

Gel-filtration indicated predominantly clean protein with few degradation bands of Ies1 and one minor degradation band of Nhp10 (Fig.55). Determination of this construct however, remained unsuccessful, as Western-Blot and other detection methods on the N-terminal His-tag remained inconclusive.

<u>Purification of His-tagged Nhp10, Ies3, Ies5, (Ies1) from H5 cells</u>

To avoid instability through degradation of the Ino80_1-450 fragment, a four-component complex containing His-tagged Nhp10, Ies3, Ies5 and Ies1 was designed. The purification yielded a stable complex of Nhp10, Ies3 and Ies1. However, Ies1 seems to be lost during the purification, indicating the necessity of the Ino80 fragment to stabilize the complex (Fig.56).

Fig.56: Purification of His-Nhp10, Ies1, Ies3, Ies5 with size exclusion chromatography and SDS-PAGE

Purification of Nhp10, Ies3, Ies5, (Ies1) with nanobody #42 from H5 cells

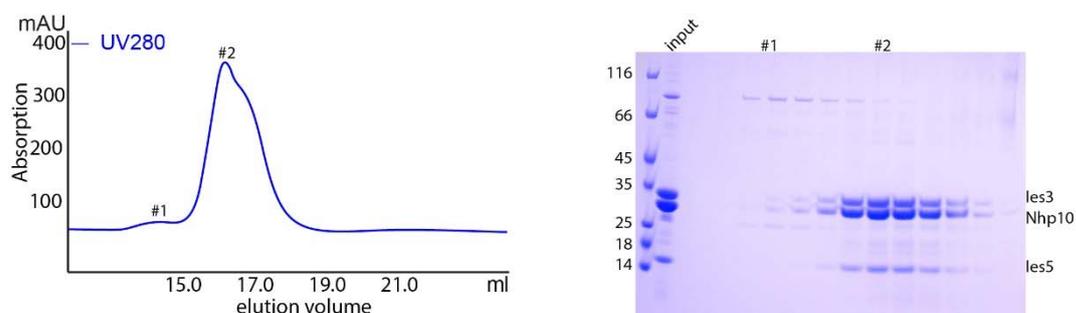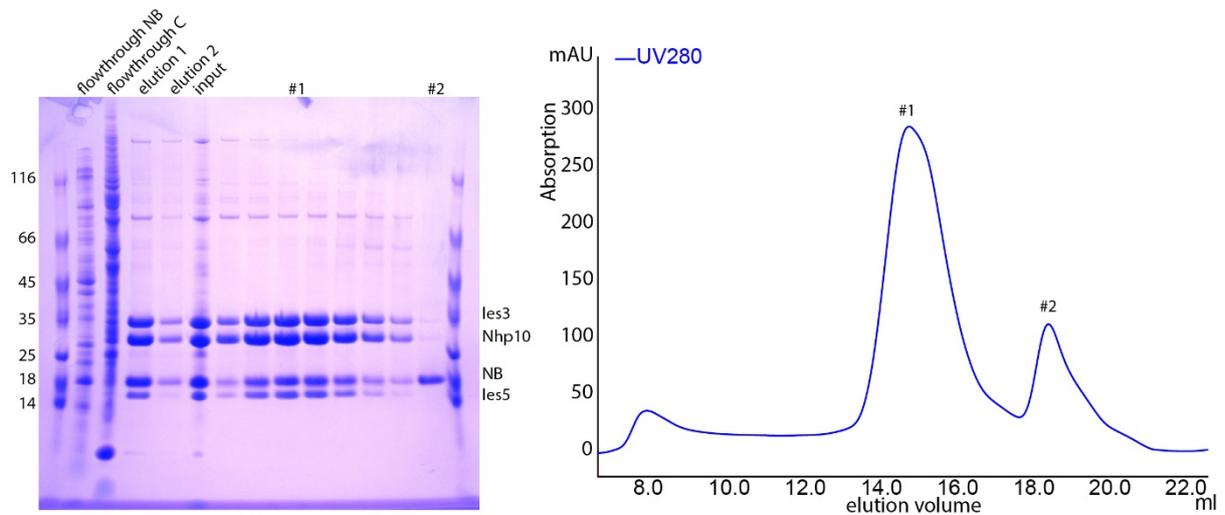Fig.57: Purification of His-Nhp10, Ies1, Ies3, NB with size exclusion chromatography and SDS-

To stabilize the Nhp10 submodule further, a nanobody (NB) derived against the Ino80 complex and specific against the Nhp10 module (Tosi PhD) was used in the purification. To obtain the most stable subcomplex endogenously present in the INO80 complex, the nanobody was employed in a pull down of insect cells expressing 10 different constructs, including the Nhp10 module. This vector is generally used in the lab for purification of the INO80 complex. The purification yielded a stable complex of Nhp10, Ies3 and Ies5 bound by the nanobody, with substoichiometric amounts of Ies1 (Fig.57).

Extensive crystallization attempts yielded no promising hits.

Purification of HisIes5, Nhp10 from Rosetta cells from two different plasmids

As Ies1 seems to be mainly transiently bound to the complex, a minimal, stable complex was searched for. With Ies1 being the only component insoluble in *Escherichia coli*, as previously shown by Sebastian Fenn, different combinations of proteins were tested for complex formation in *Escherichia coli*. The combination of His-tagged Ies5 together with Nhp10 showed complex formation in Ni-NTA pulldown, however the complex seems to undergo heavy degradation indicating the lack of an interaction partner stabilizing the complex formation (Fig.58).
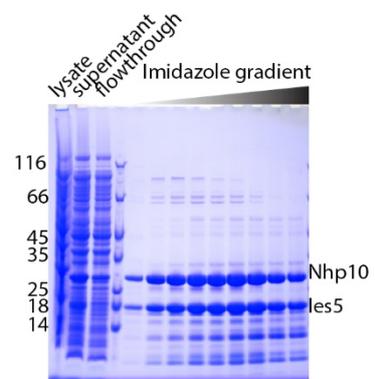
Fig.58: Testpurification of His-Nhp10, Ies5

Purification of HisNhp10, Ies3, Ies5 in Rosetta cells from three different plasmids

To stabilize the ies5-Nhp10 interaction, Ies3 was added by co-transformation of three different plasmids. However, the purification by Ni-NTA showed no pull down effect on Ies3 and Ies5 by the His-tagged Nhp10 (Fig.59).

This might indicate that the expression levels from three different plasmids are undergoing too much variation as to allow the formation of a stable complex.
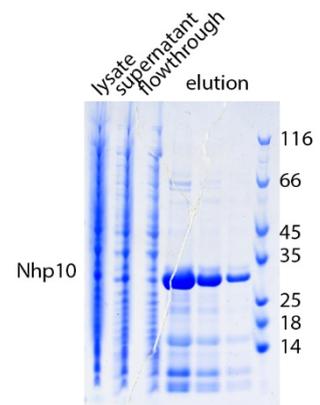
Fig.59: Testpurification of His-Nhp10, Ies3, Ies5

Purification of HisNhp10, Ies3, Ies5 in Rosetta cells from 2 plasmids with nanobody #42

To enhance complex formation and to level expression levels of the different components, cotransformation of a pET28 vector containing Nhp10 with pETDuet vector expressing Ies3 and Ies5 was tested, indicating a complex formation (Fig.60).

However, this complex was showing tremendous DNA contamination, and lost structural stability upon removal of DNA. Therefore, purification by interaction with the nanobody was employed. This yielded a stable four-component complex free of DNA contaminations (Fig.61).

Despite extensive crystal screening, no suitable hits could be found.
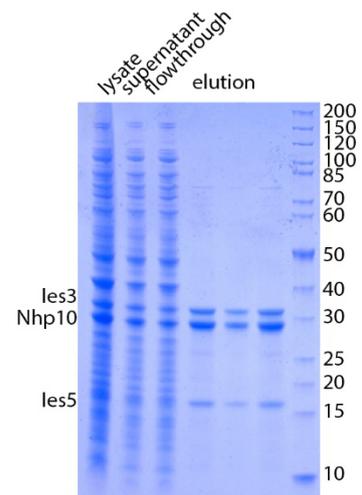
Fig.60: Testpurification of His-Nhp10, Ies3, Ies5 with
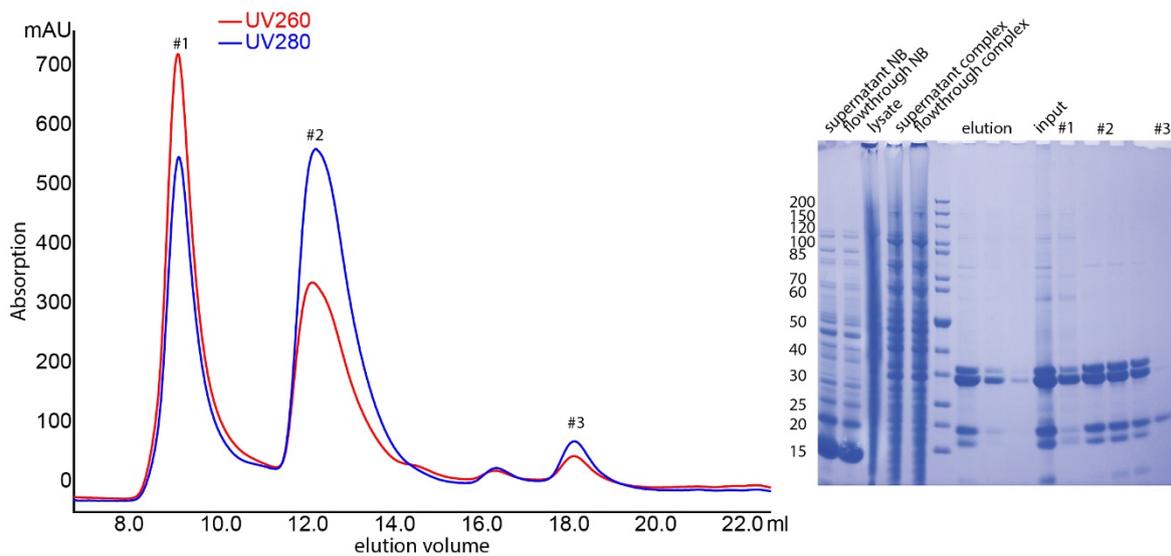
Fig.61: Size exclusion chromatography and SDS-PAGE of Purification of His-Nhp10, Ies3, Ies5 with NB

<u>DNA Binding Assays of HisNhp10, Ies3, Ies5, nanobody #42</u>

EMSAs were performed with different DNAs. First different DNA length were used. Therefore EMSAs were performed with 16mer, 20mer, 30mer and 40mer double stranded DNA with random sequences. Furthermore, to test for sequence specificity two distinct sequences were used named Reb1-seq and anti-Reb1-seq. These sequences comprise 50 bp double stranded DNA designed according to the DNA-binding specificity of Reb1. In the literature the Reb1 binding sequence was shown as specific binding sequence of Nhp10 in CHIP-exo experiments. To examine whether this specificity is due to a direct sequence specificity or based on an interaction between Nhp10 and Reb1, affinity of the Nhp10 module to the Reb1-seq was compared to anti-Reb1-seq. The DNA sequence bound to Reb1 was picked as Reb1-seq and as anti-Reb1-seq the sequence in the same distance to the nucleosomal dyade of a non-Reb1 bound nucleosome were chosen. The comparison of both DNAs on an EMSA showed no difference in affinity (Fig.62). Also no significant difference was detectable in EMSAs of the DNAs of different length (Fig.63).
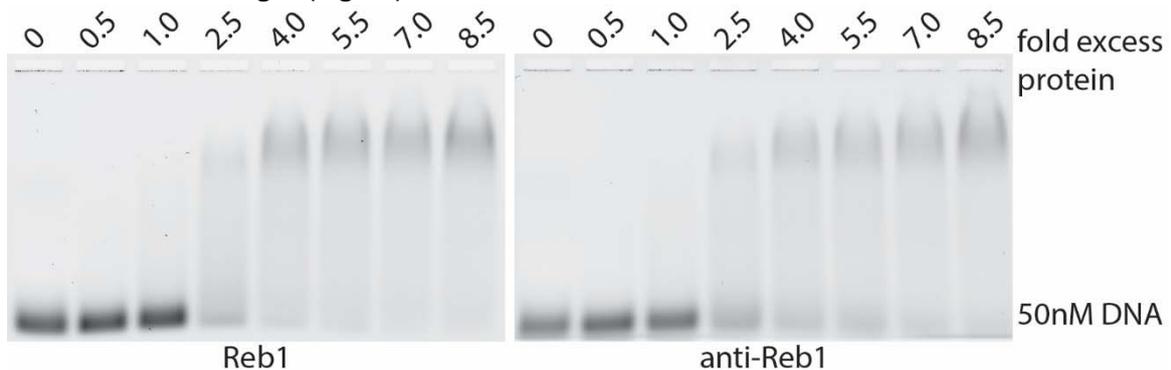
Fig.62: DNA binding assays with Reb1 and anti-Reb1 sequences
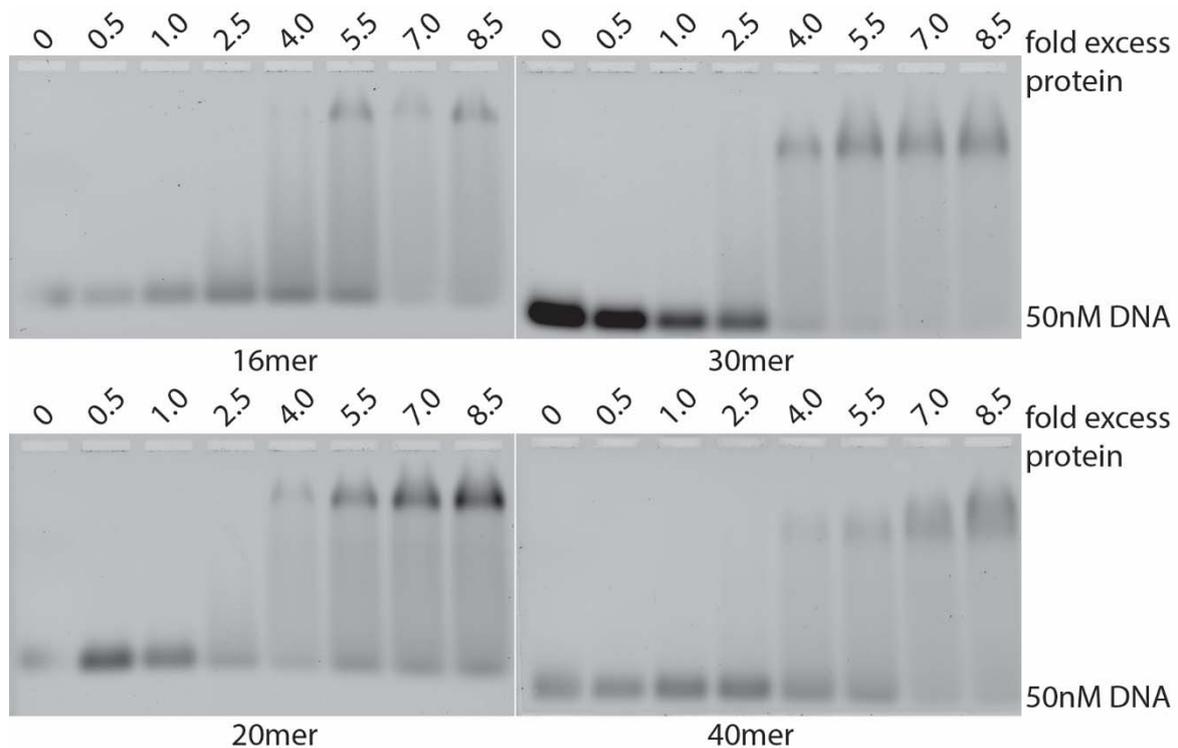
Fig.63: DNA binding assays with DNA sequences of varying length

