Chloride dependent regulation of osmolyte producing GgpS

and

Remodelling of the substrate translocating loop by ADP release in AAA^{+} protease FtsH

INAUGURAL–DISSERTATION

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Matthias Uthoff

Aus Gütersloh

2020

Berichterstatter:

Prof. Dr. Ulrich Baumann

Prof. Dr. Kay Hofmann

Zusammenfassung

Die Glukosylglycerin-Phosphat Synthase (GgpS) ist eines der beiden Enzyme, die beteiligt sind an der Darstellung von Glukosylglycerin, dem wichtigsten kompatiblen Solut im Cyanobakterium *Synechocystis sp.* Verbindungen dieserart sind osmotisch aktiv, dienen damit der Halotoleranz und sind deswegen zugleich in hohem Maße kompatibel mit dem Metabolismus, d.h. vorranging Protein stabilisierend. Glukosylglycerin und andere kompatible Solute werden in verschiedenen Pharmazeutika, insbesondere aber in kosmetischen Produkten zur „Revitalisierung" oder Hydratisierung der Haut eingesetzt. GgpS ist eine Glykosyltransferase, die die Reaktion von ADP-Glukose und Glyceryl-3-phosphat zu ADP und Glukosylglycerin-phosphat unter Beibehaltung der Stereoisomerie des anomeren Kohlenstoffs katalysiert. Frühere Veröffentlichungen berichten, dass GgpS unter Bedingungen mit niedrigem Salzgehalt inaktiv ist, aber konstitutiv exprimiert wird; hingegen jedoch aktiviert und stärker exprimiert wird nach dem Umsetzen in Medium mit hohem Salzgehalt. Bisher wurde angenommen, dass die Inaktivierung durch Bindung an DNA realisiert ist: Erst durch einen intrazellulären Salz-Schock werden GgpS und die Nukleinsäure voneinander getrennt. Die Salzionen schirmen dabei beide Bindungspartner elektrostatischen voneinander ab. Im Folgenden werde ich die erste Struktur von GgpS, einer Glykosyltransferase mit der bekannten B-Faltung, vorstellen. Überraschenderweise sind zwei Chlorid-Anionen an das konstituierte Tetramer gebunden und implizieren damit eine bislang unbekannte direkte Chlorid-Abhängigkeit von GgpS. In der Tat konnte ich einen Effekt von Chlorid auf die Aktivität validieren u.a. durch einen Vergleich mit einer Mutante die kein Chlorid an der spezifizierten Stelle mehr binden kann. Dieser Effekt könnte in der Vergangenheit entweder übersehen oder gar verwechselt worden sein mit einer Mg^{2+}-Abhängigkeit. Das Mg^{2+} war immer als Chlorid-Salz hinzugefügt worden. Zusätzlich zeige ich die Bedeutung der Chlorid-Bindung für die Reaktivierung nach DNA-

Hemmung und gebe erste Vorhersagen über den GgpS:DNA-Komplex.

Ein weiteres hier diskutiertes Projekt befasst sich mit der AAA$^+$ Protease FtsH. AAA$^+$ Proteine sind bekannt für ihre hexamere Ringstruktur und die Fähigkeit, Proteine oder Nukleinsäuren durch die gebildete Pore zu translozieren. Die Translokation entfaltet Strukturen und daher lassen sich AAA$^+$ Proteine dort finden, wo eine Entfaltung, möglicherweise vor einer korrekten Rückfaltung (Chaperone für Proteine), oder auch nur eine Bewegung entlang eines Substrats (Helikasen für Nukleinsäuren) erforderlich ist. Die Untersuchung des Translokationsmechanismus wird derzeit durch die aufkommende Methode der Kryo-Elektronenmikroskopie beschleunigt. Bei AAA$^+$ Proteasen ermöglicht ein weiterer Ring mit proteolytischer Aktivität einen vollständigen Abbau der ehemals stark gefalteten Proteine. Diese Proteasen gewährleisten daher die Kontrolle der Proteinqualität in allen Zellen und das Recycling ihrer Aminosäuren. FtsH ist eines der einfachsten Mitglieder dieser Proteinfamilie, da sein Hexamer nur aus sechs identischen Proteinen besteht. Es ist jedoch gleichzeitig die einzige membrangebundene AAA$^+$ Protease. Während meiner Masterarbeit (2015) begann ich ein neues Konstrukt aus *Aquifex aeolicus* ohne Membranteil zu kristallisieren. Die verfeinerte und nun gründlich analysierte Struktur wurde schließlich im Jahr 2018 veröffentlicht (Uthoff & Baumann, 2018). Ich beschrieb eine neue Faltung der „pore loop", welche normalerweise mit dem Substrat während der Translokation interagiert. Die alternative Faltung ist meiner Meinung nach für die korrekte Rückbewegung der Untereinheit nach der Translokation erforderlich. Um für den Mechanismus kritische Aminosäure zu validieren, entwickelte ich drei Aktivitätsassays welche ich ebenfalls hier vorstelle. Darüber hinaus haben wir eine Zusammenarbeit mit Elmar Behrmann begonnen und präsentieren die erste Struktur von volllängen-FtsH in Detergenz, die bei etwa 6 Å aufgelöst ist. Kryo-Elektronenmikroskopie mit Membranproteinen ist jedoch eine Herausforderung und die Arbeiten dauern an. Ein vielversprechender Datensatz für ein Modell mit höherer Auflösung wird derzeit prozessiert.

ABSTRACT

Glucosylglycerol-phosphate synthase (GgpS) is one of the two enzymes required for the production of glucosylglycerol, the main compatible solute in the cyanobacterium *Synechocystis sp*. These compounds are essential for halotolerance and as such are osmotic active but highly compatible with the metabolism, i.e. they are protein stabilising. Glucosylglycerol and other compatible solutes are commercialised in various pharma but especially in cosmetic products for skin 'revitalisation' and as hydrating agents. GgpS is a glycosyltransferase catalysing the reaction of ADP-glucose and glycerol-3-phosphate to ADP and glucosylglycerol-phosphate under retention of the stereoisomerism of the anomeric carbon. GgpS was found to be constitutively expressed but inactive under low salt conditions and strongly expressed and activated if the cells are shocked with medium containing high salt concentrations. Previously, inactivation was thought to be realised by binding of GgpS to DNA and thought to be activated due to electrostatic shielding of GgpS from the nucleic acid by the salt ions. Here I will present the first 3D-structure of GgpS a glycosyltransferase with the common B fold. Surprisingly, two chloride anions are bound to the tetrameric enzyme and imply a so far unknown direct chloride dependence of GgpS. I could indeed validate the effect of chloride on the enzymatic activity by comparison of the wild type with a binding deficient mutant. This effect probably has been missed due to or mixed up with an Mg^{2+}-dependency. The Mg^{2+} cation in previous studies has always been added in the form of $MgCl_2$. Additionally, I demonstrate the importance of chloride binding for DNA inhibition rescue and give first predictions of the GgpS:DNA complex.

Another project discussed here is about the AAA^+ protease FtsH. AAA^+ proteins are known for their hexameric ring structure and the ability to translocate proteins or nucleic acids through the central pore. Translocation unfolds 3D-structures, and hence AAA^+ proteins are found where unfolding, potentially prior to correct refolding (chaperone for proteins), or where just the movement along a

substrate (helicases for nucleic acids) is required. Investigation of the translocation mechanism is currently accelerated by the emerging method of cryo-electron microscopy. In AAA$^+$ proteases, another ring with proteolytic activity enables these complexes to completely degrade tightly folded proteins. Hence, they ensure protein quality control in all cells and recycling of their amino acids. FtsH is one of the simplest members of this family of proteins as its hexamer is constituted by just six identical proteins. However, it is simultaneously the only membrane bound AAA$^+$ protease in bacteria, mitochondria and plastids. During my master's thesis (2015), I started crystallising a new construct of FtsH from *Aquifex aeolicus* lacking the N-terminal trans-membrane and periplasmic part. I described a new conformation of the usually substrate-engaging 'pore loop' which I believe is required for correct subunit movement after translocation (Uthoff & Baumann, 2018). To validate critical residues, I present here the development of three activity assays. In addition, we started a collaboration with Elmar Behrmann and I will present here the first full-length structure of FtsH solubilised in detergents at around 6 Å. However, cryo-electron microscopy with membrane proteins is challenging and the work is ongoing. A very promising dataset for a higher resolution structure is currently being processed.

TABLE OF CONTENTS

ABREVIATIONS

AAA	ATPases Associated with diverse cellular Activities
ADPG	ADP-glucose
AMP-PNP	Adenylyl-imidodiphosphate
DDM	n-dodecyl-β-D-maltopyranoside
EM	Electron microscopy
G3P	Glycerol-3-phosphat
GG	Glucosylglycerol
GGP	Glucosylglycerol-phosphate
GOL	Glycerol
GT	Glycosyltransferase
ISS	Intersubunit signalling network
LPS	Lipopolysaccharide
MSP	Membrane scaffolding protein
NCS	Non-crystallographic symmetry
NMR	Nuclear magnetic resonance
PHB	Prohibitin homology domain
RMSD	Root mean square deviation
SD	Standard deviation
SE	Standard error
SEC	Size exclusion chromatography
SEC-MALS	Size exclusion chromatography with multi angle light scattering
SMA	Styrene-maleic acid copolymer
SRH	Second region of homology
SUC	Sucrose
tb-loop	Tension belt loop
VDO	Validoxylamine A 6'-O-phosphate, bisubstrate analoge of TPS

GgpS

1 GGPS – INTRODUCTION

1.1 Salt and osmosis – Opportunities and challenges

'Life is a characteristic that distinguishes physical entities that have biological processes [...] from those that do not [...]' (Wikipedia, 2020). This is the first sentence written in the 'Wikipedia' article about life and the truncated statement is apparently a circular reasoning because the term 'biology' is ancient Greek for 'the study of life'. The statement carries an important fact nonetheless: Life is a physical entity and as such is distinct of its environment. A common feature for all living entities on earth is a thin lipid bilayer separating them from their surroundings. The membrane is hypothesised to have coevolved with the metabolism from simple abiotic to increasingly complex 'biological' processes (Szathmáry, 2007). Today, (eukaryotic) cells are highly compartmentalised: Vacuoles store things, lysosomes degrade molecules, mitochondria and chloroplast generate energy, peroxisomes contain harmful reactions... All are separated by one/several lipid bilayer membrane/s which are semi-permeable. This usually means that small nonpolar molecules can freely diffuse to the other side but not charged, polar and bulky ones. Controlling the flow of ions actively, enables cells to separate charges on both sides. This generates the membrane potential required for signal transduction and energy production.

1.1.1 The problem of osmosis

Osmosis is a phenomenon occurring at semi-permeable barriers like bio membranes. The diffusion rate of water through simple lipid bilayers is actually quite low but it is also the universal solvent required for all living beings. Hence, cells evolved for rapid mostly passive water exchange by integrating aquaporins (Haines, 1994). On the side where there are more solutes (hypertonic), the water has effectively a lower concentration and accordingly a lower chemical potential, and there is therefore a net diffusion rate of water towards that side. It becomes naught if the water concentrations or more precisely the water activity becomes equal in both

spaces. Thus, not the solutes but the water diffuses to equilibrate the concentrations. However, this increases volume or, usually more importantly, the pressure on the former hypotonic side. Hence the true equilibrium is reached when the osmotic potential on both sides equalises, this includes amongst others predominantly water activity and hydrostatic pressure. Typical cells have evolved to be slightly over-pressurised (turgor), but a higher pressure might burst the cell. Controlling osmosis is therefore critical, especially for single cell organisms and it must always maintain a slightly lower potential i.e. usually a lower water activity in its interior (Campbell & Reece, 2009).

Bacteria and archaea live or survive up to 5.1 M or 30 % sodium chloride and are often divided according to their preferential growth conditions into slight (0.34–0.85 M), moderate (0.85–3.4 M) and extreme (3.4–5.1 M) halophiles (Ollivier et al., 1994). The living conditions are unfortunately not constant and can change rapidly. During low tides small ponds can evaporate and thereby decreasing considerably the water activity inside. Likewise, rain can dilute such ponds quickly, increasing their water activity. Hence, single cell organisms often had to evolve systems for rapid adaptation.

1.1.2   Pitfalls in osmoregulation – The Hofmeister phenomenon

Water activity can easily be reduced by production or uptake of osmolytes, i.e. osmotically active agents. However, some compounds may have adverse effects on the metabolism. Cells evolved for low intracellular sodium, chloride, and calcium, but higher potassium levels as required for the membrane potential. In addition, there are more, less apparent effects. Franz Hofmeister described how certain compounds modulate the solubility of proteins (Hofmeister, 1888). Theses osmolytes were later categorised as chaotropes and kosmotropes, which either salt-in or salt-out. The former are supposed to break water structure and the latter to increase it, thereby altering the protein stability. In more detail, salting-in refers to the ability of chaotropes to increase solubility of folded proteins at low concentrations while at higher

concentration the unfolded denatured state is stabilised. Eventually, also chaotropes precipitate polypeptides. The salting-out effect of kosmotropes on the other hand only stabilises folded proteins and decreases directly their solubility. The categories and their names were kept, but the idea of more or less ordered bulk water was discarded (Ball & Hallsworth, 2015; Y. Zhang & Cremer, 2006).

The Hofmeister phenomenon is a mixture of effects depending on the agent. Zhang and co-workers tested the classical Hofmeister anions on a peptide-bond model polymer (PNIPAM) and reported that all their sodium salts increase surface tension in general (Y. Zhang et al., 2005). This means all enhance separation of protein and water and have thus a classical salting-out effect, which is stronger for kosmotropes as one might expect. In addition, kosmotropes have a strong negative hydration entropy and are thus able to polarise i.e. remove the hydration shells of proteins, further increasing the salting-out effect. Chaotropes on the other hand have a low negative hydration entropy and it is their hydration shell which is removed if close to proteins. This enables them to interact with very low affinity sites on the protein directly. Hence, one may say the protein is not only solved in the water but in the chaotropic agent as well, increasing overall solubility. However, the peptide bonds are potential binding sites, too. Most of them are usually only accessible in the unfolded state and the proteins therefore start to unfold with increasing chaotrope concentrations. Since all interactions sites are finite, this effect saturates and is overcome by increased surface tension, leading eventually to precipitation.