Bioinformatic analysis of Nhp10

Nhp10 is reported in the literature [145, 146] to be composed of two HMG-boxes, HMGB-A and HMG-B, with HMG-A showing only little sequence similarity to classical HMG-boxes. Both HMG-boxes are considered to bind DNA. Submission of Nhp10 to BLAST-search reveals no obvious similarities to HMG-boxes. With MODELLER [148] implemented in the MPI Bioinformatics Toolkit [149] a 3D model was generated based on the sequence specific HMG-box protein Nhp6 (Fig.64) and analysed according to the parameters of HMG-box proteins as published in [150].
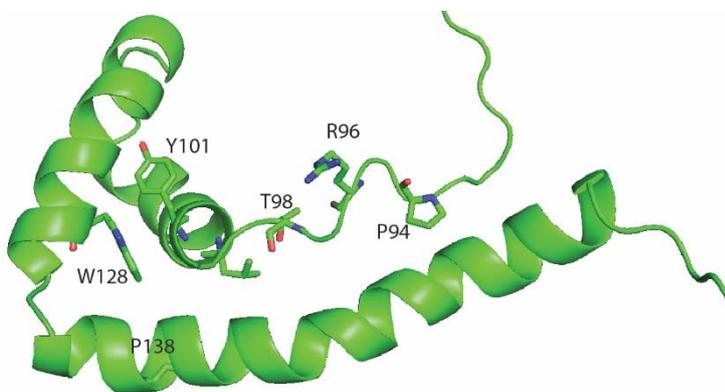
Fig.64: Model of the HMG-B box of Nhp10 based on the structure of Nhp6A (pdb:1J5N) using MODELLER

According to this publication the prolines at position 138 and 94 lead to the formation of a kink in helix 3, characteristic for sequence unspecific HMG-boxes. This analysis is in accordance with the DNA-binding studies performed by EMSA. In addition, putative amino acids involved in DNA-binding

could be identified: the Threonine at position 98 most likely interacts with DNA and the tyrosine at position 101 stabilizes the DNA through stacking. Furthermore, the work published in [150] showed, that the Tryptophan at the end of helix2 (in HMGB-B position 128) is essential for the formation of a HMG-box. However, no Tryptophan can be found in the N-terminal region of HMGB-B, indicating that the domain HMG-A, previously assigned to be an atypical HMG-box, does not form the L-shaped structure of an HMG-box. Purification attempts of a construct of Nhp10 comprising the first 95 amino acids remained unsuccessful. This indicates, that the N-terminal domain of Nhp10 might be unstructured and responsible for the interaction with interaction partners, such as other subunits of the INO80 complex or the nucleosome. An interaction with DNA, as published in [145, 146] could not be verified.

2.3.2. Fun30 chromatin remodeller

<u>Purification of full-length Fun30 from *Chaetomium thermophilum* (fl ctFun30)</u>

Purification of full-length ctFun30 yielded decent amounts of clean protein. However, few degradation bands were visible.
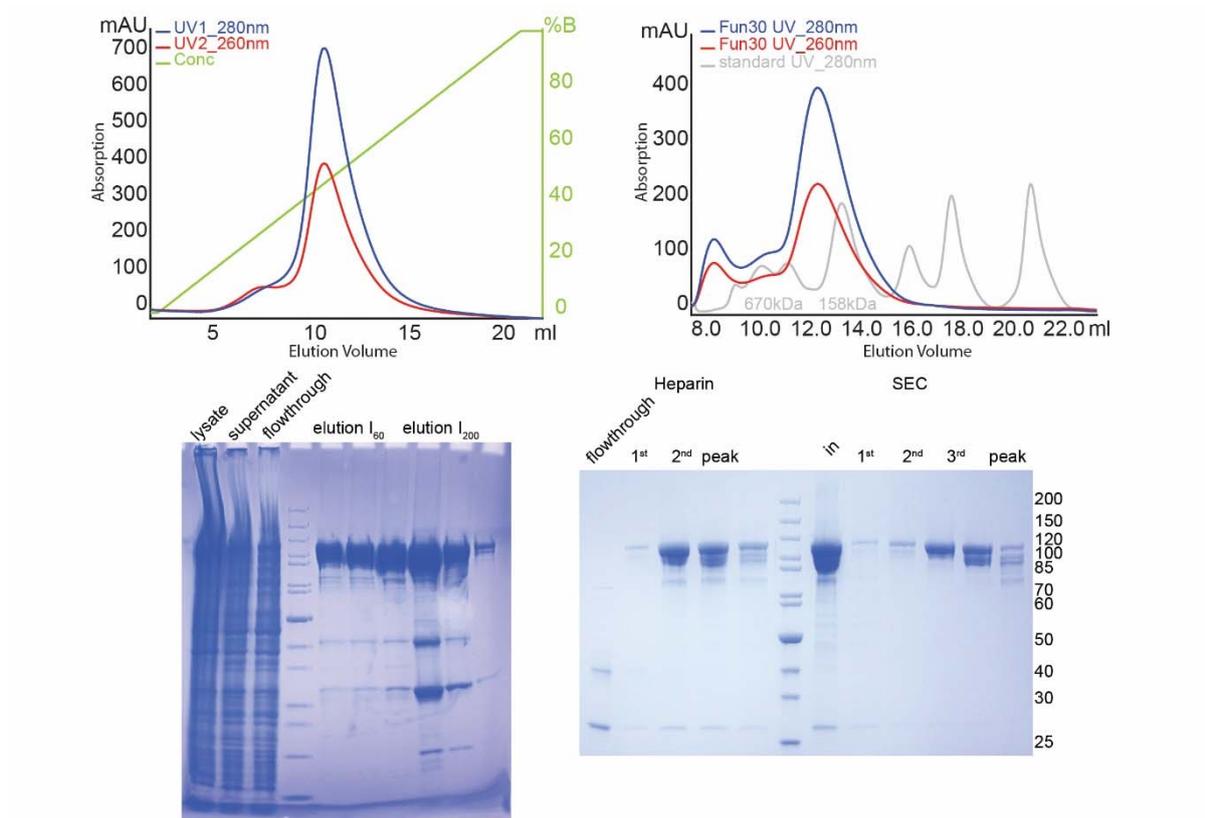
Fig.65: Purification of fl ctFun30 with Ni elution, Heparin affinity purification and size exclusion chromatography profile with corresponding SDS-PAGEs

Fl ctFun30 elutes from the size exclusion chromatography in two peaks, corresponding to around 600 kDa and 200 kDa according to the standard. Increased elution volume can either indicate an elongated shape of the protein, or the formation of multimers (Fig.65).

Mass-determination of fl ctFun30 by size exclusion coupled Right-Angle Light Scattering (RALS)

To determine the accurate mass of the eluted protein in the size exclusion chromatography profile, right angle scattering was performed on the elution of a S200 column. The experiment revealed a molecular mass of 226 kDa in the first peak and 113 kDa in the second peak. (Fig.66)

As in SDS-page analysis both peaks were shown to be fl ctFun30, this indicates the formation of monomer and dimer during expression and purification.
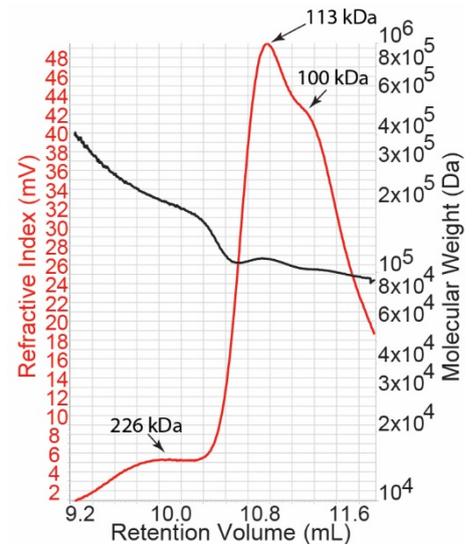
Fig.66: RALS of fl ctFun30

Thermal shift assays

With Thermal shift assays the stability of fl ctFun30 was tested in different buffer conditions.

A1: water
A7: citic Acid pH6
D12: Bicine pH9
G12: 250mM Tris pH8

H5: 50mM Hepes pH7.5, 750mM NaCl
H6: 50mM Hepes pH7.5, 1M NaCl
H9: 50mM Tris pH8, 250mM NaCl
H11: 50mM Tris pH8, 750mM NaCl
H12: 50mM Tris pH8, 1M NaCl
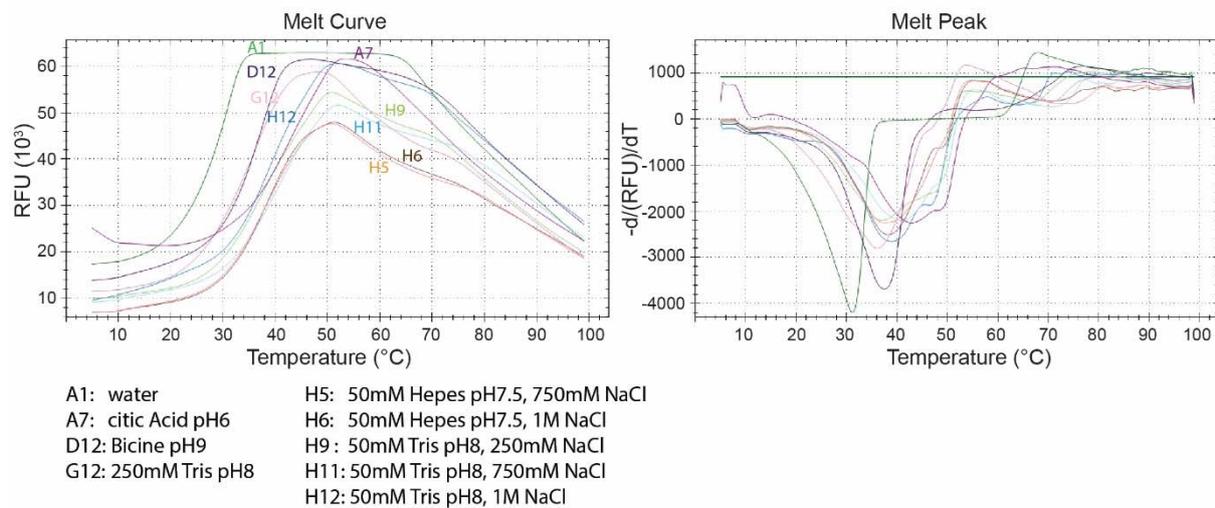
Fig.67: Melting curve of thermal shift assay from fl ctFun30

The results, shown in Fig.67 indicate that high salt conditions are favourable for the stability of the protein. Although Citrate buffer was identified as suitable candidate it was not used for purification, as it is known to be problematic in crystallisation attempts.

Activity assays of fl ctFun30

DNA-binding of fl ctFun30 was analysed by EMSA at a salt concentration of 150 mM. Random DNA sequences at the length of 16, 25 and 50 bp were used, as well as poly 25mer poly dT.

Fig.68: EMSA of fl ctFun30 with 25mer DNA in double and single strand

DNA binding can be clearly detected at concentrations above 800 nM (Fig.68). However, a shift to the pocket might be due to aggregation of sample in low-salt conditions. Binding of all random sequences with similar affinity indicates DNA binding regardless of the sequence.

Limited Proteolysis of fl ctFun30

As some degradation bands of fl ctFun30 appear during purification, limited proteolysis was applied to obtain a suitable construct for crystallization. Bioinformatic analysis predicted a mainly unstructured N-terminus of the protein, whereas the C-terminal domain includes the ATPase domain, and might be considered as fully structured. Therefore, it was concluded, that degradation bands appearing in the limited
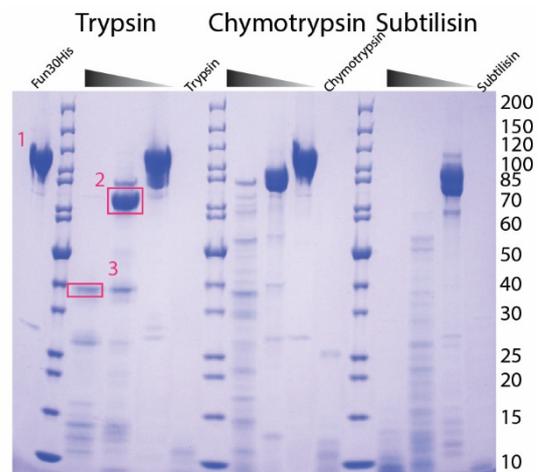
Fig.69: SDS-PAGE of limited proteolysis of fl ctFun30

proteolysis might most likely be due to N-terminal degradation. The bands marked with red boxes in Fig.69 were excised and send for N-terminal. They correspond to constructs beginning at amino acid 333 and 663. The corresponding amino acids are marked in the following disordered prediction calculated with RONN 3.2 [151] displayed in Fig.70:
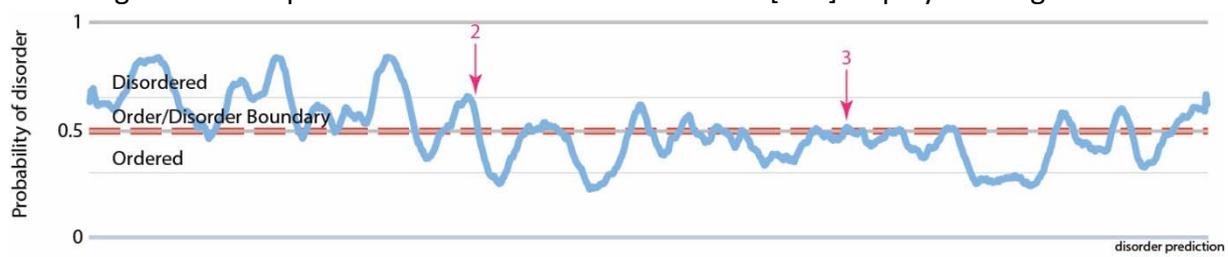
Fig.70: Disordered prediction of fl ctFun30 using RONN 3.2

As amino acid 663 is situated directly in between the N- and the C-lobe of the ATPase, this construct was not considered for Crystallization.

Purification of ctFun30_333-978

For purification of the longer construct yielded by N-terminal sequencing, Ni-NTA affinity purification and Gel filtration after tag removal was applied. The construct yielded high amounts of decently pure protein (Fig.71), suitable for extensive crystal screening.
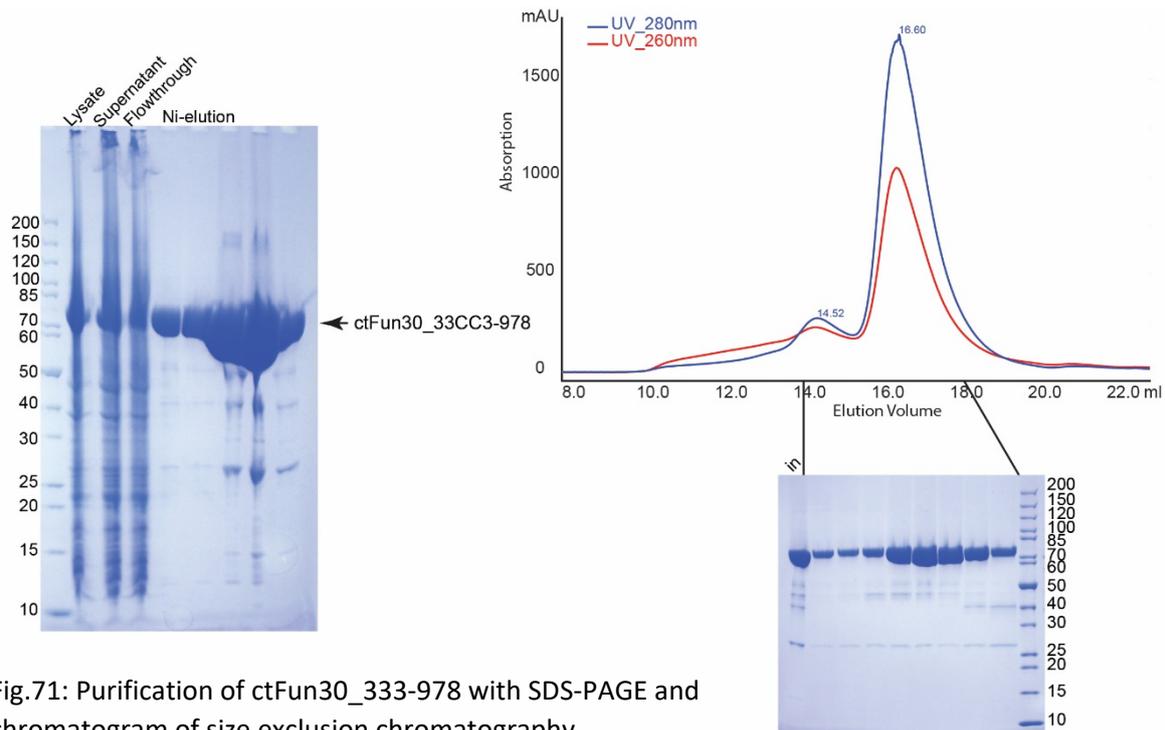
Fig.71: Purification of ctFun30_333-978 with SDS-PAGE and chromatogram of size exclusion chromatography

Mass determination of ctFun30_333-978 by size exclusion coupled Right-Angle Light Scattering

Right angle static light scattering experiments were performed on ctFun30_333-978 to analyse Dimer formation.

The experiment shows both monomer and dimer to be present in the sample (Fig.72). As the sample was derived from only the first peak of a previous size exclusion chromatography, this clearly indicates an equilibrium formation between monomer and dimer.
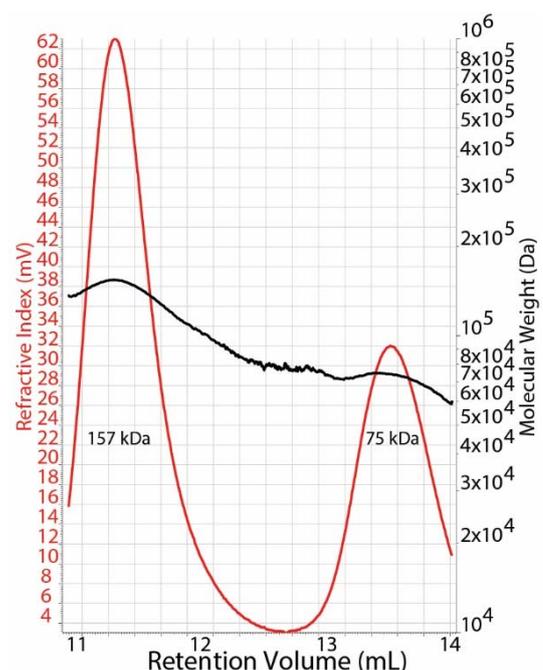
Fig.72: RALS of ctFun30_333-978

Stability assays of ctFun30_333-978

Thermal shift assays, displayed in Fig.73 showed that the purification condition used for ctFun30_333-978 (red) is amongst the best-suited conditions tested.

pink: 50mM Tris pH8 1M NaCl
red: 50mM Hepes pH7.51M NaCl
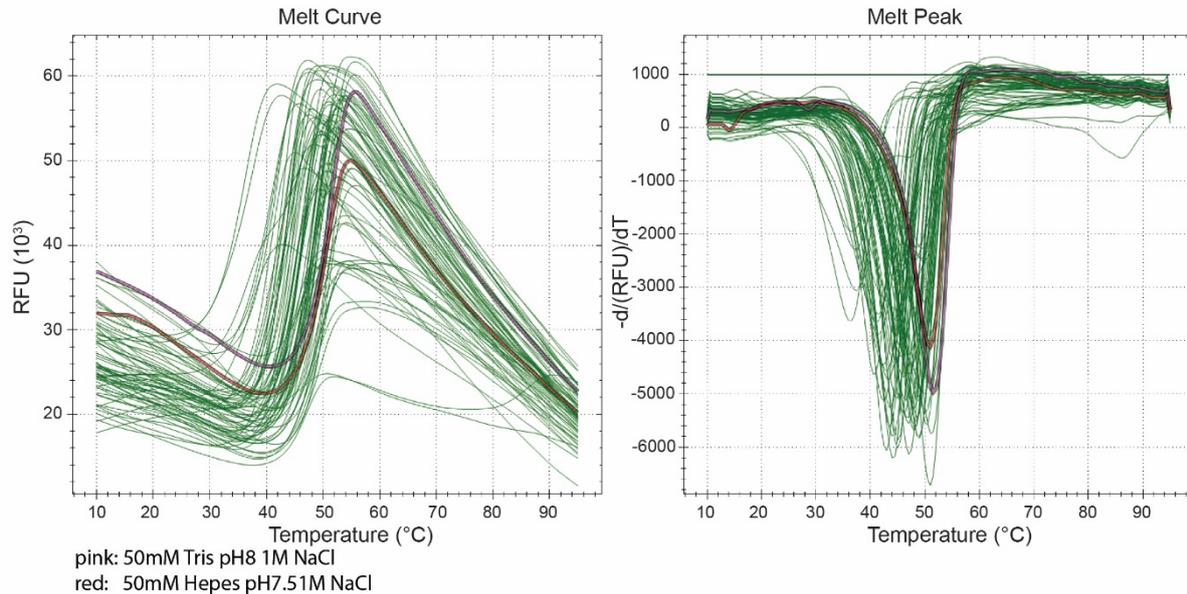
Fig.73: Thermal shift assay of ctFun30_333-978

Activity assays of ctFun30_333-978

To test the activity of the ctFun30_333-978 construct DNA and nucleosome-binding assays were performed, as well as ATPase assays.

DNA-binding assays were performed using a double stranded 25mer DNA with a sequence of TTTTTTCGTCTTCGGCAATTTTTT at a salt concentration of 150mM. DNA binding could be detected starting at a concentration of 400 nM protein (Fig.74 left).
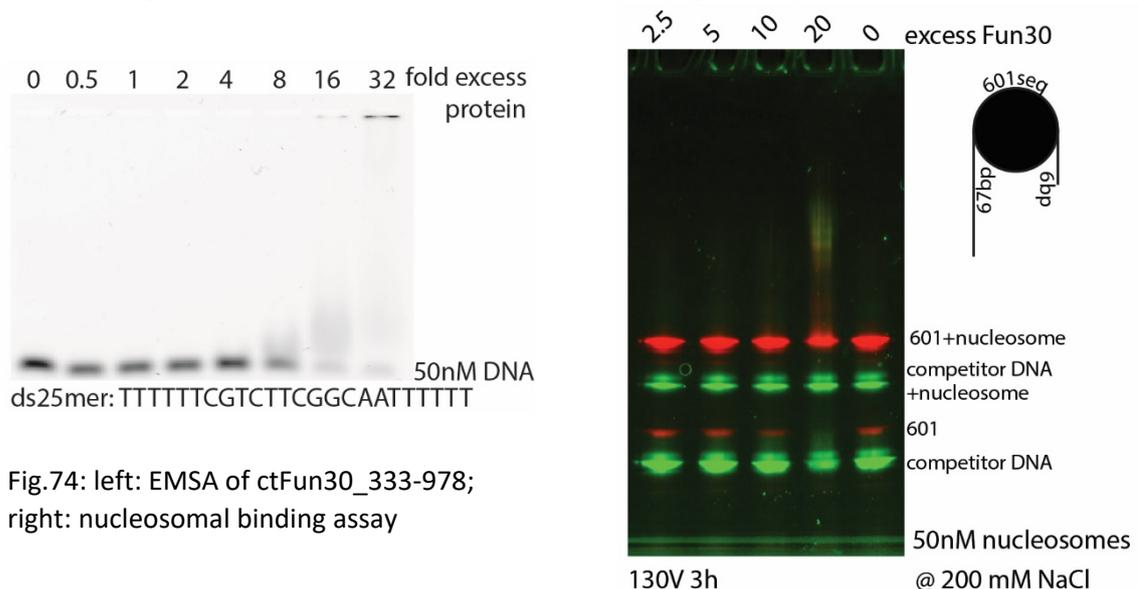
Fig.74: left: EMSA of ctFun30_333-978;
right: nucleosomal binding assay

Nucleosome-binding assays, performed at 200 mM salt indicated only very weak binding to nucleosomes and a preferred binding to remaining free DNA (Fig.74 right).

DNA-dependent ATPase activity of ctFun30_333-978 was tested by a NADH-coupled ATPase assay. The ATPase activity was measured in the absence and presence of the 25mer double stranded DNA used in the EMSA (Fig.75).

slope:
buffer   : -1.086E-5
Fun30   : -1.721E-5   Δ: -6.348E-6   ->3.314E-6 mM ATP/s   -> 0.03 ATP/s/enzyme
ds25bpDNA: -1.284E-5
Fun30+DNA: -5.858E-3   Δ: -4.574E-3   ->2.388E-5 mM ATP/s   -> 0.19 ATP/s/enzyme
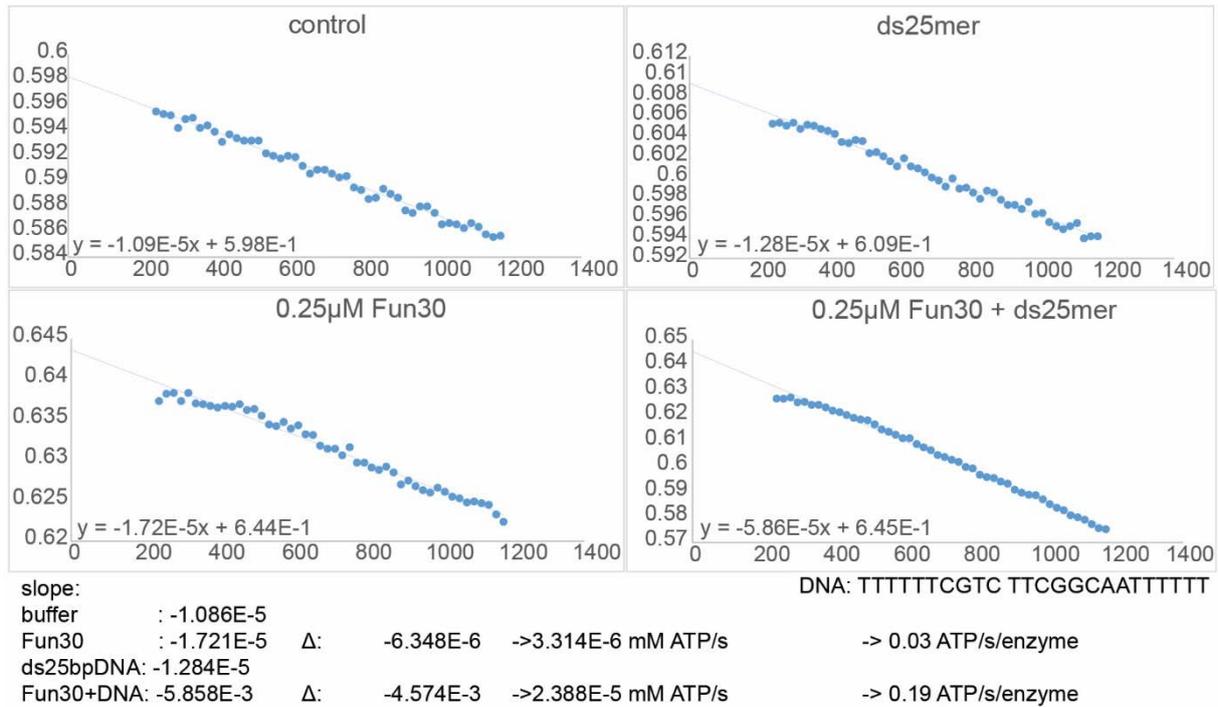
Fig.75: left: ATPase assay of ctFun30_333-978 showing DNA dependent ATPase activation

The ATPase assay revealed an ATPase activity of 0.03 ATP/s/enzyme in the absence of DNA and 0.19 ATP/s/enzyme, showing a more than 6 fold stimulation of ATPase activity by the presence of DNA.