The authors indicate their hypothesis to be not complete, especially since they did not analyse cations and the Hofmeister effect of complex compounds is certainly more complex. Urea and guanidinium for example not only interact directly but have an amphipathic surfactant-like effect, increasing their penetration capability into the protein's core (England et al., 2008). Likewise, some solutes may not be able to interact with the peptide backbone at for example due to sterically reasons. So, while urea favourably

interacts with it, sarcosine avoids it and apolar sidechains. Sarcosine is therefore excluded from the protein and its vicinity and thus increases overall protein stability (Holthauzen et al., 2010). Atomic force microscopy experiments support this. Up to 600 mM sarcosine or ectoine reduces the elasticity i.e. increases the stability of fibronectin steadily, while sodium chloride does not impose any change up to 4 M (Oberdörfer et al., 2003). Hence, solutes can be divided into protein or metabolic incompatible solutes like urea and guanidinium and compatible solutes like sarcosine and ectoine.

1.1.3   Strategies for osmoadaptation – Potassium versus compatible solutes

If a cell wants to decrease its water activity, the employed mechanism to achieve this goal and the cell's metabolism must not interfere with each other. This can be achieved by the usage of the previously introduced group of compatible solutes to concentrations of up to 2 M, or by making the metabolism compatible to the solutes in use. A simplified overview is given in Figure 1. The latter strategy is usually employed by extreme halophilic organisms as in the archaea family of *Halobacteriaceae* and referred to as salting-in strategy. They in fact require higher potassium (chloride) concentrations for survival and their proteome became eventually more acidic, i.e. negatively charged and less hydrophobic to accommodate that potassium (Oren, 2013). This strategy is especially useful under extreme conditions because it is considered as less energy requiring as no agents are actively produced (Oren, 2011). Those cells and enzymes often need 2–4 M intracellular potassium and sometimes 100 mM Mg^{2+} for peak growth or activity (Lanyi, 1974). Reviews even report 5–7 M, more than the actual solubility of potassium chloride. The salt is believed to be bound to the intracellular contents, e.g. proteins, though no primary sources are given at all. Nevertheless, these organisms are potassium dependent and the salting-in strategy is more adjusting the basal halotolerance. The second strategy in contrast requires to actively accumulate or produce compatible solutes. They often fine tune the cell to the current condition and are used in most cells. However,
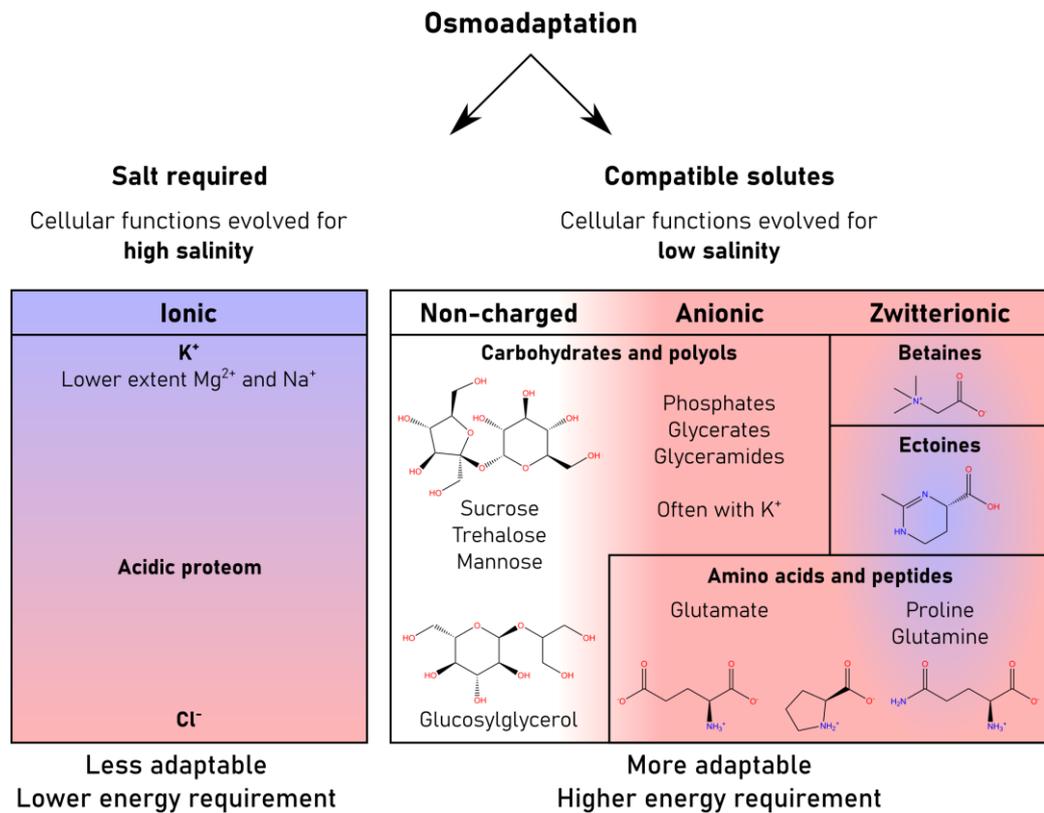
Figure 1: Simplified overview of osmoadaptation strategies and compounds used.

especially the synthesis of these products is more energy demanding.

As described, compatible solutes usually increase protein stability and in addition, since all solutes lower the freezing temperature, they also help against temperature stress and desiccation (Kunte et al., 2014; Lentzen & Schwarz, 2006). In contrasts to potassium, compatible solutes often do not need counter ions because they are neutral or zwitterionic. Ectoine and its derivative hydroxyectoine but also betaine, proline and glutamine belong to the latter group, while carbohydrates and polyols are usually neutral (Czech et al., 2018; Galinski & Trüper, 1994; Lentzen & Schwarz, 2006; Shivanand & Mugeraya, 2011; Welsh, 2000). However, phosphate, glycerate and glyceramide derivatives of carbohydrates are anionic in nature and often better suited to protect against heat stress (Faria et al., 2008). In addition, they require positive counter ions and are thus often found in combination with potassium.

Sodium uptake is usually the immediate response to salt shocks but is soon replaced against potassium while compatible solutes have to be produced first (Sleator & Hill, 2002). However, several organisms accumulate the solutes directly from the environment since they often have already been released during down shocks before. Welsh discusses this and other ecological effects in his great review (Welsh, 2000). Sleator and Hill review the response to salt shocks according to strategy (Sleator & Hill, 2002).

Compatible solutes have been shown to aid in periplasmic protein expression in *E. coli* and subsequent purification (Barth et al., 2000). Ectoine and mannosylglycerate can enhance refolding of proteins from inclusion bodies (Lentzen & Schwarz, 2006) and ectoine improved the growth of protein crystals (Harjes et al., 2004). Compatible solutes have been found helpful in neurodegenerative diseases since they can reduce prion, Aβ42 and α-synuclein aggregation in cell models (Jorge et al., 2016). Moreover, according to Merck scientists, pre-treatment of ectoine reduces 'sunburn' in organo-typical skin equivalents by UV irradiation. Also the skin of human test subjects were subsequently pre-treated and irradiated with a protective effect (Bünger et al., 2001). Glucosylglycerol (GG) enhances aquaporine levels in keratinocytes and reduces water loss in human skin (Schrader et al., 2012). In addition, it prevents human saliva bacteria from acid production thus potentially inhibits caries and it is an important natural component of Sake and Mirin (Takenaka & Uchiyama, 2000). GG can also decreases *in vitro* digestion of disaccharides and may help patients with diabetes mellitus (Takenaka & Uchiyama, 2001). At last, probiotic effects on intestine flora have been show and it is therefore discussed as an additive for functional foods (Sawangwan, 2015).

It is no surprise, that those compounds are already exploited on an industrial level, especially ectoine (Kunte et al., 2014). Itself, hydroxyectoine and glucosylglycerol are sold by the German company bito AG under the trademarks of 'Ectoin' and 'Glycoin' for skin care products (Bitop AG, 2020). Glucosylglycerol is most likely produced from sucrose and glycerol by immobilised sucrose phosphorylase (Bolivar et al., 2017), but a different method via the

enzymes GgpS and GgpP in *Corynebacterium glutamicum* was proposed recently (B. Roenneke et al., 2018).

1.2 *Synechocystis sp.* and the role of GgpS

1.2.1 Osmoadaptation in the moderate halotolerant *Synechocystis sp.* PCC6803

Cyanobacteria predominantly produce sucrose and trehalose (<0.6 M sodium chloride in medium), glucosylglycerol and glucosylglycerate (<1.7 M) or betaine and glutamate betaine (<3.0 M) as compatible solutes. The selection is not exclusive, the more tolerant species often switch production to a 'simpler' osmolyte under less extreme conditions. Hence, sucrose is found in most species at moderate conditions. However, while some species are just halotolerant, some require salt for survival (Martin Hagemann, 2011).

Synechocystis sp. PCC6714 was isolated in 1949 by G. C. Gerloff from a freshwater lake near Madison, Wisconsin, USA (JGI: G00071041). It produces glucosylglycerol as osmolyte and its response to a 500 mM sodium chloride shock was analysed (Reed et al., 1985). Adapted to BG-11' medium (5.0 mM sodium, 2.5 mM potassium, 2.0 mM chloride), the cells maintained levels of around 7 mM sodium, chloride and sucrose as well as 180 mM potassium but no glucosylglycerol. The following concentration are usually the concentration *above* these basal values as depicted in Figure 2. Within 2 min after the shock, sodium and chloride concentrations increased by 160 mM and 200 mM, respectively. While, the concentration of chloride did not change for about an hour, sodium was equimolarly exchanged against potassium within 20 min (around 340 mM *total* intracellular potassium). An hour later, ejection and replacement of chloride and potassium by sucrose and glucosylglycerol began. Thus, 24 h after the shock, ion levels stabilised at around 60 mM chloride and 30 mM potassium with 75 mM sucrose and 150 mM glucosylglycerol *above* initial values. At the end of the experiment after 48 h, sucrose concentration had dropped to 30 mM *above* initial values while GG production went
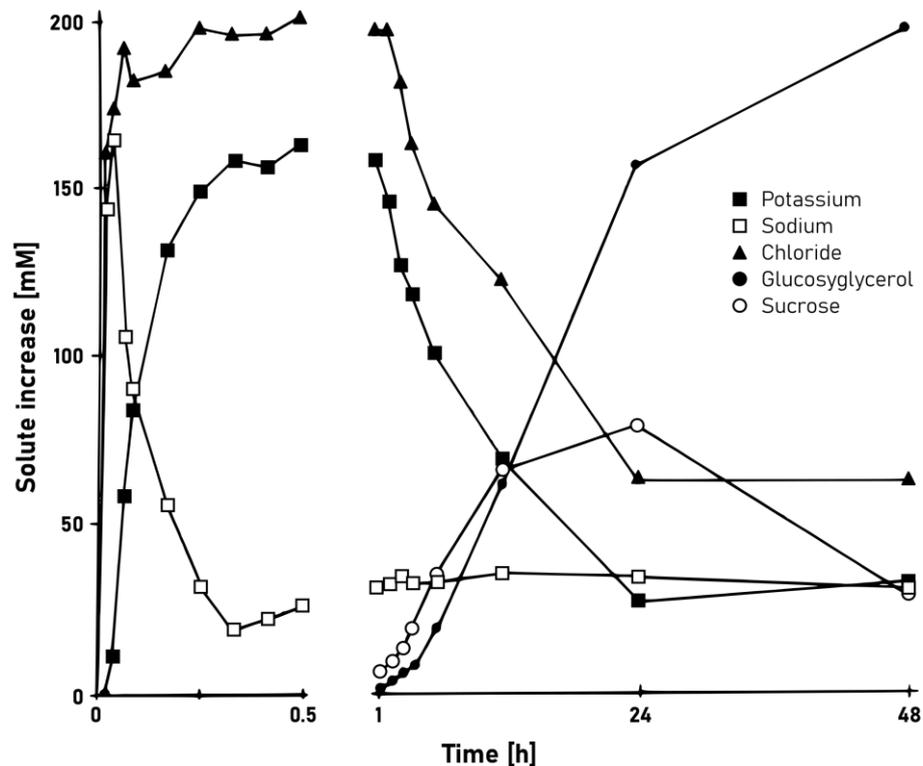
Figure 2: Potassium, sodium, chloride, glucosylglycerol and sucrose levels in *Synechocystis sp.* PCC6714.

Visually enhanced Figure 1 from (Reed et al., 1985). Changes in intracellular solute concentrations of PCC6714 cells transferred from BG-11' to a medium containing additional 500 mM sodium chloride. Data were collected within first 30 min (0.5 h) and between 1 h and 48 h after shock. Increases in glucosylglycerol and sucrose levels up to 30 min were osmotically insignificant and have not been included in first time frame. Values are represented as increases above the basal levels for cells in freshwater BG-11' medium (181.1 mM potassium, 7.8 mM sodium, 6.8 mM chloride and 7.7 mM for sucrose; glucosylglycerol was undetected in freshwater-grown cells). Thanks for permission of reuse from Oxford University Press.

on with 200 mM produced so far. The photosynthetic oxygen production was also analysed. It stopped after 2 min of the salt shock but recovered to 70 % after 30 min and 80 % after 60 min with around 10 % standard deviation or error.