Small-Angle X-ray Scattering of ctFun30_333-978 in the presence and absence of ATPγS

Small-Angle X-ray Scattering was performed with ctFun30_333-978 to test for conformational changes induced by the presence of ATPγS (Fig.76).
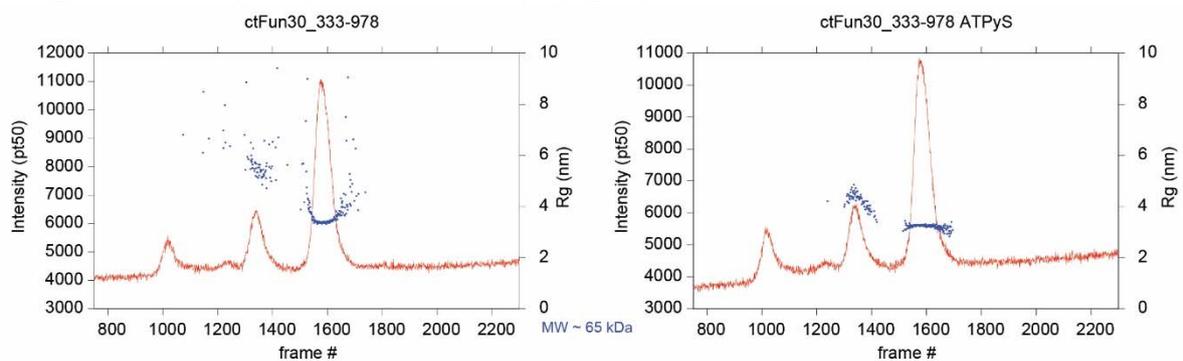
Fig.76: Small-Angle X-ray Scattering in size exclusion chromatography of Fun30_333-978 with (right) and without ATPγS (left)

93

In the scattering plot, as well as in the Kratky plot, no difference between the apo form and the ATPγS bound form are visible, indicating that no major conformational change in the molecule is induced upon binding (Fig.77).
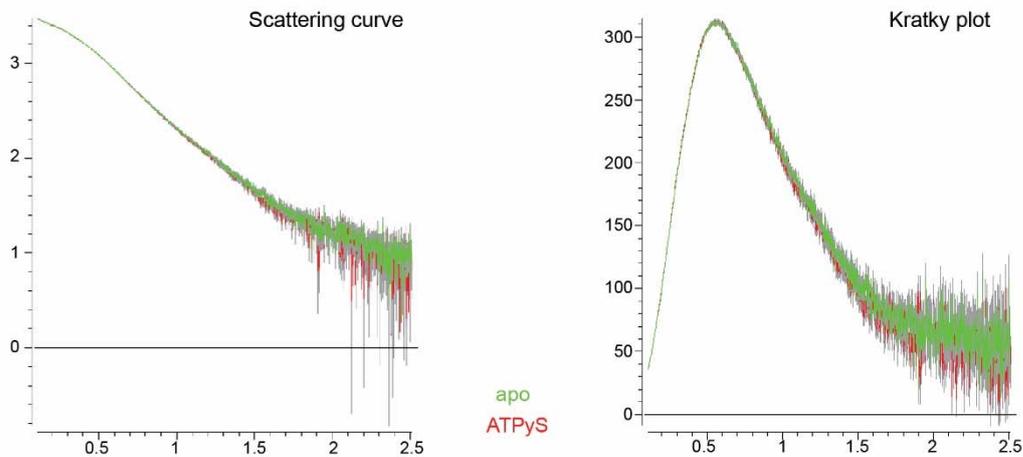
Fig.77: Scattering curve and Kratky plot from SAXS measurement with (red) and without (green) ATPγS

Crystallization of ctFun30_333-978

ctFun30_333-978 was tested extensively in crystal-screening. Several commercial screens were tested at different concentrations and with different supplementations, such as ATP, ATP-analogues and DNA constructs. Finally, crystalline aggregate in the shape of very small needles was detected in 1.4M Sodium citrate and 100 mM Hepes pH 7.5 at a concentration of about 10 mg/mL with 250mM salt supplementation in the reservoir. Intensive refinement, including seeding, additive screening and addition of ATP-analogues yielded needle formation (Fig.78). However, no diffraction pattern was obtained at the synchrotron.

Fig.78: Crystals of ctFun30_333-

Purification of further ctFun30 constructs

As no diffraction pattern could be obtained after intensive refinement of the ctFun30_333-978 constructs, further truncations of ctFun30 were tested in crystallization. In addition to the truncation of the protein, fusion to T4 Lysozyme and the C-terminus of T4 Lysozyme to ctFun30_333-978 were tested in crystallization. A schematic representation of the constructs tested is shown in Fig.79. Constructs depicted in orange were cloned but not continued. Despite extensive crystal screening, no suitable hits were found.
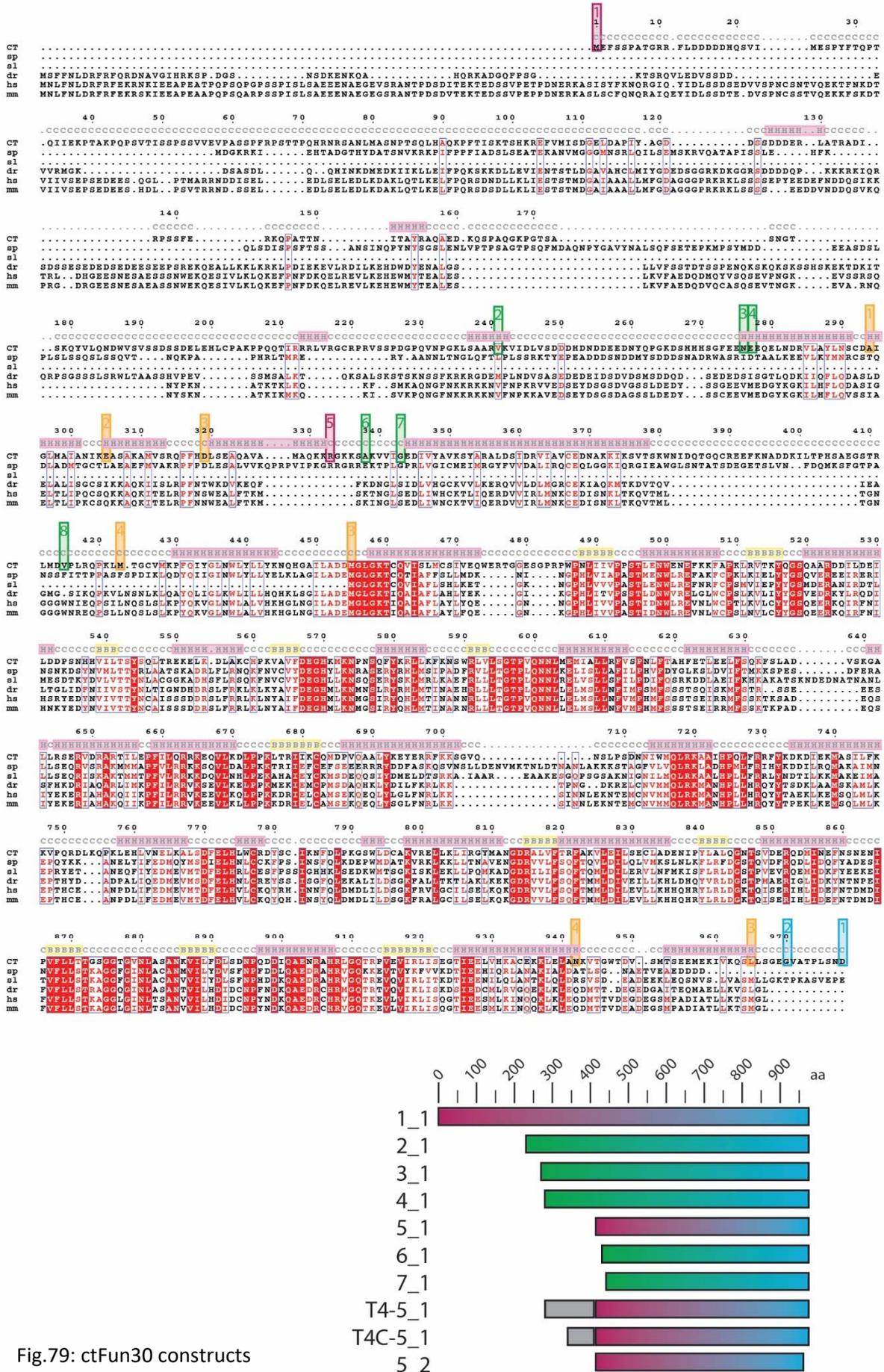
Fig.79: ctFun30 constructs
designed for crystallization trials

Fun30 constructs from different organisms

Human SMRCD1a&c

Due to sequence similarities between ctFun30_333-978 and the skin specific isoform of human SMRCD1c, human SMRCD1a and c were tested in purification. The corresponding molecular weight is 120 kDa and 71 kDa respectively. Expression in *Escherichia coli* showed low stability and a high degree of degradation or contamination (Fig.80 left).
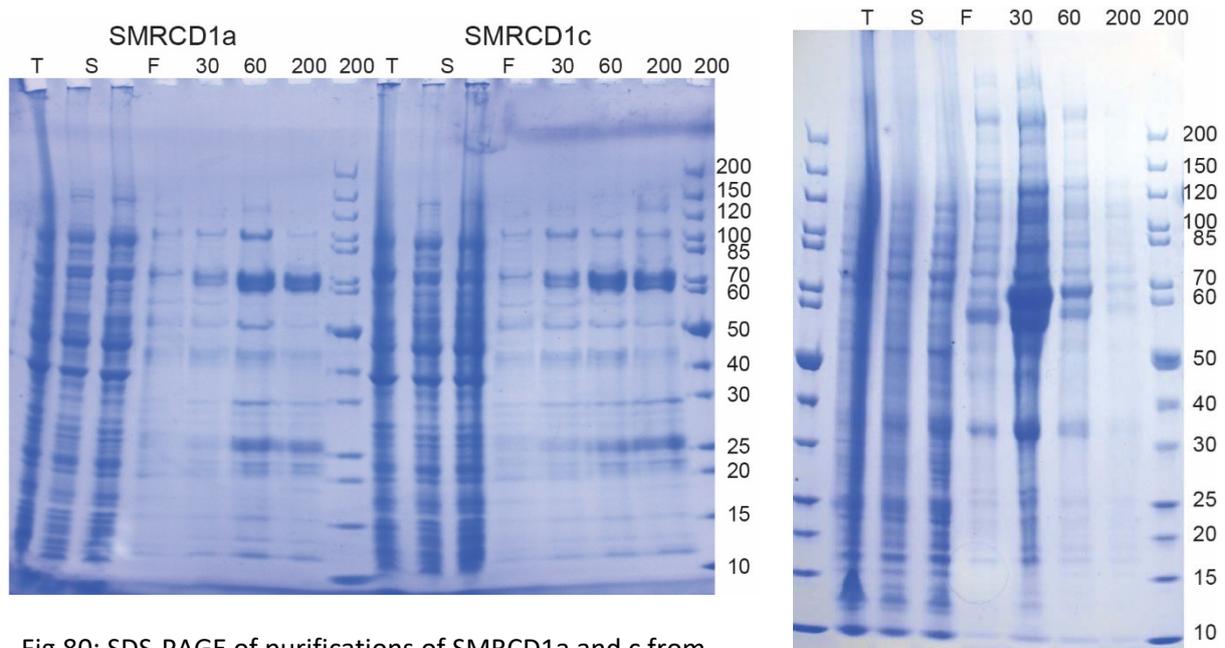
Fig.80: SDS-PAGE of purifications of SMRCD1a and c from *Escherichia coli* (left) and SMRCD1c from bacculovirus infected insect cells (right)

To improve expression level and stability, insect cell expression was tested for SMRCD1c (Fig. 80 right), however, no improvement was achieved.

Secondary columns, such as Heparin or ion exchange were able to reduce the amount of contamination to a certain amount, but due to low-salt sensibility of the protein the amounts of protein obtained after purification were too little for crystal screening.

Schizosaccharomyces pombe Fft3

Fft3 from *Schizosccharomyces pombe* was expressed as truncated construct from amino acid 336 fused to a Sumo tag, as Sumo protease allows tag removal without remaining residues. However, the construct was poorly expressed and heavily degraded, making it an unsuitable construct for crystallization (Fig.81).
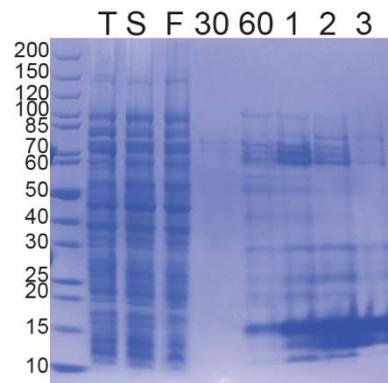
Fig.81: SDS-PAGE of purifications of Fft3 and c

Myceliophtora thermophilum Fun30

Due to high sequence similarities (Fig.82), a construct from *Myceliophtora thermophilum* was tested in crystallography.
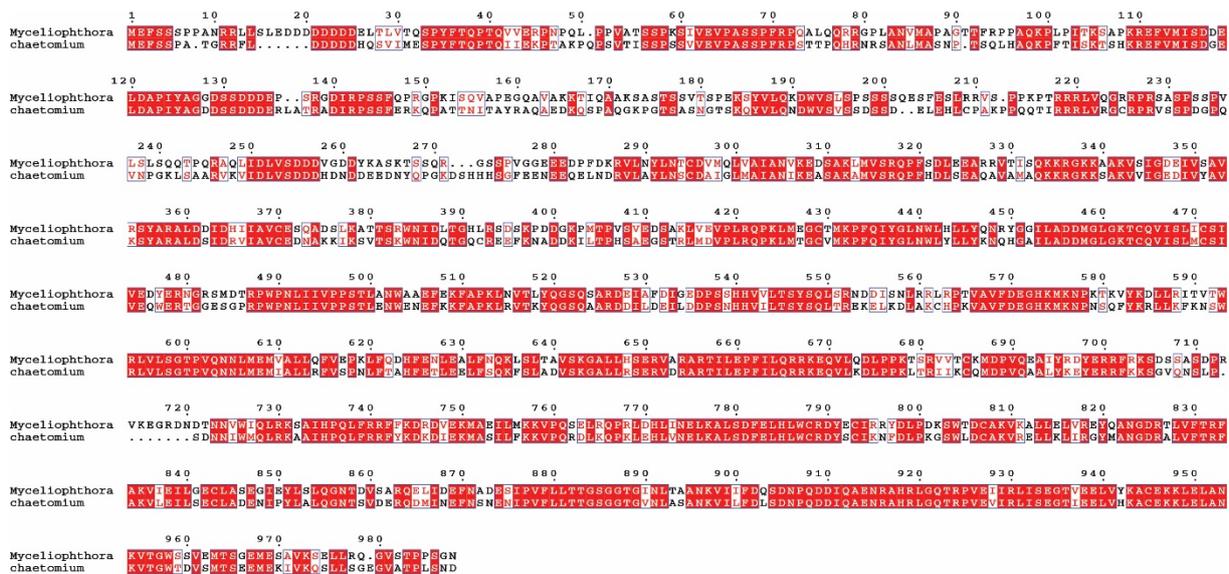
Fig.82: Sequence alignment of *Myceliophtora thermophilum* and Chaetomium thermophilum

A Sumo-tagged construct ranging from amino acid 336 to 988 of *Myceliophtora thermophilum* Fun30 was expressed in *Escherichia coli*. Purification worked comparably well to ctFun30_333-978 (Fig.83). Despite good yield and purity, crystal-screening yielded no suitable hits.
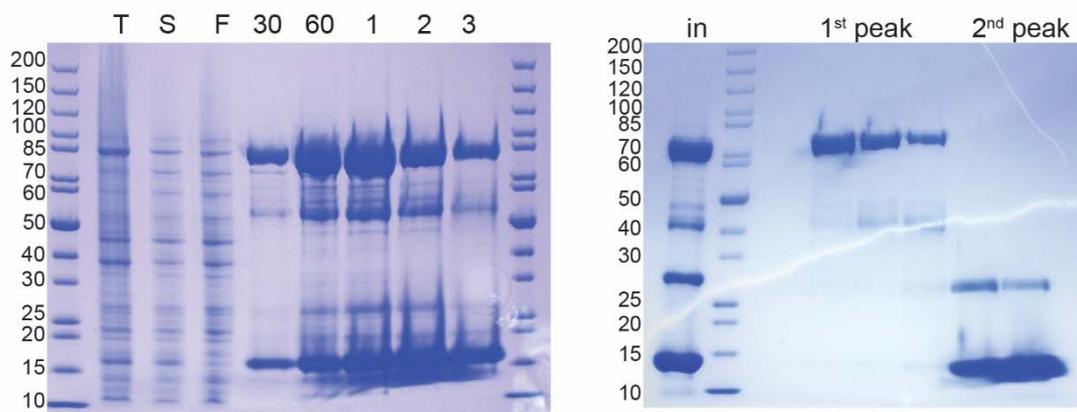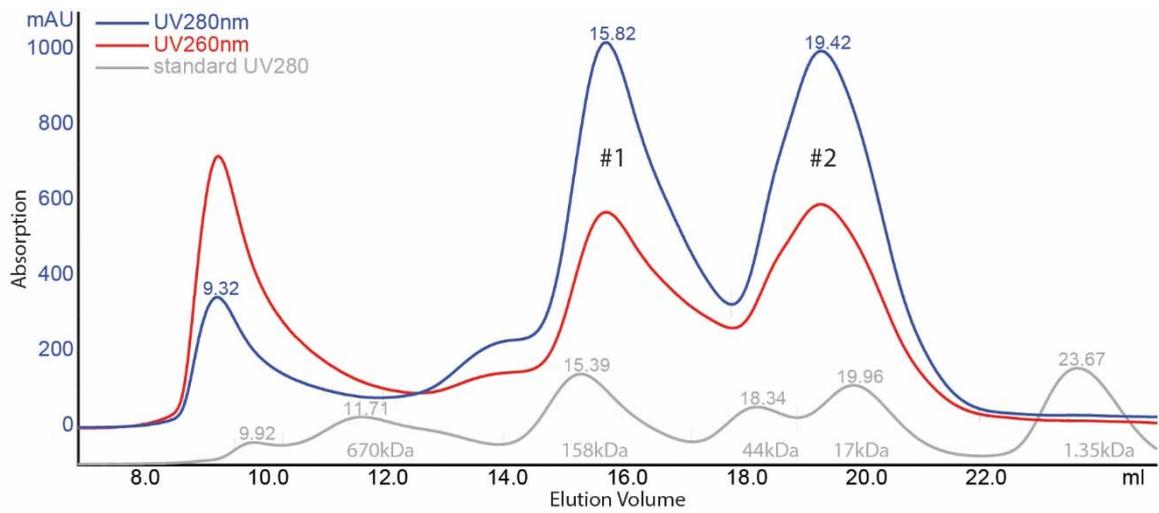
Fig.83: Purification of *Myceliophtora thermophilum* Fun30_336-988

2.4. Discussion

2.3.1. The Nhp10 submodule of INO80

<u>Crystallization attempts of the Nhp10 submodule</u>

Initial purification studies of individual proteins of the Nhp10 submodule performed in the lab revealed Ies1 to be either not expressed or not soluble in E.coli and Ies3 to be unsuitable for purification as an individual protein. A purification protocol for the Nhp10 submodule comprising Nhp10, Ies3 and Ies5 was established in the lab based on expression in insect cells. However, extensive testing showed a significant DNA contamination of the complex. To enhance the stability of the complex, allowing the application of stringent purification conditions a construct with addition of Ies1 and a fragment of Ino80p was designed. This construct represents the entire DNA-binding submodule of the Ino80 complex, as shown by [128]. The purification showed heavy degradations of Ino80p and substoichiometric amounts of Ies1 as well as minor DNA contaminations. Further attempts of improving the protocol led to a decrease in yield, rendering the protein unsuitable for crystallization. The application of an entirely new approach by implicating the step of the cytoplasmic lysis of insect cells significantly improved the quality of the obtained sample. Sadly this approach yielded protein in amounts insufficient for extensive large scale crystallization and only allowed initial screening. To improve the stability a nanobody was used for fishing of a stable construct within an insect cell lysate containing 10 proteins of the Ino80 complex but lacking the Ino80 protein as a scaffold. This revealed a stable complex comprising Nhp10, Ies3 and Ies5 but with no significant increase in yield. Therefore, a reduction of the complex to the minimal stable complex was considered and tested in E.coli. Interaction of Nhp10 and Ies5 alone was shown to be unstable, and only a complex consisting of Nhp10, Ies3 and Ies5 was stably expressed. This approach improved the yield significantly. However, the complex showed heavy degradations in the absence of DNA and nucleotide free purification was only successful in the presence of a stabilizing nanobody. Screens performed with proteins obtained from this protocol showed minor crystalline precipitation and phase separation. Despite refinement trials, no crystals could be obtained. This indicates larger unstructured regions within the complex.

Therefore, either further construct design would be necessary, or the addition of an interaction partner, which might induce further structural organization of the unstructured

regions. The necessity of a nanobody to allow stable expression and purification further supports the hypothesis of a missing interaction partner. As the Nhp10 module is part of the larger Ino80 complex, an interaction partner within the complex might be necessary to stabilize the module. Studies of the Ino80 complex with cryo-EM suggest a close proximity to the Arp module. Hence, components of this module might be suitable for coexpression and crystallization. Furthermore, the addition of another submodule would increase the size of the complex and allow structural studies using single particle cryo-EM. One further suitable interaction partner outside the Ino80 complex might be Reb. Reb1 was analysed by Chip-Exo assays and revealed a similar binding pattern as Nhp10 and other components of the Nhp10 submodule [143]. DNA binding assays and bioinformatics analysis performed in this study show a sequence unspecific binding behaviour of the Nhp10 module, and hence indicate an interaction with Reb1. A direct interaction between Reb1 and the Nhp10 module might be used for stabilizing the Nhp10 module and would therefore be suitable for crystallization trials.

Furthermore, the usage of a mutant deficient in DNA-binding might reduce the loss of protein through DNA contaminations. An increased yield in purification would allow a broader crystal screening.

<u>Implications for the Ino80 complex</u>

One of the major functions of the Ino80 complex is the sliding of nucleosomes along DNA. For this action, two interaction sites with DNA are required. A mutation of one of these sites might resemble a locked Ino80 complex on DNA and nucleosomes. This locked state might allow further studies on the interaction mechanism and facilitate structural analysis of the complex in interaction with the nucleosome. With the Nhp10 module being the major DNA binding component of the Ino80 component, it would be an ideal candidate for such studies. According to the bioinformatic analysis of Nhp10, point mutations of Y101 and T98 in the environment of the entire complex would presumably lead to the desired effect of destroying the interaction of the Nhp10 module with DNA.

2.4.2. Fun30 chromatin remodeller

The aim of this project was to investigate in the structure of a Snf2 enzyme in its active state. Mutational studies and superposition with related enzymes point towards a large conformational change of the ATPase lobe of Snf2 enzymes upon binding of ATP. To proof this hypothesis, structural studies of Fun30 in the presence and absence of ATP and nucleotide-analogues were planned.

<u>Biochemical characterization of ctFun30</u>

In the course of this study the purification of ctFun30 was optimized and now shows high yields and pure protein. Its nature as a DNA binding protein makes ctFun30 sensitive to low-salt condition in its DNA-free state. The necessity of high-salt conditions of more than 500 mM sodium chloride created severe problems in purification, as it impeded the application of a secondary column, such as Heparin or ion-exchange chromatography. Therefore, optimization of the purification schematic was necessary. The obtained protein is very well suitable for structural analysis and crystallization attempts, however, biochemical analysis are difficult. Most biochemical activity assays show salt-sensitivity of some components, preventing an application at high salt conditions. This led to difficulties in the execution and evaluation of the assays, as precipitation of Fun30 could not be excluded and would alter the results.

Furthermore, the removal of the N-terminal, presumably unstructured region (Fig.84 calculated using the MPI Bioinformatics Toolkit [149]) improved the stability and led to less degradation visible in the SDS-PAGE. In static light scattering experiments, the Fun30 constructs, examined in this work, show an equilibrium formation between monomeric and dimeric state in the purification conditions. This was the case for full-length, as well as the N-terminal truncated construct of ctFun30. This might imply a regulatory role for the region directly at the N-terminus of the ATPase region. Secondary structure predictions of this region reveal a helical pattern, indicating a structural organization beyond the ATPase domain. A functional role for this region is not known, however the further N-terminal region can be excluded as dimerization region, as the formation of the Dimer is not prevented upon truncation. Furthermore, previous structures of Snf2 family enzymes show activity of the enzymes in the monomeric state, making the formation of the Dimer in the ATPase domain of Fun30 unlikely.

Fig.84: Secondary structure prediction of ctFun30 using Quick2D from the MPI Bioinformatics Toolkit

DNA-binding assays reveal no sequence specificity of the enzyme, as expected. However, binding to nucleosomes can only be detected at very high concentration. Additionally, a preference for free or extranucleosomal DNA can be shown in nucleosome shift assays. This might indicated a necessity of activation or regulation for Fun30, as binding to nucleosomes seems essential for the functioning of a chromatin remodeller. One possibility of regulation for Fun30 might lay in the phosphorylation sites of Fun30 in the very N-terminal tail. Phosphorylation via CDK1 might increase the affinity for nucleosomes. However, it cannot be excluded, that the affinity of Fun30 for nucleosomes is affected by the truncation of the N-terminal tail itself. The activity of the ATPase domain, on the other hand, is regulated by DNA-binding of Fun30 and seems fully functional in the truncated construct. As previously reported for Snf2 enzymes, the ATPase activity of Fun30 is highly dependent on the presence of DNA and a 6-fold increase can be seen upon addition of DNA. A specificity for certain DNAs can be seen neither in DNA binding nor in the ATPase assay. The addition of ATP or ATP analogues does not show any effect in the structural conformation of Fun30, as no major conformational changes upon nucleotide addition can be detected in Small-angle X-ray Scattering. This shows, that the supplementation of ATP alone, does not induce a rotation of the second RecA lobe.

Structural characterization of ctFun30

Due to high yield and good stability ctFun30_333-978 is presumably the best construct for structural analysis, together with mtFun30-336-988. However, extensive screening remained unsuccessful. In one condition, crystalline needles were detected, which could be improved

upon seeding and refinement. Despite extensive refinement, the needles were extremely fragile and remained hollow. Therefore, no diffraction pattern could be obtained. In addition, other constructs and a change of organism to *Myceliopthora thermophilus* yielded no crystal hits in extensive screening. Other organisms, which were tested in purification, such as *Homo sapiens* and S*chizossacharomyces pombe*, showed to be unsuitable for crystallization due to low yield and decreased stability in comparison to the thermophilic fungi.

Meanwhile, Liu et al. published the crystal structure of the ATPase-C lobe of Fun30 from *Saccharomyces cerevisiae* at 1.95 Å. This structure revealed the fold of two insertion regions, one being a Snf2-specific helix insertion, and the other one showing a non-conserved helix-bundle. With increased sequence conservation of the second insertion within the Fun30 subfamily, the authors postulate a role for the insertion in specifying the function of Fun30 in distinction from other Snf2 enzymes. Furthermore, the authors report a monomeric state for the C-terminal domain of Fun30. This strengthens the hypothesis of the N-terminal region, close to the N-lobe of the ATPase domain, being responsible for dimer-formation. Therefore, structural information on the N-terminal region adjacent to the ATPase domain and of the dimer formation would still be of interest; however the impact of such a structure is greatly diminished by the presented work of [152]. Furthermore, the equilibrium formation between the dimeric state and the monomeric state are most likely impeding crystal formation. Therefore, further biochemical assays would be required to investigate in regulation and dimerization of Fun30. However, an organism, allowing purification and analysis is low salt condition would be required.

3. References

1. Arkowitz, R.A., *Chemical gradients and chemotropism in yeast.* Cold Spring Harb Perspect Biol, 2009. **1**(2): p. a001958.
2. Koolman, J., K.H. Röhm, and J. Wirth, *Taschenatlas der Biochemie*. 2003: Thieme.
3. Green, J., et al., *Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: adaptation for different ecological niches.* Curr Opin Microbiol, 2014. **18**: p. 1-7.
4. Gomelsky, M., *cAMP, c-di-GMP, c-di-AMP and now cGMP: bacteria use them all!* Mol Microbiol, 2011. **79**(3): p. 562-5.
5. Pesavento, C. and R. Hengge, *Bacterial nucleotide-based second messengers.* Curr Opin Microbiol, 2009. **12**(2): p. 170-6.
6. Dalebroux, Z.D. and M.S. Swanson, *ppGpp: magic beyond RNA polymerase.* Nat Rev Microbiol, 2012. **10**(3): p. 203-12.
7. Liu, K., A.N. Bittner, and J.D. Wang, *Diversity in (p)ppGpp metabolism and effectors.* Curr Opin Microbiol, 2015. **24**: p. 72-9.
8. Atkinson, G.C., T. Tenson, and V. Hauryliuk, *The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life.* PLoS One, 2011. **6**(8): p. e23479.
9. Steinchen, W. and G. Bange, *The magic dance of the alarmones (p)ppGpp.* Mol Microbiol, 2016. **101**(4): p. 531-44.
10. Wang, J.D., G.M. Sanders, and A.D. Grossman, *Nutritional control of elongation of DNA replication by (p)ppGpp.* Cell, 2007. **128**(5): p. 865-75.
11. Maciag, M., et al., *ppGpp inhibits the activity of Escherichia coli DnaG primase.* Plasmid, 2010. **63**(1): p. 61-7.
12. Maciag-Dorszynska, M., A. Szalewska-Palasz, and G. Wegrzyn, *Different effects of ppGpp on Escherichia coli DNA replication in vivo and in vitro.* FEBS Open Bio, 2013. **3**: p. 161-4.
13. Hauryliuk, V., et al., *Recent functional insights into the role of (p)ppGpp in bacterial physiology.* Nat Rev Microbiol, 2015. **13**(5): p. 298-309.
14. Takahashi, K., K. Kasai, and K. Ochi, *Identification of the bacterial alarmone guanosine 5'-diphosphate 3'-diphosphate (ppGpp) in plants.* Proc Natl Acad Sci U S A, 2004. **101**(12): p. 4320-4.
15. Martinez-Costa, O.H., et al., *A relA/spoT homologous gene from Streptomyces coelicolor A3(2) controls antibiotic biosynthetic genes.* J Biol Chem, 1996. **271**(18): p. 10627-34.
16. Moris, M., et al., *Effective symbiosis between Rhizobium etli and Phaseolus vulgaris requires the alarmone ppGpp.* J Bacteriol, 2005. **187**(15): p. 5460-9.
17. Bai, Y., et al., *Two DHH subfamily 1 proteins in Streptococcus pneumoniae possess cyclic di-AMP phosphodiesterase activity and affect bacterial growth and virulence.* J Bacteriol, 2013. **195**(22): p. 5123-32.
18. Mehne, F.M., et al., *Cyclic di-AMP homeostasis in bacillus subtilis: both lack and high level accumulation of the nucleotide are detrimental for cell growth.* J Biol Chem, 2013. **288**(3): p. 2004-17.
19. Huynh, T.N., et al., *An HD-domain phosphodiesterase mediates cooperative hydrolysis of c-di-AMP to affect bacterial growth and virulence.* Proc Natl Acad Sci U S A, 2015. **112**(7): p. E747-56.
20. Song, J.H., et al., *Identification of essential genes in Streptococcus pneumoniae by allelic replacement mutagenesis.* Mol Cells, 2005. **19**(3): p. 365-74.
21. Chaudhuri, R.R., et al., *Comprehensive identification of essential Staphylococcus aureus genes using Transposon-Mediated Differential Hybridisation (TMDH).* BMC Genomics, 2009. **10**: p. 291.