Glucosylglycerol is produced from ADP-glucose and glycerol-3-phosphat (G3P) in a reaction catalysed first by glucosylglycerol-phosphate synthase (GgpS, KEGG: R05328) and a final dephosphorylation by glucosylglycerol 3-phosphatase (GgpP, KEGG: R05791, Figure 3). ADP-glucose is a central component in starch and trehalose metabolism and G3P links glycolysis and the generation of glycerolipids.
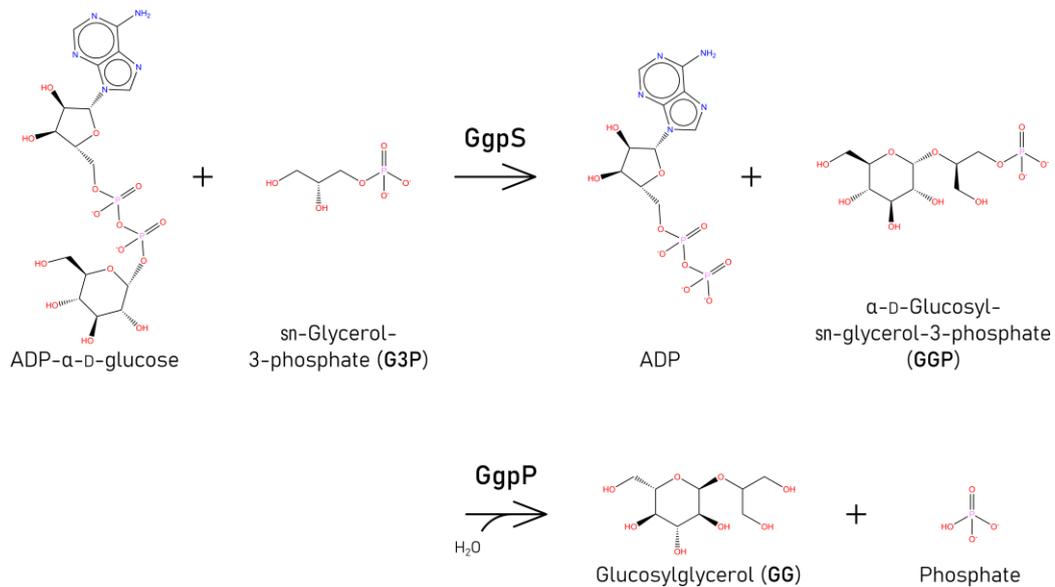
Figure 3: Synthesis of glucosylglycerol in cyanobacteria by GgpS and GgpP.
Red: Oxygen containing groups; Blue: nitrogen containing groups. (ChemDraw)

The GgpS from *Synechocystis sp.* PCC6803 was analysed in this thesis instead of PCC6714. Both strains are closely related based on 16S r-RNA (Martin Hagemann, 2011) and GgpS sequence (98 % coverage with 95 % identity). PCC6803 was isolated 1968 by Wolfgang Hess from a freshwater pond near Oakland, California, USA (JGI: G00000791).

1.2.2   Regulation of GgpS

GgpS was found to be constitutively expressed in *Synechocystis sp.* PCC6803 (Martin Hagemann et al., 1996) but salt shocks with 684 mM sodium chloride led to a tenfold increase in protein level (Marin et al., 2002). A linear relationship between mRNA, protein, glucosylglycerol and external sodium chloride concentration was found. PCC6803 tolerates more than 1.2 M sodium chloride, but photosynthesis declined and GgpS transcription was not induced directly at 1,026 mM. Nevertheless, mRNA and glucosylglycerol levels increased eventually to expected values. The authors further showed that mutation of σ^F, a transcription factor of several salt stress proteins, broke the relationship between medium-salt and mRNA concentration. However, glucosylglycerol accumulated up to 684 mM sodium chloride in the medium after five days as usual. Higher salt concentration led to diminished growth and glucosyl-glycerol accumulation (Marin et al., 2002). When shocked with

342 mM sodium chloride in presence of chloramphenicol, only half as much glucosylglycerol accumulated as without translation inhibitor after 3.5 h (Martin Hagemann et al., 1996). This clearly indicates the condition-dependent expression of GgpS as an important factor for glucosylglycerol production regulation. However, GgpS activity is directly fine-tuned especially at lower salt concentration.

GgpS in 'crude' cell extracts was found to be activated by sodium chloride (Schoor et al., 1999). All tested chloride and bromide salts activated stronger than iodide or nitrate. This fact was neglected and the focus was set on the cations and their apparent Hofmeister-like effect. Moreover, magnesium chloride was found to be exceptionally activating in low concentrations, thus laying the grounds for the belief in Mg^{2+}-dependence of GgpS. Results were refined by analysis of better purified GgpS (M. Hagemann et al., 2001) and led to the discovery of inhibition by poly-anions (nucleic acids or heparin) and subsequent sodium chloride activation (Jens F Novak et al., 2011). In this model, sodium chloride does not directly bind to the protein or the DNA but electrostatically shields both from each other. Thus, if the cells are under no salt stress, GgpS is inactivated by DNA or possibly also RNA. After a salt shock, potassium, sodium and chloride enter the cell and separate GgpS from the inhibitor, activating it, as it is now needed. This leads to the accumulation of glucosylglycerol and the intracellular concentration of inorganic ions can be reduced gradually. If the environment becomes less hypertonic or even hypotonic, more inorganic ions and eventually also the compatible solutes are released. This inactivates GgpS and adapts the cell quickly to the new condition.

1.3 Glycosyltransferases

GgpS (EC 2.4.1.213) is a glycosyltransferase (GT; EC 2.4). The 'carbohydrate-active enzyme database' (CAZy) states the following: 'these enzymes catalyse the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds' (Lombard et al., 2014). The activated donor

molecule usually refers to a monosaccharide coupled to a phosphate or a nucleotide mono- or diphosphate. GTs are therefore responsible for the formation of more complex sugars such as sucrose or cellulose and glycogen as well as the glycosylation of proteins, lipids and nucleic acids. Strictly speaking, hydrolysis and phosphorolysis are transfers to water and phosphates, respectively, and can be regarded as GTs.

There are numerous ways to classify glycosyltransferases. For example, the EC classification systems divides them into hexosyltransferases (EC 2.4.1), pentosyltransferases (EC 2.4.2), and transfer to 'other' glycosyl groups (EC 2.4.99). Based on sequences, CAZy has defined 111 families, of which few have been discontinued. A structural comparisons reveals three distinct folds: GT-A, GT-B and GT-C. These folds are also found in the GT-like hydrolases and phosphorylases as indicated above. Several reviews report about glycosyltransferases and are basis for the brief following overview (Coutinho et al., 2003; Gloster, 2014; Lairson et al., 2008; Moremen & Haltiwanger, 2019; Taujale et al., 2020).

Most GTs adopt the GT-A or the GT-B folds. A large Rossman-like domain, which is described as one composed of two Rossman motifs linked together, interacts with both substrates in GT-A proteins. In GT-B enzymes are the substrates bound in the cleft between two distinct Rossman-like subdomains. Here, the N-terminal domain usually stabilises the diverse acceptors, while the C-terminal one binds predominantly the activated sugar. A conserved feature of most GT-A enzymes is the DXD motif in the active site. It coordinates Mg^{2+} or sometimes Mn^{2+} to stabilise the phosphate leaving group of the donor. This interaction mode is replaced by arginines and lysines in GT-B proteins. Hence, most but not all GT-As are cofactor dependent, but no such dependence is known in GTs with the B fold.

Another important discriminating feature is the stereochemistry of the anomeric linkage in the product (Figure 4). The activated sugars are all α-anomers and transfers can either invert or retain this configuration. There is no correlation between fold and formed
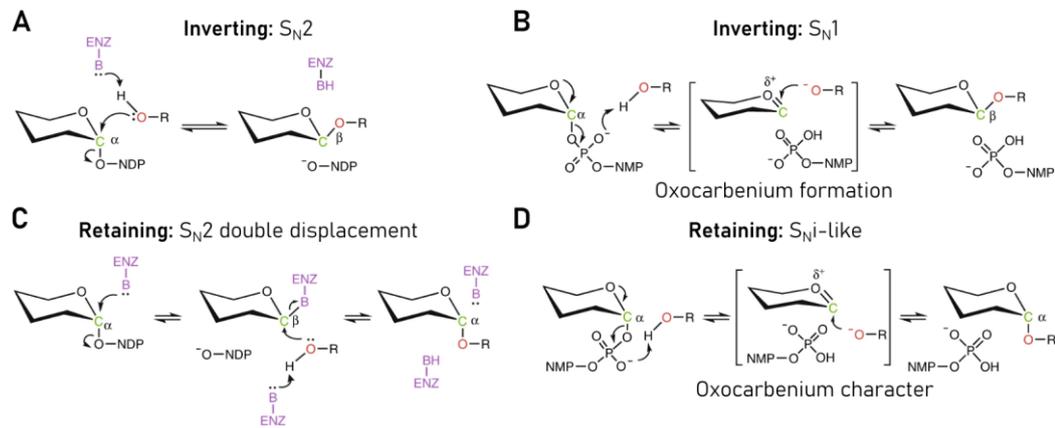
Figure 4: Reaction mechanism of inverting and retaining glycosyltransferases.

A) Typical S_N2 reaction as assumed for most inverting GTs. B) S_N1 mechanism as hypothesised for few recently found inverting GTs. C) Discarded double-displacement mechanism of retaining GTs, based on glycosidases. D) Proposed S_Ni-like mechanism of retaining GTs (Moremen & Haltiwanger, 2019, adapted). Thanks for permission of reuse from Springer Nature.

linkage type. Moreover, a deep evolutionary analysis of GT-A proteins revealed multiple lineages for inverting and retaining mechanisms (Taujale et al., 2020). The inversion of the stereoisomeric centre is explained by a simple S_N2 reaction in which the enzyme activates the acceptor and displaces the leaving group of the donor by a nucleophilic attack. However, a S_N1 mechanism has recently been discussed for few enzymes. The reaction under retention is more complicated. In glycosidases the first step is the displacement of the leaving group by a sidechain nucleophile. Subsequently, the acceptor activated by the protein attacks again nucleophilic from the front. Thus, two S_N2 reactions switch the configuration back and forth and comprise a covalently glycosylated protein intermediate. However, after long debate in the glycosyltransferase field, this was disregarded in favour of the S_Ni-like mechanism. The donor phosphate deprotonates the acceptor hydroxyl group thus further polarising the anomeric carbon. The hydroxyl group then performs a nucleophilic attack on that carbon. The protein is predominantly required for correctly orienting the substrates and the subsequent stabilisation of the leaving group by a positive charge. It is not directly involved in the mechanism (S. S. Lee et al., 2011).

As initially mentioned, CAZy defines glycosyltransferase families by sequence similarity. Therefore, same fold and mechanism

are anticipated within one family. GgpS is a member of the GT20 family. The only other members are the α,α-trehalose phosphate synthase (TPS or OtsA) and its products' phosphatase (OtsB). GgpP (StpA), the phosphatase of the product of GgpS is not characterised by CAZy. GgpP is in general not well researched. To date, Uniprot lists only 23 entries of which only the one from *Synechocystis sp.* PCC6803 has evidence at the protein level. All others are only predicted or inferred from homology. Besides GgpS and GgpP, some organisms have a protein called GgpPS. Blastp indicates that this is a form of GgpS with an additional N-terminal phosphatase domain. Again, all four Uniprot entries are only inferred from homology. GgpS and GgpP share no sequence similarity and the latter might not has the GT-B fold as it is strictly speaking not GT at all. It transfers a glycerol moiety to a water molecule, while phosphatase OtsB transfers a glucose moiety. However, both do not alter a glyosidic bond. As a side remark, a Uniprot search should not be mistaken with geranyl-geranyl pyrophosphate synthase, which is often abbreviated as GgpS as well.

The structure of TPS from *E. coli* bound to a transition state mimetic was one of the first pieces of evidence indicating the S_Ni-like mechanism to be correct (Errey et al., 2010; S. S. Lee et al., 2011), and since it has the GT-B fold, GgpS most likely assumes the GT-B fold as well. It further indicates that GgpS may not dependent on Mg^{2+} ions.

1.4 The GgpS project

It was Kay Marin with Reinhard Krämer and his team who worked on GgpS and brought this project to us. One of their important publications was about the inhibition by DNA and reactivation or rescue by high salt concentrations (Jens F Novak et al., 2011). Before the group disbanded due to retirement of Reinhard Krämer, they asked us for a collaboration to crystallise the protein and to solve the apo-structure and the complex with DNA. Since the Krämer group was able to purify the protein, crystallisation was thought to be of lower difficulty. After Benjamin Roenneke, a PhD student of their lab graduated, most

tasks were carried out by bachelor students (Monique Henschel, Georg Rabbow, Seyma Bozkus, Leon Wehrhan) as well as master a student (Monique Henschel), who were supervised by Jan Gebauer and myself. Soon it turned out, that crystallisation was much tougher than anticipated. But after exchanging the C-terminal affinity tag against an N-terminal one, crystals started to grow and I assumed more and more responsibility over the project.

1.4.1 Wait, there is a chloride! What now? – The aims of the project

As a compatible solute producing enzyme, GgpS does not have to be active all the time but must rapidly become active under conditions of salt shock, i.e. usually a strong increase of intracellular sodium chloride concentrations. The commonly assumed mechanism behind this was binding of the enzyme to nucleic acids during low salt conditions by occlusion. In the complex, the enzyme is thought to be inactive. Upon increasing intracellular salt concentrations in salt shock scenarios, the ions were thought to shield electrostatically GgpS from the nucleic acids and thus releasing and reactivating it.