22.   Woodward, J.J., A.T. Iavarone, and D.A. Portnoy, *c-di-AMP secreted by intracellular Listeria monocytogenes activates a host type I interferon response.* Science, 2010. **328**(5986): p. 1703-5.

23.   Luo, Y. and J.D. Helmann, *Analysis of the role of Bacillus subtilis sigma(M) in beta-lactam resistance reveals an essential role for c-di-AMP in peptidoglycan homeostasis.* Mol Microbiol, 2012. **83**(3): p. 623-39.

24.   Corrigan, R.M., et al., *Cross-talk between two nucleotide-signaling pathways in Staphylococcus aureus.* J Biol Chem, 2015. **290**(9): p. 5826-39.

25.   Manzanillo, P.S., et al., *Mycobacterium tuberculosis activates the DNA-dependent cytosolic surveillance pathway within macrophages.* Cell Host Microbe, 2012. **11**(5): p. 469-80.

26.   St-Onge, R.J., et al., *Nucleotide second messenger-mediated regulation of a muralytic enzyme in Streptomyces.* Mol Microbiol, 2015. **96**(4): p. 779-95.

27.   Manikandan, K., et al., *Two-step synthesis and hydrolysis of cyclic di-AMP in Mycobacterium tuberculosis.* PLoS One, 2014. **9**(1): p. e86096.

28.   Fahmi, T., G.C. Port, and K.H. Cho, *c-di-AMP: An Essential Molecule in the Signaling Pathways that Regulate the Viability and Virulence of Gram-Positive Bacteria.* Genes (Basel), 2017. **8**(8).

29.   Gundlach, J., et al., *Control of potassium homeostasis is an essential function of the second messenger cyclic di-AMP in Bacillus subtilis.* Sci Signal, 2017. **10**(475).

30.   Glass, J.I., et al., *Essential genes of a minimal bacterium.* Proc Natl Acad Sci U S A, 2006. **103**(2): p. 425-30.

31.   French, C.T., et al., *Large-scale transposon mutagenesis of Mycoplasma pulmonis.* Mol Microbiol, 2008. **69**(1): p. 67-76.

32.   Barker, J.R., et al., *STING-dependent recognition of cyclic di-AMP mediates type I interferon responses during Chlamydia trachomatis infection.* MBio, 2013. **4**(3): p. e00018-13.

33.   Lluch-Senar, M., et al., *Defining a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium.* Mol Syst Biol, 2015. **11**(1): p. 780.

34.   Commichau, F.M., et al., *A jack of all trades: the multiple roles of the unique essential second messenger cyclic di-AMP.* Mol Microbiol, 2015. **97**(2): p. 189-204.

35.   Bai, Y., et al., *Mycobacterium tuberculosis Rv3586 (DacA) is a diadenylate cyclase that converts ATP or ADP into c-di-AMP.* PLoS One, 2012. **7**(4): p. e35206.

36.   Corrigan, R.M., et al., *Systematic identification of conserved bacterial c-di-AMP receptor proteins.* Proc Natl Acad Sci U S A, 2013. **110**(22): p. 9084-9.

37.   Witte, G., et al., *Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates.* Mol Cell, 2008. **30**(2): p. 167-78.

38.   Rosenberg, J., et al., *Structural and biochemical analysis of the essential diadenylate cyclase CdaA from Listeria monocytogenes.* J Biol Chem, 2015. **290**(10): p. 6596-606.

39.   Nicolas, P., et al., *Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis.* Science, 2012. **335**(6072): p. 1103-6.

40.   Mehne, F.M., et al., *Control of the diadenylate cyclase CdaS in Bacillus subtilis: an autoinhibitory domain limits cyclic di-AMP production.* J Biol Chem, 2014. **289**(30): p. 21098-107.

41.   Corrigan, R.M., et al., *c-di-AMP is a new second messenger in Staphylococcus aureus with a role in controlling cell size and envelope stress.* PLoS Pathog, 2011. **7**(9): p. e1002217.

42.   Dengler, V., et al., *Mutation in the C-di-AMP cyclase dacA affects fitness and resistance of methicillin resistant Staphylococcus aureus.* PLoS One, 2013. **8**(8): p. e73512.

43.   Gandara, C. and J.C. Alonso, *DisA and c-di-AMP act at the intersection between DNA-damage response and stress homeostasis in exponentially growing Bacillus subtilis cells.* DNA Repair (Amst), 2015. **27**: p. 1-8.

44.   Bejerano-Sagie, M., et al., *A checkpoint protein that scans the chromosome for damage at the start of sporulation in Bacillus subtilis.* Cell, 2006. **125**(4): p. 679-90.

45.   Oppenheimer-Shaanan, Y., et al., *c-di-AMP reports DNA integrity during sporulation in Bacillus subtilis.* EMBO Rep, 2011. **12**(6): p. 594-601.

46.   Campos, S.S., et al., *Interaction of apurinic/apyrimidinic endonucleases Nfo and ExoA with the DNA integrity scanning protein DisA in the processing of oxidative DNA damage during Bacillus subtilis spore outgrowth.* J Bacteriol, 2014. **196**(3): p. 568-78.

47.   Zhang, L. and Z.G. He, *Radiation-sensitive gene A (RadA) targets DisA, DNA integrity scanning protein A, to negatively affect cyclic Di-AMP synthesis activity in Mycobacterium smegmatis.* J Biol Chem, 2013. **288**(31): p. 22426-36.

48.   Carrasco, B., et al., *Genetic recombination in Bacillus subtilis 168: contribution of Holliday junction processing functions in chromosome segregation.* J Bacteriol, 2004. **186**(17): p. 5557-66.

49.   Marie, L., et al., *Bacterial RadA is a DnaB-type helicase interacting with RecA to promote bidirectional D-loop extension.* Nat Commun, 2017. **8**: p. 15638.

50.   Zhang, L., W. Li, and Z.G. He, *DarR, a TetR-like transcriptional factor, is a cyclic di-AMP-responsive repressor in Mycobacterium smegmatis.* J Biol Chem, 2013. **288**(5): p. 3085-96.

51.   Vieira-Pires, R.S., A. Szollosi, and J.H. Morais-Cabral, *The structure of the KtrAB potassium transporter.* Nature, 2013. **496**(7445): p. 323-8.

52.   Bai, Y., et al., *Cyclic di-AMP impairs potassium uptake mediated by a cyclic di-AMP binding protein in Streptococcus pneumoniae.* J Bacteriol, 2014. **196**(3): p. 614-23.

53.   Schar, J., et al., *Pyruvate carboxylase plays a crucial role in carbon metabolism of extra- and intracellularly replicating Listeria monocytogenes.* J Bacteriol, 2010. **192**(7): p. 1774-84.

54.   Sureka, K., et al., *The cyclic dinucleotide c-di-AMP is an allosteric regulator of metabolic enzyme function.* Cell, 2014. **158**(6): p. 1389-1401.

55.   Gundlach, J., et al., *Identification, characterization, and structure analysis of the cyclic di-AMP-binding PII-like signal transduction protein DarA.* J Biol Chem, 2015. **290**(5): p. 3069-80.

56.   Campeotto, I., et al., *Complex structure and biochemical characterization of the Staphylococcus aureus cyclic diadenylate monophosphate (c-di-AMP)-binding protein PstA, the founding member of a new signal transduction protein family.* J Biol Chem, 2015. **290**(5): p. 2888-901.

57.   Muller, M., K.P. Hopfner, and G. Witte, *c-di-AMP recognition by Staphylococcus aureus PstA.* FEBS Lett, 2015. **589**(1): p. 45-51.

58.   Choi, P.H., et al., *Molecular basis for the recognition of cyclic-di-AMP by PstA, a PII-like signal transduction protein.* Microbiologyopen, 2015. **4**(3): p. 361-74.

59.   Zeth, K., O. Fokina, and K. Forchhammer, *An engineered PII protein variant that senses a novel ligand: atomic resolution structure of the complex with citrate.* Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 8): p. 901-8.

60.   Ereno-Orbea, J., I. Oyenarte, and L.A. Martinez-Cruz, *CBS domains: Ligand binding sites and conformational variability.* Arch Biochem Biophys, 2013. **540**(1-2): p. 70-81.

61.   Baykov, A.A., H.K. Tuominen, and R. Lahti, *The CBS domain: a protein module with an emerging prominent role in regulation.* ACS Chem Biol, 2011. **6**(11): p. 1156-63.

62.   Watson, P.Y. and M.J. Fedor, *The ydaO motif is an ATP-sensing riboswitch in Bacillus subtilis.* Nat Chem Biol, 2012. **8**(12): p. 963-5.

63.   Nelson, J.W., et al., *Riboswitches in eubacteria sense the second messenger c-di-AMP.* Nat Chem Biol, 2013. **9**(12): p. 834-9.

64.   Gao, A. and A. Serganov, *Structural insights into recognition of c-di-AMP by the ydaO riboswitch.* Nat Chem Biol, 2014. **10**(9): p. 787-92.

65.   Jones, C.P. and A.R. Ferre-D'Amare, *Crystal structure of a c-di-AMP riboswitch reveals an internally pseudo-dimeric RNA.* EMBO J, 2014. **33**(22): p. 2692-703.

66.   Ren, A. and D.J. Patel, *c-di-AMP binds the ydaO riboswitch in two pseudo-symmetry-related pockets.* Nat Chem Biol, 2014. **10**(9): p. 780-6.

67.   Block, K.F., M.C. Hammond, and R.R. Breaker, *Evidence for widespread gene control function by the ydaO riboswitch candidate.* J Bacteriol, 2010. **192**(15): p. 3983-9.

68. Kaplan Zeevi, M., et al., *Listeria monocytogenes multidrug resistance transporters and cyclic di-AMP, which contribute to type I interferon induction, play a role in cell wall stress.* J Bacteriol, 2013. **195**(23): p. 5250-61.

69. Tadmor, K., et al., *Listeria monocytogenes MDR transporters are involved in LTA synthesis and triggering of innate immunity during infection.* Front Cell Infect Microbiol, 2014. **4**: p. 16.

70. Ye, M., et al., *DhhP, a cyclic di-AMP phosphodiesterase of Borrelia burgdorferi, is essential for cell growth and virulence.* Infect Immun, 2014. **82**(5): p. 1840-9.

71. Rao, F., et al., *YybT is a signaling protein that contains a cyclic dinucleotide phosphodiesterase domain and a GGDEF domain with ATPase activity.* J Biol Chem, 2010. **285**(1): p. 473-82.

72. Cho, K.H. and S.O. Kang, *Streptococcus pyogenes c-di-AMP phosphodiesterase, GdpP, influences SpeB processing and virulence.* PLoS One, 2013. **8**(7): p. e69425.

73. Taylor, B.L. and I.B. Zhulin, *PAS domains: internal sensors of oxygen, redox potential, and light.* Microbiol Mol Biol Rev, 1999. **63**(2): p. 479-506.

74. Rao, F., et al., *Unusual heme-binding PAS domain from YybT family proteins.* J Bacteriol, 2011. **193**(7): p. 1543-51.

75. Tan, E., et al., *Solution structure of the PAS domain of a thermophilic YybT protein homolog reveals a potential ligand-binding site.* J Biol Chem, 2013. **288**(17): p. 11949-59.

76. Rallu, F., et al., *Acid- and multistress-resistant mutants of Lactococcus lactis : identification of intracellular stress signals.* Mol Microbiol, 2000. **35**(3): p. 517-28.

77. Aravind, L. and E.V. Koonin, *The HD domain defines a new superfamily of metal-dependent phosphohydrolases.* Trends Biochem Sci, 1998. **23**(12): p. 469-72.

78. Liu, S., et al., *A cold-sensitive Listeria monocytogenes mutant has a transposon insertion in a gene encoding a putative membrane protein and shows altered (p)ppGpp levels.* Appl Environ Microbiol, 2006. **72**(6): p. 3955-9.

79. Muller, M., et al., *Structural analysis of the diadenylate cyclase reaction of DNA-integrity scanning protein A (DisA) and its inhibition by 3'-dATP.* Biochem J, 2015. **469**(3): p. 367-74.

80. Deepthi, A., et al., *Structure of a diguanylate cyclase from Thermotoga maritima: insights into activation, feedback inhibition and thermostability.* PLoS One, 2014. **9**(10): p. e110912.

81. Drexler, D.J., et al., *Structural and Biophysical Analysis of the Soluble DHH/DHHA1-Type Phosphodiesterase TM1595 from Thermotoga maritima.* Structure, 2017. **25**(12): p. 1887-1897 e4.

82. Wang, F., et al., *Structural and biochemical characterization of the catalytic domains of GdpP reveals a unified hydrolysis mechanism for the DHH/DHHA1 phosphodiesterase.* Biochem J, 2018. **475**(1): p. 191-205.

83. Kabsch, W., *Integration, scaling, space-group assignment and post-refinement.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 133-44.

84. Kabsch, W., *Xds.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.

85. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.

86. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 235-42.

87. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.

88. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

89. Zheng, S.Q., et al., *MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy.* Nat Methods, 2017. **14**(4): p. 331-332.

90. Scheres, S.H., *Processing of Structurally Heterogeneous Cryo-EM Data in RELION.* Methods Enzymol, 2016. **579**: p. 125-57.

91. Rohou, A. and N. Grigorieff, *CTFFIND4: Fast and accurate defocus estimation from electron micrographs.* J Struct Biol, 2015. **192**(2): p. 216-21.

92. Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic Acids Res, 2018. **46**(W1): p. W296-W303.

93. Goddard, T.D., C.C. Huang, and T.E. Ferrin, *Visualizing density maps with UCSF Chimera.* J Struct Biol, 2007. **157**(1): p. 281-7.

94. Sehnal, D., et al., *MOLE 2.0: advanced approach for analysis of biomacromolecular channels.* J Cheminform, 2013. **5**(1): p. 39.

95. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nat Protoc, 2015. **10**(6): p. 845-58.

96. Bienert, S., et al., *The SWISS-MODEL Repository-new features and functionality.* Nucleic Acids Res, 2017. **45**(D1): p. D313-D319.

97. Guex, N., M.C. Peitsch, and T. Schwede, *Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective.* Electrophoresis, 2009. **30 Suppl 1**: p. S162-73.

98. Benkert, P., M. Biasini, and T. Schwede, *Toward the estimation of the absolute quality of individual protein structure models.* Bioinformatics, 2011. **27**(3): p. 343-50.

99. Bertoni, M., et al., *Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology.* Sci Rep, 2017. **7**(1): p. 10480.

100. Liu, C., et al., *Insights into biofilm dispersal regulation from the crystal structure of the PAS-GGDEF-EAL region of RbdA from Pseudomonas aeruginosa.* J Bacteriol, 2017.

101. Haber, J.E., *A Life Investigating Pathways That Repair Broken Chromosomes.* Annu Rev Genet, 2016. **50**: p. 1-28.

102. Liu, T. and J. Huang, *Quality control of homologous recombination.* Cell Mol Life Sci, 2014. **71**(19): p. 3779-97.

103. Bunting, S.F., et al., *53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks.* Cell, 2010. **141**(2): p. 243-54.

104. Lazzaro, F., et al., *Histone methyltransferase Dot1 and Rad9 inhibit single-stranded DNA accumulation at DSBs and uncapped telomeres.* EMBO J, 2008. **27**(10): p. 1502-12.

105. Trovesi, C., et al., *Distinct Cdk1 requirements during single-strand annealing, noncrossover, and crossover recombination.* PLoS Genet, 2011. **7**(8): p. e1002263.

106. Hammet, A., et al., *Rad9 BRCT domain interaction with phosphorylated H2AX regulates the G1 checkpoint in budding yeast.* EMBO Rep, 2007. **8**(9): p. 851-7.

107. Bantele, S.C., et al., *Targeting of the Fun30 nucleosome remodeller by the Dpb11 scaffold facilitates cell cycle-regulated DNA end resection.* Elife, 2017. **6**.

108. Chen, X., et al., *The Fun30 nucleosome remodeller promotes resection of DNA double-strand break ends.* Nature, 2012. **489**(7417): p. 576-80.

109. Costelloe, T., et al., *The yeast Fun30 and human SMARCAD1 chromatin remodellers promote DNA end resection.* Nature, 2012. **489**(7417): p. 581-4.

110. Eapen, V.V., et al., *The Saccharomyces cerevisiae chromatin remodeler Fun30 regulates DNA end resection and checkpoint deactivation.* Mol Cell Biol, 2012. **32**(22): p. 4727-40.

111. Chen, X., et al., *Enrichment of Cdk1-cyclins at DNA double-strand breaks stimulates Fun30 phosphorylation and DNA end resection.* Nucleic Acids Res, 2016. **44**(6): p. 2742-53.

112. Lee, J., et al., *Chromatin remodeller Fun30(Fft3) induces nucleosome disassembly to facilitate RNA polymerase II elongation.* Nat Commun, 2017. **8**: p. 14527.

113. Neves-Costa, A., et al., *The SNF2-family member Fun30 promotes gene silencing in heterochromatic loci.* PLoS One, 2009. **4**(12): p. e8111.

114. Awad, S., et al., *The Snf2 homolog Fun30 acts as a homodimeric ATP-dependent chromatin-remodeling enzyme.* J Biol Chem, 2010. **285**(13): p. 9477-84.

115. Agmon, N., et al., *Effect of nuclear architecture on the efficiency of double-strand break repair.* Nat Cell Biol, 2013. **15**(6): p. 694-9.

116. Horigome, C., et al., *PolySUMOylation by Siz2 and Mms21 triggers relocation of DNA breaks to nuclear pores through the Slx5/Slx8 STUbL.* Genes Dev, 2016. **30**(8): p. 931-45.

117. Alatwi, H.E. and J.A. Downs, *Removal of H2A.Z by INO80 promotes homologous recombination.* EMBO Rep, 2015. **16**(8): p. 986-94.

118. Lademann, C.A., et al., *The INO80 Complex Removes H2A.Z to Promote Presynaptic Filament Formation during Homologous Recombination.* Cell Rep, 2017. **19**(7): p. 1294-1303.

119. Watanabe, S., et al., *A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme.* Science, 2013. **340**(6129): p. 195-9.

120. Watanabe, S. and C.L. Peterson, *Response to Comment on "A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme".* Science, 2016. **353**(6297): p. 358.

121. Wang, F., et al., *Comment on "A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme".* Science, 2016. **353**(6297): p. 358.

122. Krietenstein, N., et al., *Genomic Nucleosome Organization Reconstituted with Pure Proteins.* Cell, 2016. **167**(3): p. 709-721 e12.

123. Ebbert, R., A. Birkmann, and H.J. Schuller, *The product of the SNF2/SWI2 paralogue INO80 of Saccharomyces cerevisiae required for efficient expression of various yeast structural genes is part of a high-molecular-weight protein complex.* Mol Microbiol, 1999. **32**(4): p. 741-51.

124. Korber, P. and S. Barbaric, *The yeast PHO5 promoter: from single locus to systems biology of a paradigm for gene regulation through chromatin.* Nucleic Acids Res, 2014. **42**(17): p. 10888-902.

125. Alcid, E.A. and T. Tsukiyama, *ATP-dependent chromatin remodeling shapes the long noncoding RNA landscape.* Genes Dev, 2014. **28**(21): p. 2348-60.

126. Poli, J., S.M. Gasser, and M. Papamichos-Chronakis, *The INO80 remodeller in transcription, replication and repair.* Philos Trans R Soc Lond B Biol Sci, 2017. **372**(1731).

127. Lafon, A., et al., *INO80 Chromatin Remodeler Facilitates Release of RNA Polymerase II from Chromatin for Ubiquitin-Mediated Proteasomal Degradation.* Mol Cell, 2015. **60**(5): p. 784-796.

128. Tosi, A., et al., *Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex.* Cell, 2013. **154**(6): p. 1207-19.

129. Singleton, M.R., M.S. Dillingham, and D.B. Wigley, *Structure and mechanism of helicases and nucleic acid translocases.* Annu Rev Biochem, 2007. **76**: p. 23-50.

130. Chen, L., et al., *Subunit organization of the human INO80 chromatin remodeling complex: an evolutionarily conserved core complex catalyzes ATP-dependent nucleosome remodeling.* J Biol Chem, 2011. **286**(13): p. 11283-9.

131. Chen, L., R.C. Conaway, and J.W. Conaway, *Multiple modes of regulation of the human Ino80 SNF2 ATPase by subunits of the INO80 chromatin-remodeling complex.* Proc Natl Acad Sci U S A, 2013. **110**(51): p. 20497-502.

132. Watanabe, S., et al., *Structural analyses of the chromatin remodelling enzymes INO80-C and SWR-C.* Nat Commun, 2015. **6**: p. 7108.

133. Ayala, R., et al., *Structure and regulation of the human INO80-nucleosome complex.* Nature, 2018. **556**(7701): p. 391-395.

134. Eustermann, S., et al., *Structural basis for ATP-dependent chromatin remodelling by the INO80 complex.* Nature, 2018. **556**(7701): p. 386-390.

135. Farnung, L., et al., *Nucleosome-Chd1 structure and implications for chromatin remodelling.* Nature, 2017. **550**(7677): p. 539-542.

136. Sundaramoorthy, R., et al., *Structure of the chromatin remodelling enzyme Chd1 bound to a ubiquitinylated nucleosome.* Elife, 2018. **7**.

137. Gerhold, C.B., et al., *Structure of Actin-related protein 8 and its contribution to nucleosome binding.* Nucleic Acids Res, 2012. **40**(21): p. 11036-46.

138. Shen, X., et al., *Involvement of actin-related proteins in ATP-dependent chromatin remodeling.* Mol Cell, 2003. **12**(1): p. 147-55.

139. Szerlong, H., et al., *The HSA domain binds nuclear actin-related proteins to regulate chromatin-remodeling ATPases.* Nat Struct Mol Biol, 2008. **15**(5): p. 469-76.

140. Knoll, K.R., et al., *The nuclear actin-containing Arp8 module is a linker DNA sensor driving INO80 chromatin remodeling.* Nat Struct Mol Biol, 2018. **25**(9): p. 823-832.

141. Zhou, C.Y., et al., *The Yeast INO80 Complex Operates as a Tunable DNA Length-Sensitive Switch to Regulate Nucleosome Sliding.* Mol Cell, 2018. **69**(4): p. 677-688 e9.

142. Yen, K., et al., *Genome-wide nucleosome specificity and directionality of chromatin remodelers.* Cell, 2012. **149**(7): p. 1461-73.

143. Yen, K., V. Vinayachandran, and B.F. Pugh, *SWR-C and INO80 chromatin remodelers recognize nucleosome-free regions near +1 nucleosomes.* Cell, 2013. **154**(6): p. 1246-56.

144. Morrison, A.J., et al., *INO80 and gamma-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair.* Cell, 2004. **119**(6): p. 767-75.

145. Ray, S. and A. Grove, *Interaction of Saccharomyces cerevisiae HMO2 domains with distorted DNA.* Biochemistry, 2012. **51**(9): p. 1825-35.

146. Ray, S. and A. Grove, *The yeast high mobility group protein HMO2, a subunit of the chromatin-remodeling complex INO80, binds DNA ends.* Nucleic Acids Res, 2009. **37**(19): p. 6389-99.

147. Kiianitsa, K., J.A. Solinger, and W.D. Heyer, *NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities.* Anal Biochem, 2003. **321**(2): p. 266-71.

148. Webb, B. and A. Sali, *Protein Structure Modeling with MODELLER.* Methods Mol Biol, 2017. **1654**: p. 39-54.

149. Zimmermann, L., et al., *A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core.* J Mol Biol, 2018. **430**(15): p. 2237-2243.

150. Masse, J.E., et al., *The S. cerevisiae architectural HMGB protein NHP6A complexed with DNA: DNA and protein conformational changes upon binding.* J Mol Biol, 2002. **323**(2): p. 263-84.

151. Yang, Z.R., et al., *RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins.* Bioinformatics, 2005. **21**(16): p. 3369-76.

152. Liu, L. and T. Jiang, *Crystal structure of the ATPase-C domain of the chromatin remodeller Fun30 from Saccharomyces cerevisiae.* Acta Crystallogr F Struct Biol Commun, 2017. **73**(Pt 1): p. 9-15.

4. Abbreviations

Abbreviations not explained in the text

5´pApA	5´Phosphoadenylyl (3´->5´) adenosine
AMP/ADP/ATP	Adenosine mono/di/triphosphate
ATPγS	Adenosine 5´-[γ-thio]triphosphate
bp	Base pairs
bsu	*Bacillus subtilis*
cAMP	Cyclic adenosine monophosphate
c-di-AMP	Cyclic diadenosine monophosphate
c-di-GMP	Cyclic diguanosine monophosphate
cGAMP	Cyclic guanosine monophosphate - adenosine monophosphate
cGMP	Cyclic guanosine monophosphate
cryo-EM	cryo Electron Microscopy
DAC	Diadenylate cyclase
db	Desthiobiotin
dHJ	Double Holliday Junction
DSB	Double strand break
dT	Desthiobiotin
DTT	1,4-dithiothreitol
E	EDTA
EMSA	Electrophoretic Mobility Shift Assay
G	Glycerol
GDP/GTP	Guanosine di/triphosphate
H	Hepes (4 (2 hydroxyethyl)piperazine 1 ethanesulfonic acid)
HMG	High mobility group
HR	Homologous recombination
I	Imidazole
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
mtub	*Mycobacterium tuberculosis*
N	Sodium chloride
OD600	Optical density at 600 nm
P	Phosphate buffer
PDE	Phophodiesterase
pGpG	5´Phosphoguanosyl (3´->5´) guanosine
ppGpp	Guanosine tetraphosphat
pppGpp	Guanosine pentaphosphat
RCK_N/C	RCK N/C-terminal domain
S	Saccharose
SDS	Sodium dodecyl sulfate
ssDNA	Single stranded DNA
T	Tris (Tris(hydroxymethyl) aminomethane)
tma	*Thermotoga maritima*
β	β-Mercaptoethanol

5. Acknowledgements

It is of great importance to me, to express my sincere thanks to the people who helped me through my PhD time:

First of all, I want to thank my family for support and tolerance. Especially Liliana for accepting a big share of my time being spend with the laptop rather than with her.´´

I want to thank my "Doktorvater" Prof. Dr. Karl-Peter Hopfner for giving me the chance to join his lab. I am especially grateful for his support during my parental leave.

Furthermore, I want to give heartfelt thanks to PD Dr. Gregor Witte for giving me the opportunity to join the c-di-AMP team. His support and supervision were very much appreciated. I am furthermore thankful for personal guidance in pregnancy and parenting time.

I want to express my sincere gratitude to Elena Conti as TAC member and reviewer of my thesis. I am thankful for all the advice over the years.

A sincere thank deserves Adrian Bandera for letting me join his project – I really enjoyed the work on GdpP.

I am also very thankful for the help I got in the lab during pregnancy – especially from Franziska Kunert, David Drexler and Adrian Bandera.

I want to thank all the "reviewers", Charlotte Lässig, David Drexler, Marianne Schwarz, Adrian Bandera, Nicole Eisele and Gregor Witte for spending their time on reading my thesis and helping me to improve it.

Additionally, I want to thank Marianne Schwarz for countless cups of tea during our tee time!

Thanks to the Graduate School for Integrated Analysis of Macromolecular Complexes and Hybrid Methods in Genome Biology and the SFB 1064 collaborative research center for Chromatin Dynamics for funding.