The initial scope of the project changed after solving the protein structure and finding two potential chloride ions bound to the tetramer. It challenges the established theory about salt regulation, since those sites may act as chloride sensors and directly regulate the activity bypassing any nucleic acid interactions. A brief literature research also reveals an inconsistency in the DNA-inhibition hypothesis: The cytosol at low salt cultivating conditions contains ions, predominantly potassium, at roughly 100 mM ionic strength (Reed et al., 1985). However, 100–150 mM ionic strength added in the form of sodium chloride to GgpS *in vitro* is enough to lift the DNA inhibition (Jens F Novak et al., 2011). So, how can the inhibition be possible if it is purely ionic strength dependent and independent of the ion species? This part of my thesis concentrates therefore on these potential chlorides and the aims were redefined as follows:

1. To validate the ion species unambiguously (as chloride)
2. To characterise the structural effects of non-, half and full chloride occupation
3. To correlate the structural effects of chloride binding with enzymatic activity
4. To re-evaluate the nucleic acid regulation mechanism
5. To co-crystallise GgpS with DNA and to identify the binding mode

In order to better assay chloride binding and its effects, we generated a binding-deficient mutant. Since the anions bind via the amino subgroup of the amide bond of Trp110, only the mutation to proline (W110P) was feasible. Fortunately, the actual Ramachandran angles of Trp110 are already located in the allowed regions for prolines, indicating the mutation to have only a minor effect on the overall structure.

Sucrose turned out to be the best cryo-protectant for GgpS, probably because it binds at several positions within the structure. In addition, we analysed crystals soaked with ADP and *sn*-glycerol-3-phosphate cryo-protected with mineral oils. Moreover, I crystallised GgpS in the presence of sulphate since some biochemical data suggested that only one sulphate can bind per tetramer. This could enable us to characterise the GgpS bound to only one anion, in contrast to the anion-free W110P mutant and the two-chloride occupied wild type.

2 GGPS – RESULTS AND DISCUSSION

2.1 Purification

A Strep-tag with a linker containing a TEV cleavage site was added to the N-terminus of GgpS from *Synechocystis sp.* PCC6803 (Uniprot: P74258). In addition, the construct I obtained had its very last amino acid removed. This version of the wild type protein and the corresponding W110P mutant version were both heterologously expressed in BL21(DE3) *E. coli* cells. The bacteria were lysed twice at 2.5 kbar using a Cell Disruptor (Constant Systems) and GgpS proteins were further purified via a Strep-tag:Strep-Tactin affinity and Superdex 200 size exclusion chromatography. A highly critical step is the removal of nucleic acids on the affinity chromatography column. When the protein was bound to Strep-Tactin or to Ni-NTA by using a different construct with a His_6-tag, nucleic acids have to be removed with up to 30 column volumes of buffer containing 500 mM sodium chloride. In theory, only the positively charged Ni-NTA interacts with nucleic acids, but in our experience with other proteins, Strep-Tactin does as well. After this washing step, the protein appears reliably monodisperse as a tetramer every time if kept in 150 mM sodium chloride and 10 mM HEPES/NaOH pH 7.6 or for assays 30 mM HEPES/KOH pH 7.6 (Figure 5). The tag could have been removed by cleavage with TEV protease but that has never been done. Moreover, the tag is probably required for crystallisation.

The protein was always produced on the same day of usage for biochemical assays. Freezing in the size exclusion buffer usually led to loss of around 15 % of the protein after thawing and centrifugation. A fast (4 h) small scale purification with a smaller affinity column (0.5 ml) for the assays yielded 0.2 mg protein per 1 g bacterial pellet. The process was optimised for fast production of fresh protein for the assays of the current day and sometimes for the next day (stored at 20°C). Larger scale purifications were more efficient with 1.2–2.1 mg/g due to usage of considerably
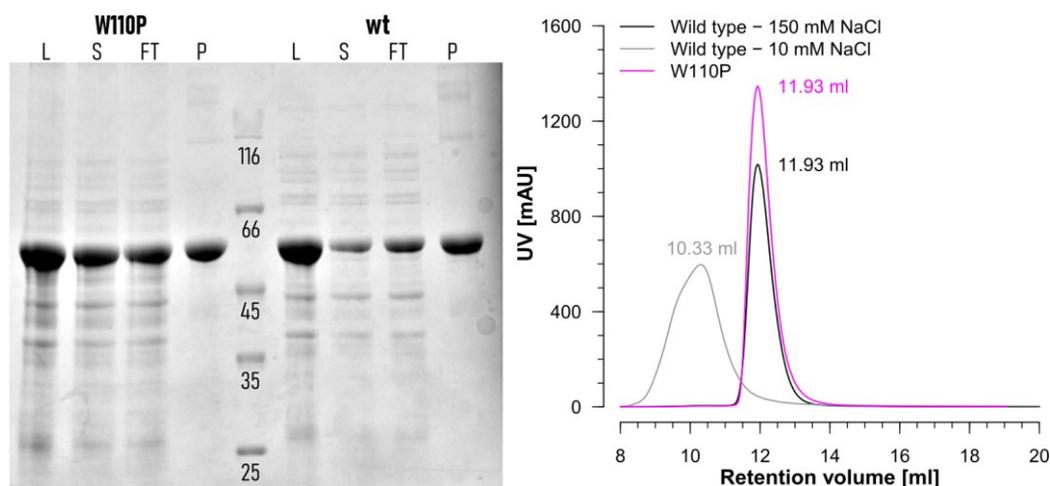
Figure 5: Purification of GgpS wild type and W110P mutant.

The proteins were purified via Strep–tag:Strep–Tactin affinity and size exclusion chromatography (Superdex 200). The Coomassie stained gel depicts the lysate (L), supernatant (S), affinity chromatography flow–through (FT) and the final pooled and concentrated protein after size exclusion chromatography (P). Size: 59.30 kDa, runs between 45 kDa and 66 kDa. The plot depicts the size exclusion chromatograms at 150 mM or 10 mM sodium chloride and 20 mM HEPES/NaOH pH 7.6. The proteins were usually concentrated with a 100 kDa molecular weight cut–off ultrafiltration device, hence lower sized particles might have been removed beforehand. Yet, no new monomeric GgpS was formed. As comparison, a 47 kDa protein elutes usually from this column at around 16 ml.

slower but larger affinity columns (2 ml) with same pellet size. An expression usually yielded 1.9 g pellets per 1 l culture in LB medium.

2.2 Overview and general description of the crystallographic models

2.2.1 Crystallisation and data processing

Here I give a brief overview of all important crystallographic models. All data about used proteins, crystal type, data collection, reduction and refinement can be found in Table 12. The crystals, datasets and models are named by the used protein (wild type and W110P mutant) as well as an index composed of the anion present during crystallisation and the PDB code of the ligand supplied during soaking (SUC: sucrose, ADP: adenosine diphosphate (glucose), G3P: sn-glycerol-3-phosphate, GOL: glycerol). $\text{W110P}_{\text{Cl,SUC}}$ for example refers to the crystal of the W110P mutant grown in presence of chloride and sucrose. However, the ion may not by present in the final crystallographic model. A slight

exception is ADP-glucose. Only the ADP moiety was found in the active site and the models are only have an 'ADP' in the index.

The proteins were purified as explained in section 2.1. Crystals were grown under several but similar conditions as listed in Table 12. If crystallisation with a different anion than chloride was desired, the sodium salt of that ion was used at latest during size exclusion chromatography with the same concentration. In addition, crystallisation conditions with chloride ions were avoided or in case of an employed Tris buffer, it was titrated with malonic acid instead of hydrochloric acid.

Most crystals appeared morphologically as columns, sometimes with a recognisable hexagonal base. Initial crystals often had hollow tips und were used after 1–2 days for micro seeding of conditions with less precipitant and no crystals yet. Under sulphate conditions, crystals grew as simple prisms with hexagonal base and the W110P mutant crystals predominantly in needle clusters. From the latter, individual crystals were detached and mounted.

All crystallization conditions turned out to require further cryo-protection. Most crystals were eventually cryo-protected in a 50 % saturated sucrose solution made from reservoir solution or reservoir solution of an adjacent condition. Crystals soaked in ADP-glucose and G3P were protected with Paratone N or paraffin oil. The very first dataset ($wt_{Cl,suc}$) was phased by molecular replacement with coordinates from PDB entry 2wtx, a homologous GT20 glucosyltransferase with 31 % sequence identity (Errey et al., 2010).

An inconsistency at position 463 was found during model building of the initial wild type structures. The electron density is clearly indicating a small amino acid like cysteine or serine, perhaps the former due to shape, however, by sequence an arginine was supposed to be there. The anomalous dataset of $wt_{Cl,ADP}$ further indicated the presence of an anomalous scatterer at that position. A thorough analysis of the last sequencing result eventually revealed a potential mixed plasmid stock of wild type and the unintentional mutant R463C. This cysteine mutation was probably

also present in the preceding plasmid preparation and hence persisted in very low and unnoticed amounts two such purifications from *E. coli*. During expression, however, that mutant was picked or at least predominantly expressed. All assays and crystallisation trials in bromide were performed with a corrected wild type. The eventually used proteins are listed in the dataset statistics (Table 12) as well.

2.2.2   Asymmetric unit: GgpS is a proper member of the GT20 family

GgpS wild type with chloride ($wt_{Cl,SUC}$) and bromide ($wt_{Br,SUC}$) as well as the W110P mutant in chloride buffer ($W110P_{Cl,SUC}$) crystallised in space group $P6_{5}22$ with cell dimensions of $a = b = 104$ Å and $c = 364$ Å. Datasets were truncated to 1.88 Å, 1.80 Å and 3.35 Å resolution, respectively. The asymmetric unit is composed of two copies of GgpS and contains about 50 % solvent.

The first GgpS dataset was solved by molecular replacement with a model of the well-studied trehalose-6-phosphate synthase (TPS) from *E. coli* (PDB: 2wtx). Sequence identity is mediocre with 31 % for 397 out of 499 amino acids (blastp). However, the structure is conserved. As TPS, GgpS has two Rossman fold-like subdomains, where the active site is found in the cleft between both. The N-terminal domain has ten β-strands and the C-terminal domain has six strands (Figure 6). While aligning the first chains of $wt_{Cl,SUC}$ and TPS with Chimera, 313 C_{α}-atom pairs have an root mean square deviation (RMSD) of 0.99 Å. The structures differ primarily in the N-terminal regions up to position 84 and between residues 276–293 of GgpS (TPS: 1-68, 238–254). At least the first region coincides with strong sequence dissimilarity (Figure 69). On the other hand, regions of higher similarity include foremost the active site. GgpS has an additional two strand β-sheet (β1/2) in positions 14–26 prominently protruding the structure and two extended β-strands in the N-terminal Rossman β-sheet (225–246,
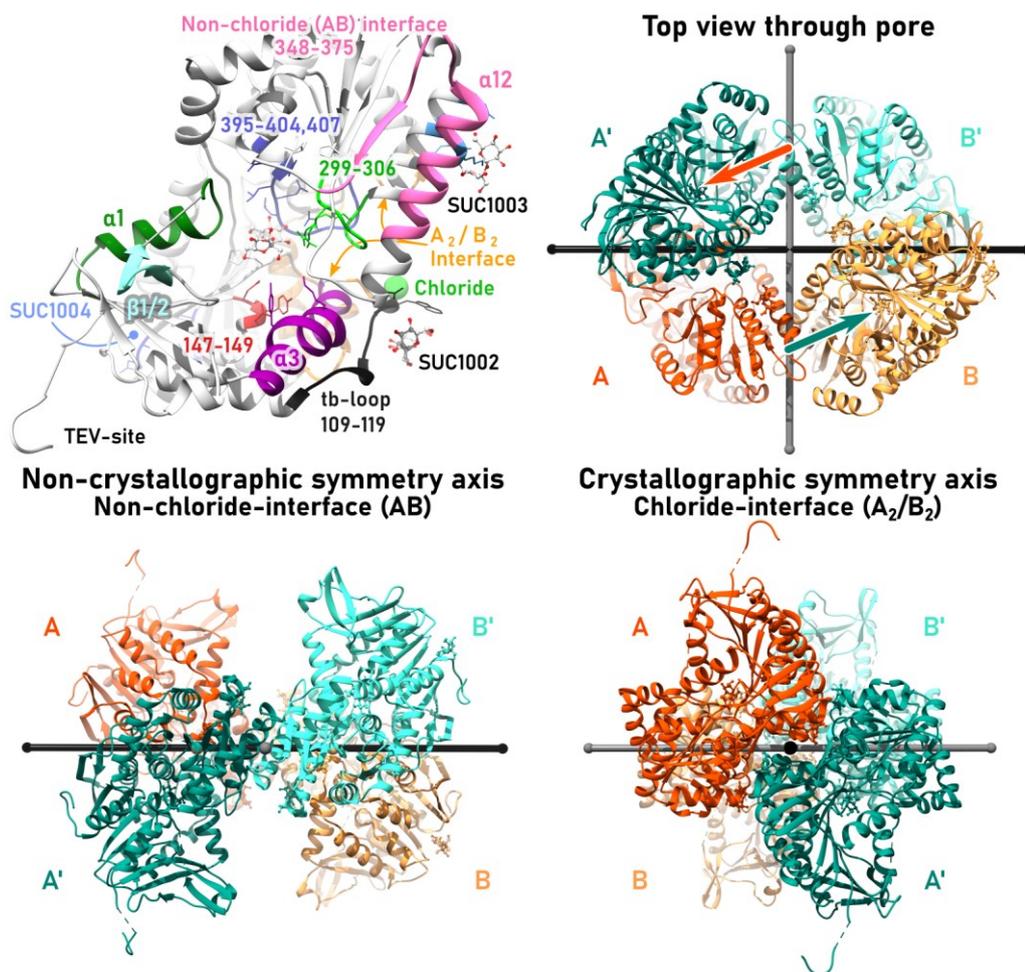
Figure 6: The monomer and the tetramer of $wt_{Cl,SUC}$ with highlighted features and symmetry axes, respectively.

Representatively depicted is chain A, which has no SUC1004, however, its position in chain B is indicated. The monomer is shown with view into the active site with SUC1001. This perspective is indicated by arrows in the top view of the tetramer. The other two sites are accessible from the bottom. The asymmetric units are coloured in either cyan or orange while chains B are brighter than chains A. The crystallographic twofold axis through the bound chlorides is coloured in black, whereas the non-crystallographic axis is grey.

TPS: 208). Both insertions are partially visible in the alignment (Figure 69).

The asymmetric unit of the $P6_522$ crystals has two GgpS molecules. Both are virtually identical with an RMSD of less than 0.33 Å for all C_α-atoms excluding 18 pairs with distances larger than 2 Å (calculated using Chimera and $wt_{Cl,SUC}$). The structure with bromide is remarkably similar, with RMSD values as low as 0.116 Å and 0.114 Å across all C_α-atom pairs for the molecules A and B, respectively if compared with $wt_{Cl,SUC}$. Moreover, the average residue B factors are very similar at each position. However, the

electron density map of the bromide bound model was subjectively superior, and this model was finally used to phase all other datasets. Most images depicted here represent the $wt_{Br,suc}$ model.

As mentioned in the introduction, initial crystallisation experiments with a C-terminally tagged GgpS failed. The crystals appeared were very irregular and intergrown with weak diffraction properties. After switching the tag to the other terminus, better crystals grew. The built model explains why changing the terminus was important. The complete TEV-cleavage site between the Strep-tag and GgpS is indeed part of a crystal contact. It protrudes from the tetramer and probably facilitates crystal growth (Figure 6, bottom left of monomer and the 'hooks' protruding the tetrameric views at the bottom).

Large peaks were found in the electron density map near Trp110. The initial interpretation as chlorides could be confirmed via anomalous scattering difference density maps (section 2.2.4). These chloride anions are bound to the amino subgroup of the peptide bond linking Pro109 and Trp110. They are localised on one of the twofold crystallographic axes and were constrained to that special position. Due to the symmetry also Trp110' of the symmetry equivalent GgpS subunit interacts with the anion. In contrast, the W110P mutation successfully prevents anion binding of any kind at this position and we concluded that it is a good model for a chloride-free GgpS (section 2.3.3).

Cryo-protection with sucrose led to incorporation of up to seven sugar molecules bound by both GgpS molecules of the asymmetric unit. Four different sites were identified: The true active site (SUC1001, Figure 13) and secondary sites at $\eta 1$ near the chloride binding site (SUC1002, Figure 8), at $\alpha 11$ and $\alpha 12$ (SUC1003, Figure 7) and at the loop between $\alpha 5$, $\beta 8$ and the C-terminus (SUC1004, Figure 7). The latter is only found in chain B and its electron density is clearly visible but less pronounced indicating a lower occupancy or flexibility in general. However, it is part of a crystal contact which does not occur in chain A. Therefore chain A has only three bound sucrose molecules. In the W110P mutant this sucrose
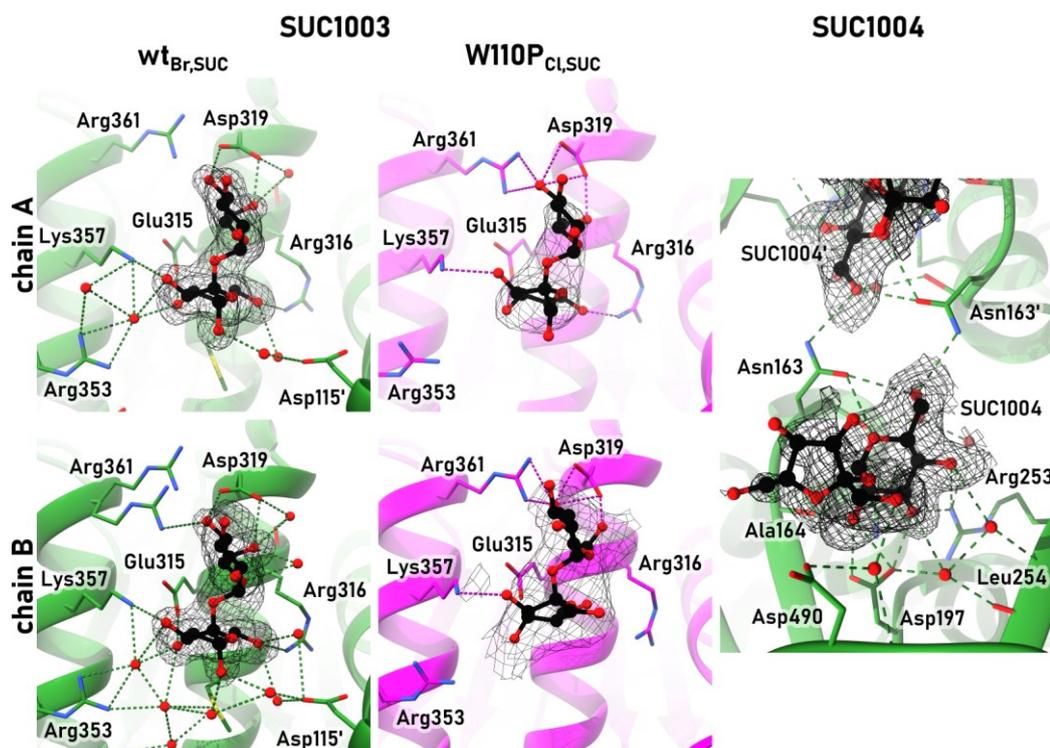
Figure 7: SUC1003 and SUC1004 sites.

Three of the seven sucrose sites in space group $P6_522$. The fourth site is only present in chain B and while there is electron density justifying the position and placement of the fructose moiety, it is weaker. Polder maps (Liebschner et al., 2017) of the wild type are depicted at 3 σ RMSD and 1.5 σ for the W110P mutant owing to its lower resolution. Polar interactions below 3.5 Å distance are indicated by dashed lines.

site cannot be seen probably due to lower resolution, but also the sucrose in the active site of chain B is likely not present since its position there is taken by amino acids 399–403 (section 2.4.1).

Crystals soaked with 10 mM ADP-glucose (wt$_{Cl,ADP}$) and 6 mM G3P (wt$_{Cl,G3P}$) yielded similar models as with sucrose. However, the large c-axis was reduced from 364 Å to 350 Å and 337 Å, respectively. Only the ADP moiety of ADP-glucose was discernible in the active site. The glucose part was either hydrolysed or not visible at 3.30 Å resolution. Since the dataset was measured at a wavelength of 1.70 Å, the general phosphate location and ADP orientation could be verified by anomalous dispersion. Glycerol-3-phosphate was found bound at the 3_{10}-helix η1 near the chloride binding site at the position of SUC1002 (Figure 8). Since sucrose and glycerol-3-phosphate both appear there, this can be a potential allosteric inhibition site for compatible solutes. Several mono- and
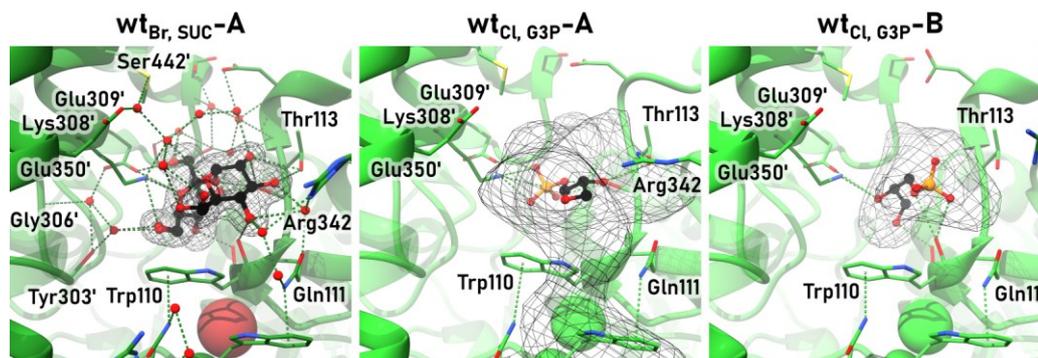
Figure 8: SUC$_{1002}$ site can also be occupied by glycerol–3–phosphate.

Red–brown sphere: Bromide; Green sphere: Chloride.

disaccharides inhibit GgpS (B. M. E. Roenneke, 2014). In addition, very weak density appeared in the active site. This may be the phosphate group of the G3P, however, resolution with 3.75 Å is too low to place any small molecule confidently at this position.

Crystals not treated with sucrose diffracted worse than sucrose-soaked crystals in general. The very best was indeed completely ligand-free but with both chloride anions bound in the structure (wt$_{Cl}$). The crystal diffracted to 2.77 Å and its cell shrunk from 104 Å, 104 Å, 364 Å to 102 Å, 102 Å, 344 Å.

Purifying with and crystallising in sulphate while cryo-protecting with either sucrose or glycerol (wt$_{SO4,SUC}$ or wt$_{SO4,GOL}$) yielded GgpS crystals in space group P6$_1$22 with cell dimensions of 157 Å, 157 Å and 602 Å or 155 Å, 155 Å and 592 Å, respectively. They diffracted to 3.43 Å and 3.20 Å but suffered from anisotropy with diffraction limits of about 3.0 Å to 4.0 Å. With the different cell, also the asymmetric unit changed. It contains six copies of GgpS with 59.4 % solvent. At the given resolution, no sucrose or glycerol molecules were found. Furthermore, although crystals grew in sulphate, none were incorporated into the model eventually due to ambiguous electron density as detailed in section (2.4).

2.2.3 Biological unit: Always a tetramer, but chloride dependent?

While there are only two molecules in the asymmetric unit, the protein appears to be tetrameric in solution (SEC-MALS in collaboration with Lutz Schmitt, University of Düsseldorf). The two copies of GgpS within the asymmetric unit (A and B) of space group P6$_5$22 form a non-crystallographic C2 dimer (AB) via the α12β15 interface

Table 1: Total and buried surface area, ΔG^{int} as well as ΔG^{diss} for all assemblies of models wt_{Cl} and $W110P_{Cl,SUC}$ (sucrose removed) as analysed by PDB ePISA (Krissinel & Henrick, 2007).

Assemblies with a $\Delta G^{diss} > 0$ are considered thermodynamically stable. There were no stable assemblies detected in the sulphate models.

Protein	Assembly	Total Surface [Å^2]	Buried Surface [Å^2]	ΔG^{ins} [kcal/mol]	ΔG^{diss} [kcal/mol]
Wild type	$A_2B_2[Cl]_2$	68,990	11,540	−72.0	14.7
Wild type	A_2B_2	69,230	11,050	−34.7	1.7
Wild type	$A_2[Cl]_2$	35,470	3,920	−21.8	13.4
Wild type	$B_2[Cl]_2$	37,310	3,840	−22.0	14.2
Wild type	A_2 or B_2		Not stable		
Wild type	AB	38,240	1,890	−14.1	0.9
W110P	AB	40,230	1,850	−12.0	−1.7

(Figure 6). It predominantly includes residues 348–375 that belong to this α-helix and β-strand and is relatively small with 1,890 Å^2 (Table 1). The AB-dimer lies parallel to the same crystallographic twofold axis on which the chloride ions are located. The final (partially crystallographic) D2 symmetric GgpS tetramer is constituted by the bespoken dimer and its crystallographic symmetry mate, bridged by the ions (Figure 6). However, the entire AB:AB interface comprises an area of 7,760 Å^2 ($A_2[Cl] + B_2[Cl]$: 3,920 Å^2 + 3,840 Å^2) of which the ion-bridges are only a part off. Moreover, the interface is effectively split in two areas, each consisting of the symmetry equivalent dimers only. The tetramer forms a pore with a diameter of smaller than 20 Å while the tetramer is around 100 Å in diameter; measured along the crystallographic twofold axis.

Besides buried surface area, PDBePISA estimates the ΔG^{diss}, which 'indicates the free energy of assembly dissociation' in solution (Krissinel & Henrick, 2007). Assemblies with $\Delta G^{diss} > 0$ kcal/mol are considered thermodynamically stable. According to the analysis, the dimers across the ion bridge ($A_2[Cl]$: 13.4 kcal/mol; $B_2[Cl]$: 14.2 kcal/mol) are more stable than the dimer found in the asymmetric unit (AB: 0.9 kcal/mol). Altogether, the tetramer with ions is judged stable with 14.7 kcal/mol. In contrast, the ΔG^{diss} of the tetramer whose chloride haven been removed decreases to 1.7 kcal/mol. This still indicates the tetramer to be stable but raises the question about ion/chloride dependent
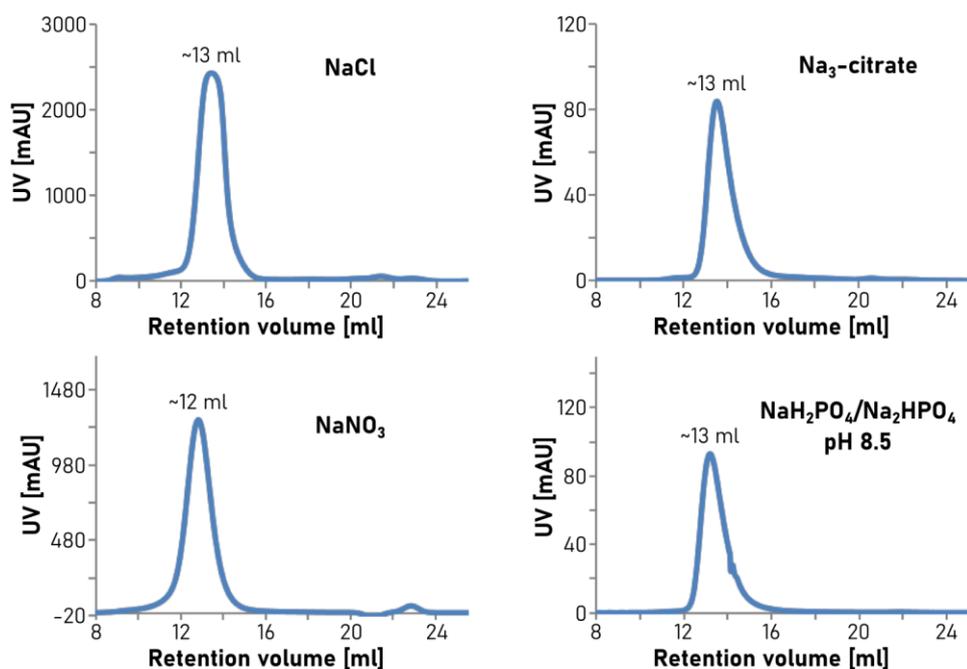
Figure 9: Size exclusion with sodium salts of chloride, citrate, nitrate and phosphate.