6. Appendix

<u>Vector sequences</u>

Phen6-nanobody #42

GctggaTtGttatTACTCGCGGccaGCCGGCCAtgGCTCAGGTGCAGCTGGTGGAgtCTGGGGGAGGATTGGTGCA
GGCTGGGGGCTCTCTGAAACTCTCCtgtGCAGCCTCTGGACGCACCGGCACTATCTCtgTCATGGGtTGGTTCCG
CCAGGCTCCAGGGAAAGAGCGTGAGTTTGTAGGACGTAGACTTTTGACTGGTGACATCACAAGGTATGCAGA
TTCCGTGAAGGACCGGTTCACCATCTCCAGAGACAACGCCAGAAGCACGGTCTATCTTGAAATGAACAGCCTG
AAACCTGAGGACACGGCCGTTTATTACTGTGCAGCGGGGAAATATTTCTCACGCAACGATCAAACATATGACT
ACTGGGGCCAGGGGACCCAGGTCACCGTCTCCTCAAAGGGATCCTCCGCTTGGAGCCACCCGCAGTTCGAGA
AAGGTGGAGGTTCCGGAGGTGGATCGGGAGGTTCGGCGTGGAGCCACCCGCAGTTCGAAAAATAACTCGAG
GCGCGAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTT
GCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGC
GCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATA
CGTCAAAGCAACCATAGTACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCG
TGACCGCTACACTTGCCAGCGCCTTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGG
CTTTCcCCGTCAAGCTCTAnATCGGGGGGGCTCCCTTTTAGGGTTCCGatntTAGTGCTtTacngGCACCTcgnaCcCC
AAAAAAaCTtngATTTGGgGTGAtgGGtTCACGTAantGgGGCCATC…

pET21Ies5

caccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaacc
gcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactatagggga
attgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatataCATATGCCTAGTAAAGATCCAGAGA
GCGTTATAGACAAAGAAATACGGAAGATATCAGCAAGAAATGATGAGCTTATCAAACAGGATGGGACTTTAA
AAAGGGAATACACTACATTATTGAGAAAAGTATCGAGCGTGATAACGGTACTGAACAGTATTGATGATGCAGA
TACTGGTTCTGCTGAAACGGAACTCCCACGCTTAATATCTCAAGCAACAGTAGAGAAAGTACCCGAATTGAAG
TGGTACAATGATCAGATATCATTGATTACAGAAAAGTTGGAGGACGATGAAGATATTGAGGTTCCCGAGGAG
CTAATGGATGCTTATACTCTGTACAAGGAAACCCCATTATTATATAATGATACGCACACACCATAACTCGAGcac
caccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataacccc
ttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggat…

pET21-Nhp10

gatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaa
tacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatatacatatgTCAGTTG
AAGAAAAAAAGCGCAGACTTGAAGAATTGAAAGATCAGAATGTGGTGCTTGGACTTGCAATCCAAAGGTCAC
GACTTTCGGTGAAAAGACTCAAATTAGAATACGGTGTTTTACTGGAGAGGTTGGAATCCAGGATAGAACTTGA
TCCTGAGTTAAACTGTGAGGATCCATTGCCGACTCTGGCCTCCTTCAAACAGGAACTATTGACGAAGCCATTTC
GAAAATCTAAAACTAAAAGACACAAAGTAAAGGAAAGAGATCCGAATATGCCCAAGAGGCCCACAAATGCAT
ACTTATTGTACTGTGAAATGAATAAGGAAAGGATTAGGCAGAATGGCTCTCTGGACGTGACAAGAGATCTGG
CAGAAGGTTGGAAAAATCTTAACGAACAAGATAGGAAACCGTACTATAAGCTTTATAGTGAGGACAGAGAAA
GATACCAGATGGAAATGGAAATATATAATAAAAAGATATCTAATATAGATGCGGATGATGATAAAGAAGAGA
ATGAACAGAAGATAAAAAATAACGAGGAAGGATCATCTACAAAAGTAGCTGATTCTAAAGGAGGTGAAGATG
GAAGTTTAGTTTCCTCTAACTAAGgcggccgcactcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaag
gaagctgagttggctgctgccaccgctgagcaataactagcataacccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggagg
aactatatccggat…

pCDF-Ies3

ggggaattgtgagcggataacaattcccctgtagaaataattttgtttaactttaataaggagatataccATGGGCAAGTTCGAAGACCT
CTTGGCAACCAATAAACAGgtGCAGTTTGCTCATGCAGCTACTCAACATTACAAAAGTGTCAAAACTCCAGATTT
CTTGGAGAAAGACCCACATCACAAAAAATTTCATAACGCAGaTGGCTTGAATCAACAAGGTTCTTCGACCCCAT
CTACAGCTACCGATGCAAATGCCGCCTCAACTGCTTCCACTCACACAAACACCACGACATTCAAAAGACACATT
GTTGCCGTGGATGACATATCGAAAATGAACTATGAAATGATAAAGAATTCTCCAGGGAATGTAATAACGAACG
CTAACCGAGATGAGATAGATATAAGCACCCTCAAGACAAGGTTGTACAAGGACAATCTTTACGCTATGAATGA
CAACTTTTTACAAGCTGTCAATGACCAAATTGTAACATTAAATGCAGCGGAGCAGGACCAAGAAACTGAGGAC
CCTGACTTATCAGATGATGAAAAAATTGATATACTGACCAAAATCCAAGAGAACTTGCTTGAAGAGTACCAAA
AACTTTCACAGAAAGAAAGAAAGTGGTTTATATTGAAAGAACTACTGTTGGATGCAAATGTAGAACTCGATTT
GTTCAGTAATCGTGGCAGGAAGGCAAGCCATCCAATCGCATTCGGAGCAGTAGCCATACCAACGAACGTCAAT
GCGAATTCACTAGCATTTAACAGAACCAAGAGAAGAnAAATTAATAAGAACGGCCTTTTGGAAAACATTCTTTA
gcggccgcataatgcttaagtcgaacagaaagtaatcgtattgtacacggccgcataatcgaaattaatacgactcactataggggaattgtgag
cggataacaattccccatcttagtatattagttaagtataagaaggagatatacatatggcagatctcaattggatatcggccggccacgcgatcg
ctgacgtcggtaccctcgagtctggtaaagaaaccgctgctgcgaaatttgaacgccagcacatggactcgtctactagcgcagcttaattaacct
aggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaacctcaggcatttgaga
agcacacggtcacactgcttccggtagtcaataaaccggtaaaccagcaatagacataagcggctattt…

pETDuet-Ies3Ies5

ggggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatataccatgggggCCTAGTAAAGATCCA
GAGAGCGTTATAGACAAAGAAATACGGAAGATATCAGCAAGAAATGATGAGCTTATCAAACAGGATGGGACT
CTAAAAAGGGAATACACTACATTATTGAGAAAAGTATCGAGCGTGATAACGGTACTGAACAGTATTGATGATG
CAGATACTGGTTCTGCTGAAACGGAACTCCCACGCTTAATATCTCAAGCAACAGTAGAGAAAGTACCCGAATT
GAAGTGGTACAATGATCAGATATCACTGATTACAGAAAAGTTGGAGGACGATGAAGATATTGAGGTTCCCGA
GGAGCTAATGGATGCTTATACTCTGTACAAGGAAACCCCATTATTATATAATGATACGCACACACCATAAGCGG
CCGCATAATGCTTAAGTCGAACAGAAAGTAATCGTATTGTACACGGCCGCATAATCGAAATTAATACGACTCAC
TATAGGGGAATTGTGAGCGGATAACAATTCCCCATCTTAGTATATTAGTTAAGTATAAGAAGGAGATATACAT
ATGAAGTTCGAAGACCTCTTGGCAACCAATAAACAGGTGCAGTTTGCTCATGCAGCTACTCAACATTACAAAAG
TGTCAAAACTCCAGATTTCTTGGAGAAAGACCCACATCACAAAAAATTTCATAACGCAGATGGCTTGAATCAAC
AAGGTTCTTCGACCCCATCTACAGCTACCGATGCAAATGCCGCCTCAACTGCTTCCACTCACACAAACACCACG
ACATTCAAAAGACACATTGTTGCCGTGGATGACATATCGAAAATGAACTATGAAATGATAAAGAATTCTCCAG
GGAATGTAATAACGAACGCTAACCAAGATGAGATAGATATAAGCACCCTCAAGACAAGGTTGTACAAGGACA
ATCTTTACGCTATGAATGACAACTTTTTACAAGCTGTCAATGACCAAATTGTAACATTAAATGCAGCGGAGCAG
GACCAAGAAACTGAGGACCCTGACTTATCAGATGATGAAAAAATTGATATACTGACCAAAATCCAAGAGAACT
TGCTTGAAGAGTACCAAAAACTTTCACAGAAAGAAAGAAAGTGGTTTATATTGAAAGAACTACTGTTGGATGC
AAATGTAGAACTCGATTTGTTCAGTAATCGTGGCAGGAAGGCAAGCCATCCAATCGCATTCGGAGCAGTAGCC
ATACCAACGAACGTCAATGCGAATTCACTAGCATTTAACAGAACCAAGAGAAGAAAAATTAATAAGAACGGCC
TTTTGGAAAACATTCTTTAGcatatggcagatctcaattggatatcggccggccacgcgatcgctgacgtcggtaccctcgagtctggtaa
agaaaccgctgctgcgaaatttgaacgccagcacatggactcgtctactagcgcagcttaattaacctaggctgctgccaccgctgagcaataac
tagcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggattggcgaatgggacgcgccctgt…

pET28-Fun30_1-978

atggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaag
gaatggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaa
gtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccgg
cgtagaggatcgagatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctagaaataattttgtt
taactttaagaaggagatatacatATGGAGTTCTCATCGCCGGCCACTGGCCGCCGCTTTCTGGATGACGATGATGACC
ACCAAAGCGTCATCATGGAATCACCATACTTTACACAGCCCACTCAGATCATTGAAAAACCGACCGCAAAGCCA
CAGCCCTCTGTGACAATTTCTTCGCCAAGTTCTGTTGTCGAAGTTCCCGCTTCCTCACCCTTTCGTCCCTCTACCA

CACCGCAACACAGAAATCGTTCAGCAAACTTGATGGCTTCAAATCCCACGTCTCAGCTTCACGCGCAAAAGCCC
TTTACAATCTCTAAAACTTCTCATAAACGAGAGTTTGTCATGATATCTGATGGTGAGCTGGATGCTCCCATCTAT
GCTGGAGATGACTCGTCCGATGATGATGAGCGGTTGGCGACTCGAGCCGATATTCGCCCTTCTTCGTTCGAAA
GGAAGCAGCCAGCTACAACTAATATTACTGCATACAGAGCTCAGGCCGAAGATAAACAGAGCCCAGCGCAGG
GCAAGCCCGGCACCTCAGCTTCCAATGGAACTTCCAAGCAATATGTGTTGCAGAATGACTGGGTTTCAGTCTCT
TCTGATTCTTCAGATGAGCTCGAGCACCTGTGCCCAGCAAAACCACCGCAGCAAACCATCCGGCGGCGTCTTGT
TCGAGGTTGTCGTCCTCGTGTCTCGAGTCCGGACGGTCCTCAAGTTAATCCTGGCAAATTATCTGCAGCTCGTG
TTAAGGTCATCGATCTCGTTAGTGATGACGACCACGATAACGACGACGAAGAGGACAACTACCAACCTGGCAA
AGACTCCCATCACCACTCAGGCTTTGAAGAGAATGAAGAACAGGAACTAAATGATCGCGTGCTGGCTTATCTC
AATTCTTGTGATGCGATTGGTCTCATGGCTATTGCCAACATTAAAGAGGCGAGTGCGAAGGCCATGGTGTCGC
GGCAGCCCTTTCATGACCTCTCCGAGGCACAGGCCGTCGCAATGGCCCAGAAAAAGCGGGGCAAGAAGTCAG
CCAAAGTTGTCATTGGTGAGGATATCGTCTACGCCGTCAAGTCCTATGCACGTGCGTTGGACTCTATCGATCGT
GTCATCGCAGTTTGCGAGGATAACGCCAAAAAGATCAAGTCTGTGACTTCGAAGTGGAATATCGATCAGACGG
GCCAATGCCGCGAGGAATTCAAGAACGCCGACGACAAGATTCTCACTCCGCACAGCGCTGAAGGCTCGACGA
GGTTGATGGATGTCCCGCTGCGTCAGCCAAAGCTCATGACGGGCTGTGTCATGAAGCCCTTTCAGATATACGG
ACTCAATTGGCTGTATCTTCTGTATAAGAATCAGCATGGCGCCATCCTTGCTGACGACATGGGTCTTGGCAAGA
CATGTCAGGTCATCTCTTTGATGTGCTCCATTGTTGAGCAGTGGGAGAGGACCGGCGGCGAGAGTGGTCCTCG
TCCCTGGCCGAATCTTATCATCGTTCCACCTTCAACGTTGGAAAACTGGGAGAACGAATTCAAGAAATTTGCCC
CGAAGCTGCGAGTCACCAAATACCAAGGCAGTCAGGCAGCCAGAGATGATATATTGGATGAAATCTTGGATG
ACCCCTCGAATCACCATGTCATCTTGACTTCTTACTCTCAGCTTACTCGGGAAAAGGAGCTGAAAGACCTTGCC
AAATGCCACCCGAAAGTCGCAGTCTTCGATGAAGGTCACAAGATGAAGAACCCAAACTCACAGTTCTACAAGC
GACTGCTCAAGTTCAAGAATTCGTGGCGTCTGGTCTTGTCTGGAACTCCAGTCCAGAACAACCTGATGGAAAT
GATTGCGCTGCTCAGGTTTGTGTCACCAAACCTTTTCACGGCCCACTTCGAGACTCTGGAGGAGCTGTTCTCTC
AAAAGTTCTCTCTCGCCGATGTGTCTAAGGGGGCGCTCCTTCGTAGCGAGCGAGTCGACCGCGCTCGAACGAT
CTTGGAGCCGTTCATTTTGCAACGGCGGAAAGAACAGGTCTTGAAGGATCTCCCGCCTAAACTCACCAGGATT
ATTAAATGTCAGATGGATCCTGTCCAAGCAGCTCTGTACAAGGAATACGAGCGTCGGTTTAAGAAGTCAGGCG
TTCAAAATTCTTTGCCTTCCGACAACAACATCTGGATGCAACTGCGCAAAGCCGCCATCCACCCACAGCTTTTCC
GCAGGTTCTACAAAGATAAGGATATTGAAAAGATGGCCAGCATTCTCTTCAAAAAAGTCCCTCAACGTGACCTC
AAACAGCCCAAACTGGAACATCTCGTCAACGAATTGAAGGCCCTGTCAGACTTCGAGCTGCATCTTTGGTGCA
GAGATTATTCATGTATCAAGAACTTCGATCTTCCGAAGGGCTCTTGGTTGGACTGTGCGAAGGTCAGGGAGCT
TTTGAAGCTCATCCGGGGGTACATGGCCAATGGAGATCGAGCACTGGTTTTTACACGTTTTGCCAAGGTCCTTG
AGATTCTTTCTGAATGTCTTGCAGACGAGAACATCCCTTACTTGGCTTTGCAAGGCAACACCAGCGTGGACGAG
CGACAGGATATGATTAACGAGTTCAATTCAAACGAAAACATTCCGGTGTTCCTCCTCACCACTGGCTCAGGAG
GCACAGGTGTTAACCTGGCGTCTGCAAACAAGGTTATTCTTTTCGATCTTTCTGATAATCCACAGGATGATATTC
AGGCTGAGAATAGGGCCCATCGCCTTGGCCAGACACGGCCAGTCGAAGTTATCAGGCTCATCAGCGAGGGTA
CCATCGAAGAACTTGTGCACAAGGCGTGCGAGAAGAAGCTCGAACTAGCCAACAAGGTCACAGGTTGGACCG
ACGTCTCGATGACTTCCGAGGAAATGGAGAAGATCGTTAAGCAATCGTTGCTTTCGGGTGAAGGTGTTGCTAC
TCCTCTCTCCAACGACCTGGAAGTTCTGTTCCAGGGGCCCCATCATCATCATCATCATCATCATTAAaagcttgcgg
ccgcactcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaata
actagcataaccccttggggcctctaaacgggtcttgaggggtttttttgctgaaaggaggaactatatccggat…