Adapted from master's thesis of Monique Henschel (2017).

tetramerisation. Accordingly, PDBePISA cannot find any stable assembly for the W110P mutant apart from the AB dimer which 'may or may not be stable in solution' with a value as low as -1.7 kcal/mol. However, the mutant behaves identical to the wild type during size exclusion chromatographies and the wild type in presence of only 10 mM sodium chloride has unexpectedly a lower retention volume corresponding to octamers (Figure 5). This higher oligomerisation was reversible and judged as less important. Size exclusions chromatography runs in presence of the sodium salts of citrate, nitrate, and phosphate instead of chloride demonstrated no change of the oligomeric state, thus no chloride dependent tetramerisation (Figure 9). Hence, low ionic strength potentially leads to higher oligomerisation but there is no detectable chloride dependent association.

The activity and chloride-interaction assays (section 2.5.3 and 2.5.4) yielded unexpected results and let me revaluate the octamerisation. A simple dynamic light scattering experiment confirmed the octamerisation, however by roughly doubling the HEPES/KOH content I could almost completely suppressed this phenomenon (Figure 10F). The implication of this for the
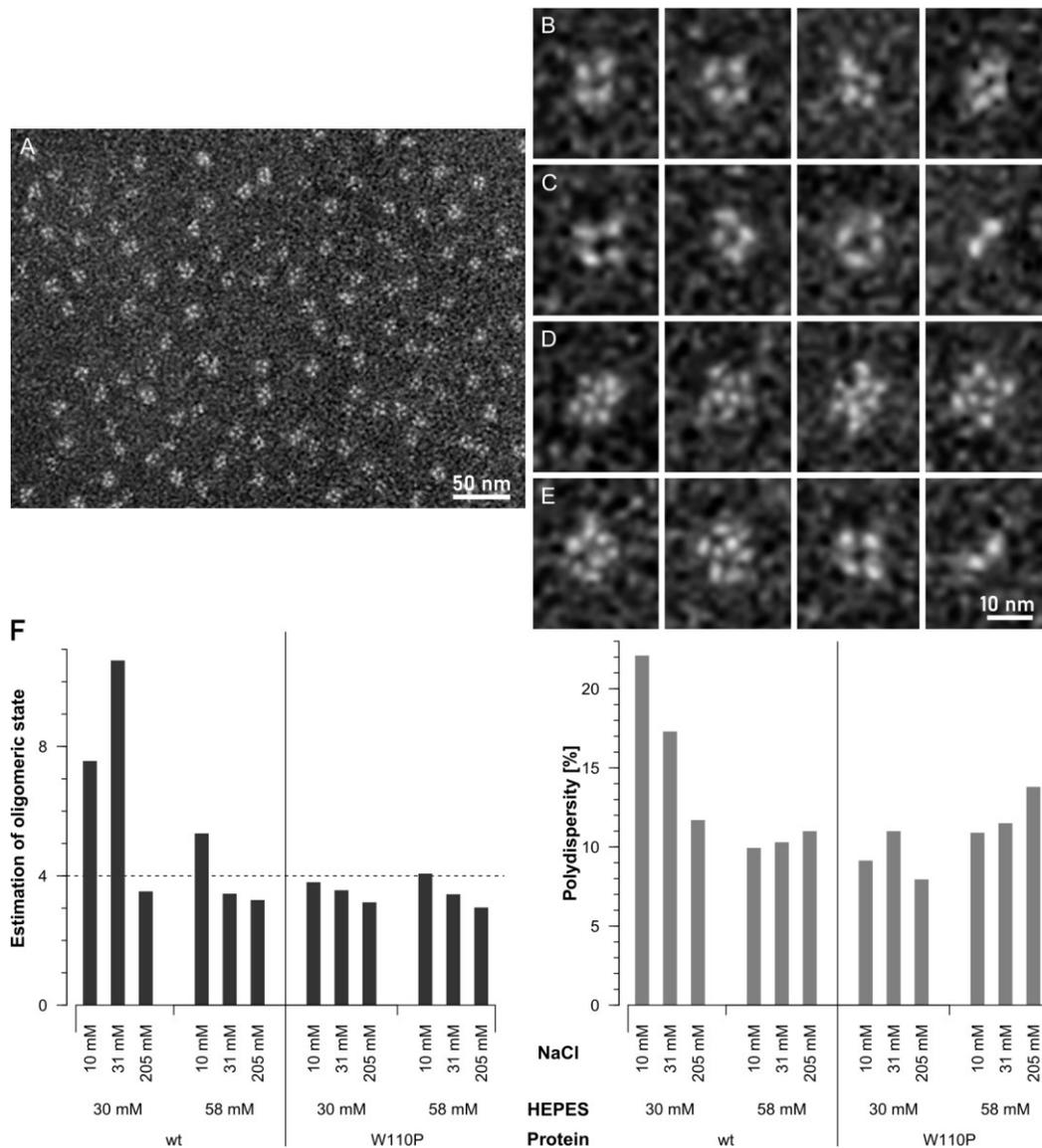
Figure 10: Negative stain electron micrographs and dynamic light scattering results of GgpS wild type and W110P mutant at low and high salt conditions.

A) Overview of the wild type at 150 mM sodium chloride and 20 mM HEPES/NaOH pH 7.9. B) Focus on four particles of A). C) Focus on particles of the W110P mutant in the same buffer. D) Wild type particles at approximately 5 mM sodium chloride and 20 mM HEPES/NaOH pH 7.9. E) W110P mutant under same low salt conditions as D). F) Dynamic light scattering results of both proteins at different sodium chloride concentrations and HEPES/KOH pH 7.6 background. The HEPES background was chosen to be identical to chloride interaction assay (30 mM) and similar to the activity assay (58 mM). Similar means assay substrates were exchanged against HEPES so the ionic strengths matches, assuming all HEPES molecules have a singular charge. However, half the HEPES molecules at pH 7.6 have one negative charge or a negative and a positive charge. The latter are excluded from the actual ionic strength calculation, yet may interfere with octamerisation.

interpretation of the assays is discussed in the summary and conclusion section (3.4.1). A very recent and so far preliminary negative stain electron microscopy analysis of GgpS at 20 mM

HEPES/NaOH pH 7.9 and roughly 5 mM sodium chloride by Matthias Mörgelin (Colzyx AB, Lund, Sweden) further supports the octamerisation. The wild type forms predominantly octamer-like oligomers (Figure 10A, B, D). However, the mutant exists apparently as octamers and to some extent as tetramers and dimers simultaneously at low salt conditions (Figure 10E). These dimers are also visible at 150 mM sodium chloride (Figure 10C). There is, as of now, no statistical evaluation. Dimers were not detected in the dynamic light scattering experiment at all.

The crystal structures of GgpS in space group $P6_122$ (grown in the presence of sulphate) do not differ strongly from those in $P6_522$. The asymmetric unit is composed of six copies and independent of the used cryo-protectant, tetramers are easily recognisable. There are two different D2 tetramers in this space group. The first has no crystallographic symmetry (ABCD) and the second utilises again one crystallographic twofold axis (EF; Figure 70). However, this symmetry is different: prior, the molecules across the ion bridge were crystallographically equivalent; now the molecules across the weak $\alpha12\beta15$ interface are equivalent. A PDBePISA assembly analysis was also employed with these models and no stable assemblies are to be expected in solution. However, the analysis may be compromised because, the anion binding sites have become more flexible and amino acids have often been deleted due to insufficient electron density. Moreover, the structure was modelled at 3.43 Å with underlying anisotropy, hence side chains conformations are not very reliable due to high B factors.

The general constitution i.e. the orientation of each wild type monomer in the tetramer with chloride/bromide is usually not depended on the presence/absence of ligands, as visualised in Figure 11 ('sulphate crystals': Figure 70). The tetramers were aligned only on their subunit A and the deviation of the mainchain atoms of each residue from the reference tetramer ($wt_{Br,SUC}$) to the equivalent residue/subunit in the analysed tetramer was evaluated. An example: If chain B' (symmetry equivalent of B) of $wt_{Cl,ADP}$ is moved within the complex relative to chain B' of $wt_{Br,SUC}$
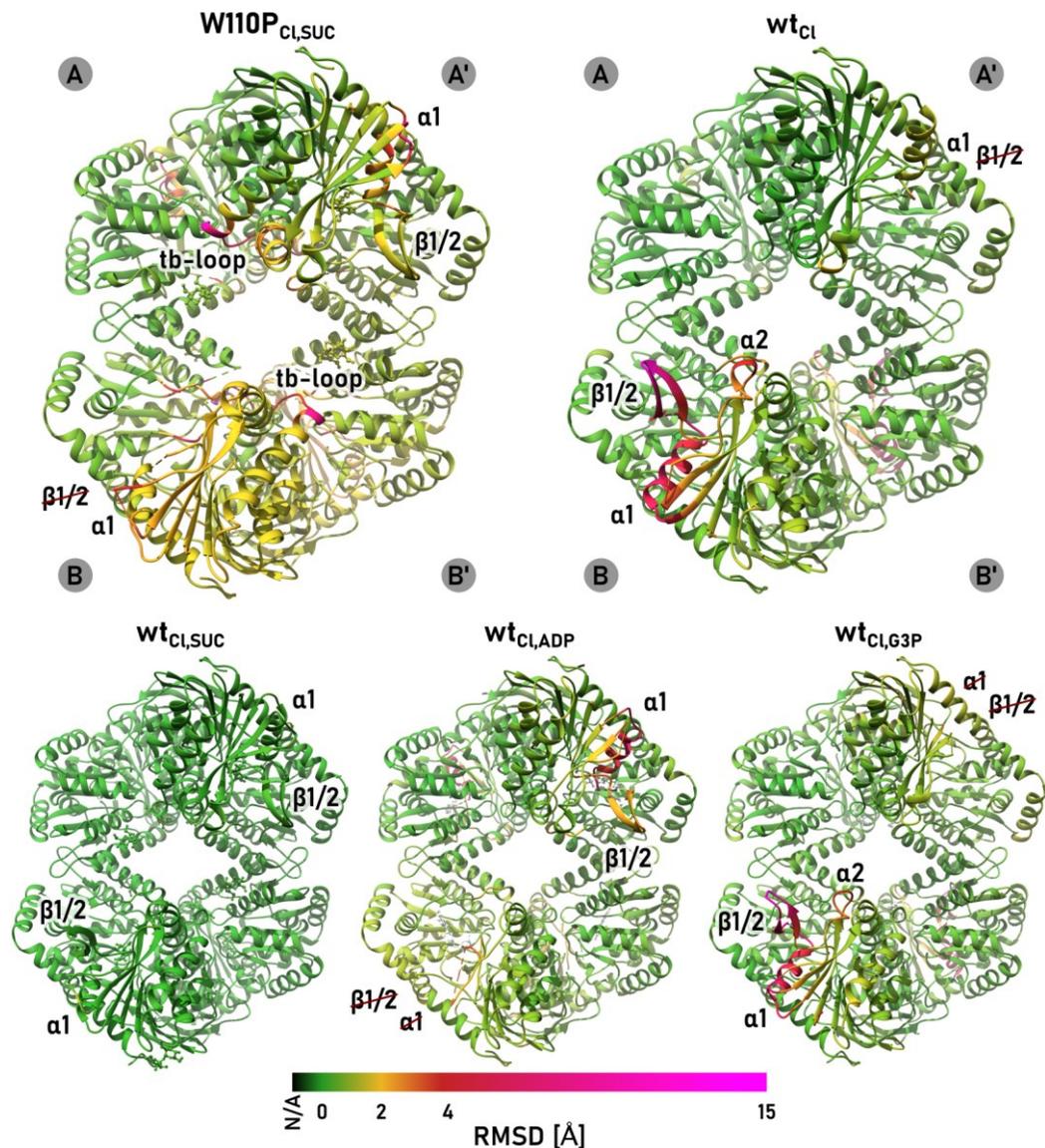
Figure 11: Deviations of the tetramers from the two-bromide bound state.

Grey circles identify the subunit. All tetramers were aligned to chain A (always top left) from the $wt_{Br,SUC}$ model. The RMSD of the mainchain atoms of each residue of each model from the equivalent chain in the $wt_{Br,SUC}$ model is visualised, e.g. deviation between wt_{Cl}-B' and $wt_{Br,SUC}$-B'. Non-sucrose bound tetramers have differently positioned β-sheet β1/2, α1 and α2. The $W110P_{Cl,SUC}$ tetramer deviates stronger on the opposite site of the alignment (chains B; bottom). Most deviations are found in the tension belt (tb) loop (closer to the pore) and α1 yet again. Not build elements are crossed out. The tb-loop is only annotated in the $W110P_{Cl,SUC}$. Annotations as in Figure 6. 'Suphate' structures: Figure 70.

it should exhibit a higher deviation in general and should be coloured more reddish over all. This not case, only isolated movements of for example sheet β1/2 and helix α1 become visible. However, subunit wide differences are indeed present in the W110P mutant with a magnitude of up to 2 Å RMSD and in the 'sulphate crystals' with 2–4 Å RMSD. There is the impression of a
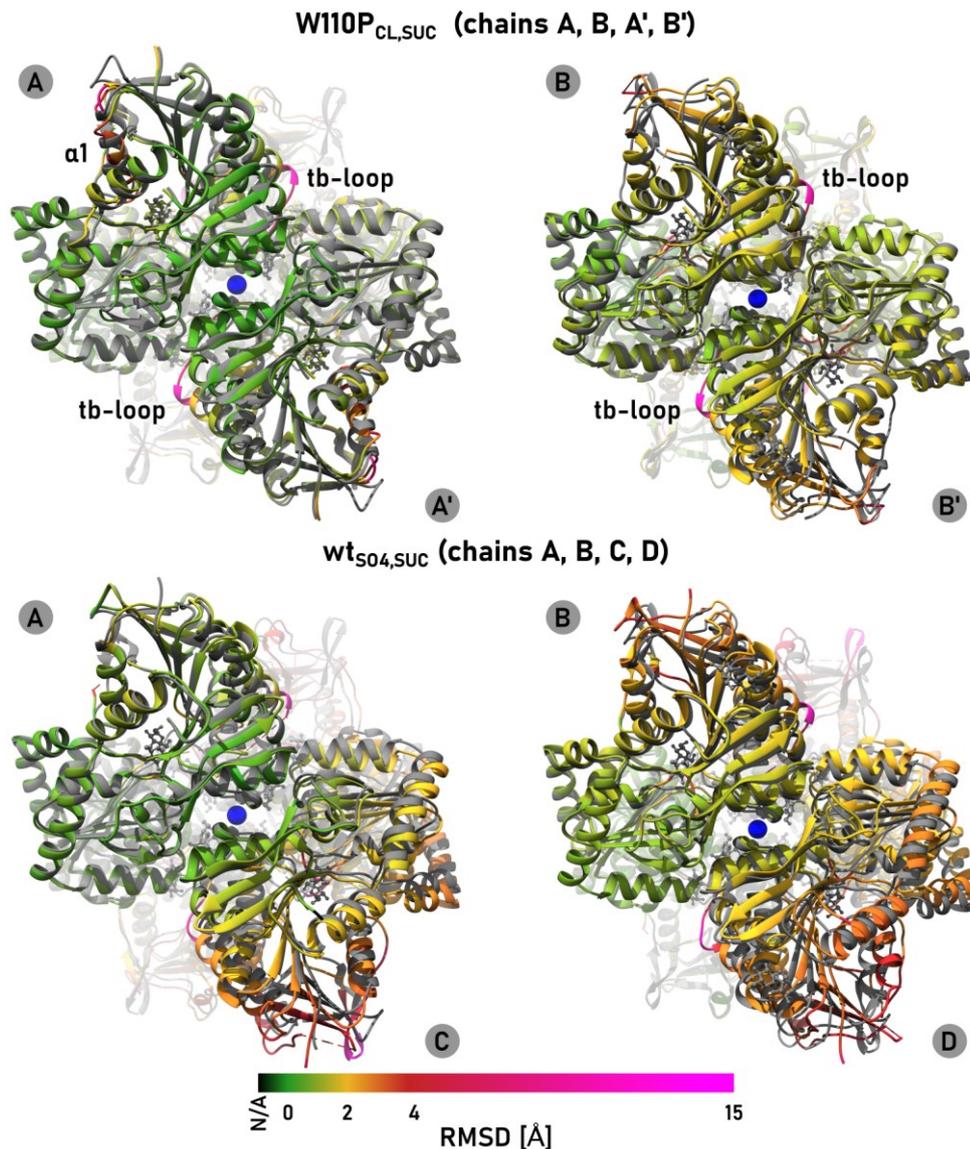
Figure 12: Deviations of the tetramers from the two-bromide bound state with focus on the anion binding site.

Grey circles with letter identify the subunit. Perspective in relation to Figure 11: From the bottom upwards. The bromide/chloride is the blue sphere in the centre of the dimer. All tetramers were aligned to chain A from the $wt_{Br,SUC}$ model. The RMSD of the mainchain atoms of each residue of each model from the equivalent chain in the $wt_{Br,SUC}$ model is visualised, e.g. deviation between wt_{Cl}-B' and $wt_{Br,SUC}$-B'. A rotation most prominent in the 'sulphate' crystals around the anion binding site is visible.

slight rotation around the anion-site connecting axis (i.e. crystallographic axis in the mutant) of the dimer consisting of subunit B and its symmetry mate B' or the equivalent BD dimer of the depicted $wt_{SO4,SUC}$, respectively (Figure 12). However, DynDom3D (Poornam et al., 2009) cannot detect any movements. Moreover, the resolution of those datasets is between 3 Å and 4 Å

with anisotropy in case of the wt_{SO4} crystals, hence the data are less reliable.

Taken together, GgpS wild type and W11OP mutant appear as tetramers during size exclusion chromatographies and indeed constitute D2 symmetric tetramers in the crystal. This D2 assembly in the crystals is usually a mixture of non-crystallographic and crystallographic symmetry. Most notably, losing both bound ions dramatically impairs tetramer stability as calculated by PDBePISA but does not lead to its breakdown or of the apparent symmetry as amongst others the W11OP mutant demonstrates. However, preliminary negative stain electron micrographs revealed dimers of the W11OP mutant independent of ionic strength, i.e. sodium chloride concentration. In addition, low ionic strength in general leads to octamerisation.

2.2.4   The active site is GT20 like

Most GgpS crystals were cryo-protected with sucrose and one sucrose molecule was usually found in the active site (SUC1001, Figure 13). As GgpS converts ADP-glucose and G3P to ADP and GGP I tried to soak either of both substrates into the crystals without sucrose cryo-protection. However, only ADP without glucose moiety was found in the active site ($\text{wt}_{Cl,ADP}$, Figure 13), while G3P only bound to the known second sucrose site at η1 ($\text{wt}_{Cl,G3P}$, Figure 8). The product GGP has never been co-crystallised or soaked into existing crystals. Therefore, the exact position of G3P or glucosyl-glycerol-phosphate in the active site is unknown. Nevertheless, it can be inferred from the homologue TPS and by docking. TPS catalyses the reaction of UDP-glucose and glucose-6-phosphate to UDP and trehalose-6-phosphate. The model 2wtx of TPS with UDP and a bisubstrate analogue, validoxylamine A 6′-O-phosphate (VDO), was used to study the S_{Ni}-like mechanism of retaining glycosyltransferases (Errey et al., 2010; S. S. Lee et al., 2011). A structural alignment of this and our sucrose as well as ADP bound models exhibits a great agreement in position of the respective equivalent ligands and residues (Figure 13).
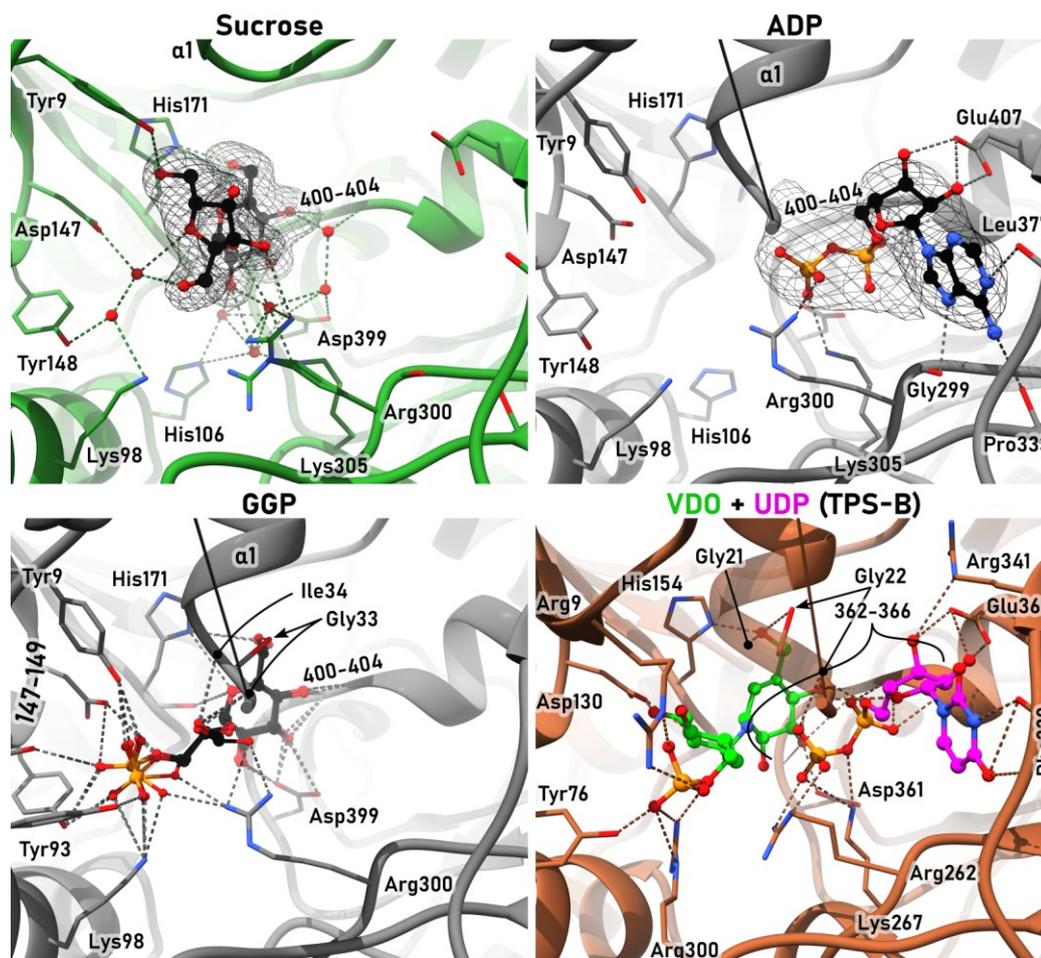
Figure 13: Active site of GgpS models wt$_{Br,SUC}$, wt$_{Cl,ADP}$, wt$_{Cl,DockG3P}$ (chain A) and TPS (chain B) (2wtx) with bound ligands.

The camera is positioned outside the protein with perspective into the binding pocket. α1 (top, middle) is forced upwards due to the fructose moiety, but slides down after correct substrate binding. The orientation of GGP is a docking result. Four of the best fitting molecules are shown (details in the text). TPS in 2wtx has bound the (pseudo-)products validoxylamine A 6'-O-phosphate (VDO) and UDP (Errey et al., 2010). Polder maps are modelled at 3 σ RMSD and polar interactions below 3.5 Å distance are indicated by dashed lines.

The glucose moiety of the sucrose is mostly bound by the sidechains or by the amides from the loop 399–402. In addition, histidine 171 binds to the O6, thus ensuring the orientation of the glucose. The binding mode is similar to the homologue TPS, where the loop of 361–365 and His154 the glucose moiety (Figure 14). Comparing the sucrose and ADP bound models emphasises a big but the only difference. The whole helix α1 is shifted roughly by a turn upwards, freeing space for the fructose moiety, which would otherwise overlap with its N-terminus (Figure 11, Figure 13, Figure 19). Further comparison with the ligand-free model reveals this movement to be a feature than just a sucrose induced artefact. α1
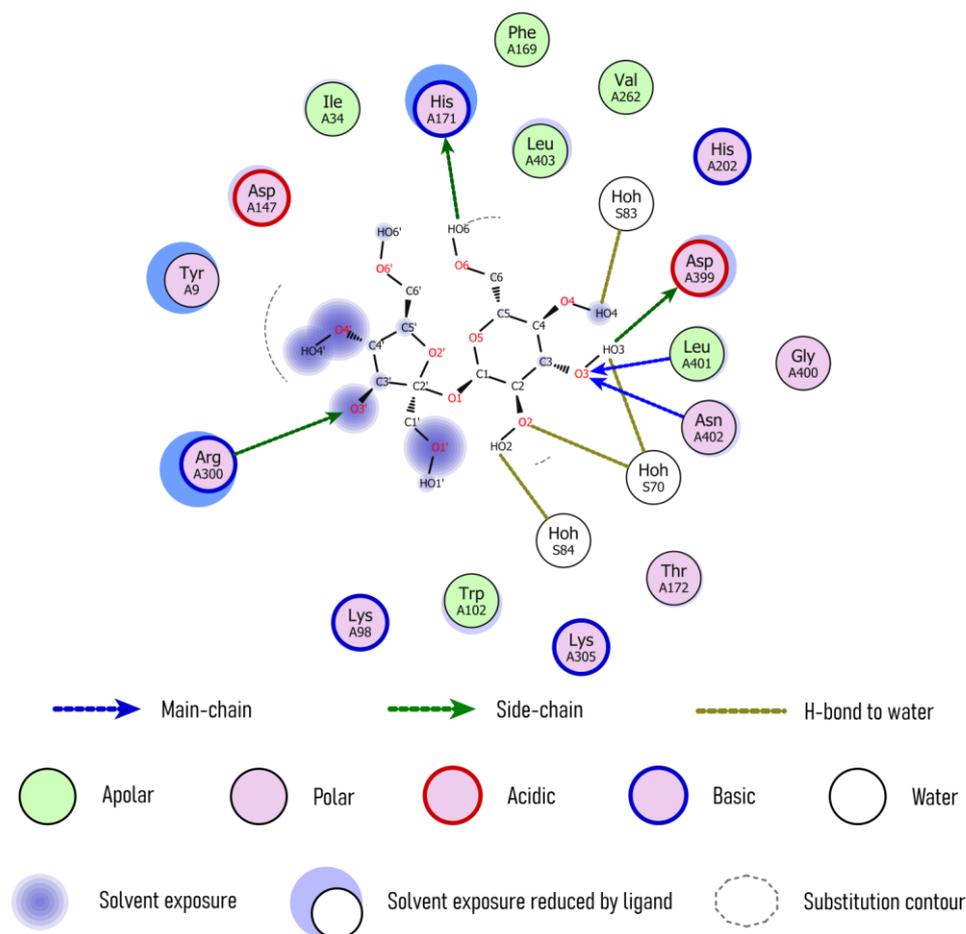
Figure 14: 2D representation of SUC1001 in the active site with coordination shell (by Coot LIDIA, Emsley et al., 2010).

can be in an open or closed conformation but it moves towards the substrates upon binding and locks them in place. In addition, β-sheet β1/2 and the linker to helix α1 change their positions/orientations accordingly but also independently to some extent (Figure 11).

Gly21 and Gly22 at the N-terminus of α1 interact with the phosphates of the UDP in TPS. This is helix dipole-nucleotide interaction is a common feature in retaining GTs (Lairson et al., 2008). In GgpS, Gly33 and Ile34 could take over this role but as currently modelled are too far away. However, resolution is low with 3.30 Å, the ligand bound is apparently only the product and the electron density is low so positions 27–31 could not built at all. Interestingly, one unique feature of GgpS compared to TPS is the β-sheet β1/2 constituted by amino acids 14–26. This hints to an unknown regulative mechanism linking a highly exposed flexible structural element to the active site. The phosphates are further
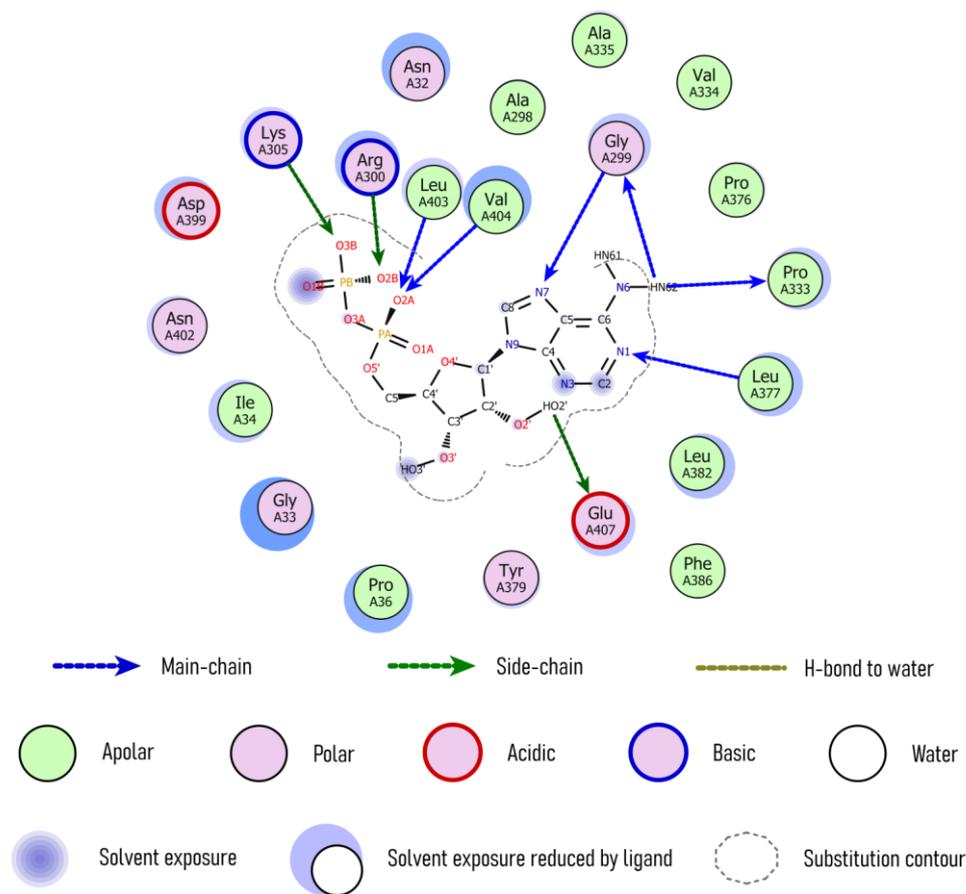
Figure 15: 2D representation of the ADP in the active site with coordination shell (by Coot LIDIA, Emsley et al., 2010).

interacting with Arg300, Lys305 and potentially Asp399 (TPS: Arg262, Lys267, Asp361) as well as with the amide amino subgroups of Leu403 and Val404 (Leu365 and Val366). The arginine and the lysine take on the role of leaving group stabilisation (Figure 15) and replace the Mg^{2+} cation in GT-B proteins. Similar is also the interaction of the ribose with the sidechain of Glu407 (Glu369) while Arg341 of TPS is substituted with the not interacting Tyr379 in GgpS. Since TPS utilises the pyrimidine derivative UDP as donor instead of the purine derivative ADP, coordination obviously deviates. The adenine is bound by the amide bonds of Gly299, Pro333 as well as Leu377 in GgpS.

In order to infer the interactions of G3P or the product glucosylglycerol-phosphate (GGP), I compared GgpS first with the homologue TPS. As described in the two paragraphs before, the glucose moiety and the nucleotide diphosphate moieties are bound in a very similar way by GgpS and TPS. Since the GGP is relatively

similar to the pseudo-product VDO in 2wtx, the phosphate group of G3P is probably at a similar position, i.e. in proximity to Tyr9, Tyr93, Lys98 and Asp147, based on similar positioned residues Arg9, Tyr76, Arg300 and Asp130 in TPS.

Next, using AutoDock Vina with UCSF Chimera (Pettersen et al., 2004; Trott & Olson, 2009), I tried to dock GGP, into chain A of the ADP-bound model. I chose this chain of this model because α1 is in similar position as in TPS, instead of being disordered as in chain B or shifted as in the sucrose bound models (Figure 13, Figure 19). Nevertheless, the active site is missing a few amino acids—the loop including amino acids 27–31 could not be modelled. The docking results vary drastically, however, since I knew the exact position and orientation of the glucose moiety due to the TPS substrate analogue and the bound sucrose, I excluded any results with a differently positioned glucose moiety (RMSD of the ring atoms bigger than 1 Å). Nevertheless, the glucose moiety is usually shifted away from the β-phosphate of the ADP. Still, the distance to the oxygen of the glycosidic bond, i.e. the bond broken during catalysis, is only 1.2 Å. A comparison with UDP from TPS, shows the β-phosphate to be similar but slightly differently positioned and thus closer to the position of the glucose moiety. This might be the correct position of ADP after catalysis or more likely an artefact owing to low resolution.

I repeated the docking and removed the ADP beforehand. The original dockings in presence of ADP yielded every time at least one valid docked GGP. Now 200 dockings (different seeds) ignoring the ADP yielded only four valid placed conformations. Eventually, in both scenarios the interactors are very similar but I focus on the docking ignoring ADP, since the glucose position is usually better (Figure 13). Similar to the expectations, the phosphate of the GGPs interact with Tyr9, Tyr93, Lys98, Asp147, Tyr148 and Asn149. Furthermore, the hydroxyl group of the glycerol in position 1 interacts with either Arg300, an ADP phosphate interactor or the amide amino subgroups of Gly33 and Ile34 of helix α1. However, coordination by Arg300 seems unlikely as it is critical for interaction and charge mitigation of the β-phosphate of the ADP.

Just as a reminder, the ADP was ignored for this docking and thus probably allows this interaction. From the position of the substrate in the active site, especially the glucose moiety, I assumed that ADP-glucose binds first before the G3P. All four glucose moieties of the docked GGPs are shifted similarly by a RMSD of 0.56–0.45 Å. They interact as the crystallised moiety with His171 and amino acids 399–402 but also with the amide bond of Leu403 and potentially with the Gly33 of α1.

2.3 Validation of chloride sites and chloride binding deficient mutant

2.3.1   GgpS binds reversibly chloride and bromide

A strong peak in the electron density map of the $wt_{Cl,SUC}$ dataset was interpreted as an ion coordinated by the amino subgroup of the amide bond between Pro109 and Trp110. The nature of the coordinator hinted towards an anionic ligand, of which chloride was the most abundant anion in the crystallisation condition. A subsequent refinement with chloride revealed no suspiciously low occupancy, high B factors or difference density around the peak.

The $wt_{Cl,SUC}$ dataset was measured at a wavelength of 1.00 Å and no significant anomalous signal of chloride is present. In contrast, a dataset from a new crystal (with ADP instead of SUC, $wt_{Cl,ADP}$) collected at 1.70 Å, exhibits anomalous density around the ion, further supporting the assumption that it is chloride (Figure 16). While the density drawn at 3 σ RMSD fits the placed chloride very good in chain A, the peak is lower and off in chain B due to higher B factors in that whole chain. In addition, since the K absorption edge of chlorine is at 4.39 Å, the weak anomalous signal is not unambiguos.

Therefore, we purified and crystallised the protein again but exchanged all chloride against bromide. As already mentioned, these models are remarkably similar. We analysed one crystal with bromide at a wavelength of 0.9197 Å ($wt_{Br,SUC,highAno}$) slightly above the bromine K absorption edge (0.9202 Å) as well as at 1.00 Å ($wt_{Br,SUC,lowAno}$). As expected, $wt_{Br,SUC,highAno}$ has a much stronger
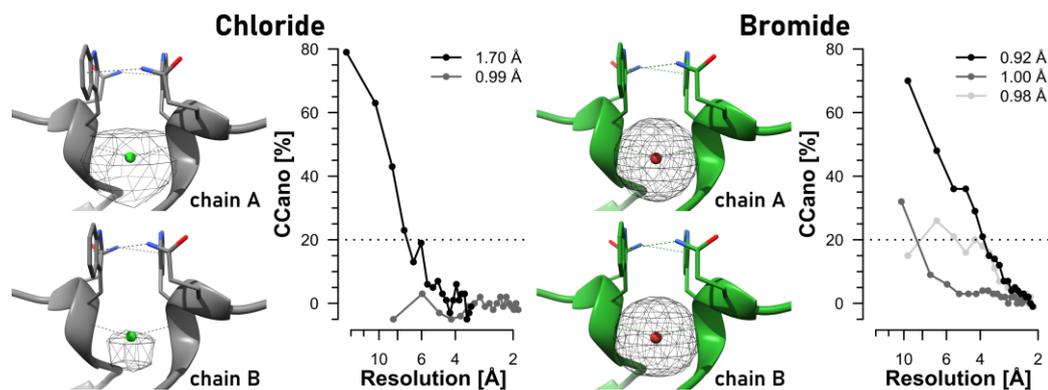
Figure 16: The chloride–binding site between the amide amino groups of Trp110 and its symmetry partner.

Two different crystals with chloride were measured at 0.99 Å ($wt_{Cl,SUC}$) and 1.70 Å ($wt_{Cl,ADP}$). As expect there is a higher anomalous signal at the latter wavelength since it is closer to the chlorine K edge at 4.39 Å. However, also phosphor (ADP) and sulphur (Cys, Met) scatter anomalously and are present in the crystal. Another crystal whose protein was purified and crystallised with any chloride exchanged against bromide was measured at 0.9197 Å ($wt_{Br,SUC,highAno}$) and 1.00 Å ($wt_{Br,SUC,lowAno}$) as well as a second crystal with bromide ($wt_{Br,SUC,Ano}$) at 0.98 Å. As expected the highest anomalous signal is visible in the first dataset, measured on the high energy side of the bromine K edge at 0.9202 Å. Depicted are the maps at 3 σ rmsd with a high resolution cut–off corresponding to a CC_{ano} of 20 % as estimated by XSCALE and indicated in the plots (Cl: 7.00 Å, Br: 3.75 Å). Only both molecules of the asymmetric unit of the high signal datasets are shown.

anomalous signal. However, since the low signal dataset was collected at the same crystal position after the first one, the anomalous signal might be reduced due to excessive radiation. Therefore, I further analysed anomalously the $wt_{Br,SUC}$ dataset which is a combination of two measured at a wavelength of 0.98 Å. The CC_{ano} of all resolution shells of the merged as well as unmerged high-resolution datasets are consistently low or even lower. The Figure 16 depicts only the dataset with the strongest anomalous signal ($wt_{Br,SUC,Ano}$).

Since *in vivo* chloride is usually more prevalent than bromide, we conclude that there is a chloride anion bound to the amide amino subgroup between Pro109 and Trp110. However, the coordination is not very strong and the chloride is released or substituted under suitable conditions without irreversible structural breakdown. In addition, if an anion is bound or released under physiological conditions, bromide should have a similar effect as chloride,

though K_ds might differ since bromide is larger with a thermo-chemical radius of 1.96 Å compared to 1.81 Å of chloride (Simoes et al., 2017).

2.3.2   Coordinating chloride by only two ligands is unusual and the motif is unique

GgpS coordinates the chloride by the amide amino subgroup of the amide bond between Pro109 and Trp110 and its C2 symmetry mate. Thus, the anion serves as bridge between two molecules of the biological assembly and the coordination number is two. In addition, the chloride is completely shielded from the solvent and the coordinating amide amino subgroups are part of a three amino acid long 3_{10}-helix. These properties appeared to us as very unusual. Oliviero Carugo analysed in 2014 all models deposited in the protein database in regards of how chloride ions are bound (Carugo et al., 2014). As I was also interested to find proteins with a similar binding site/motif, I repeated his analysis. However, the following values are mostly taken from his paper since he filtered for non-redundancy, which was not done in my analysis. Nevertheless, most data except from the coordination numbers are very similar.

In contrast to our initial thoughts, chloride ions being shielded from the solvent is uncommon, almost 20 % of all chloride anions found in all models in 2014 had a residual solvent accessible surface area of lower than 4 %. On the other hand, 80 % of the chlorides are coordinated by three, four or five ligands while a coordination number of two is indeed uncommon (10 %). Carugo notes that the average distortion from the ideal coordination angle of 180° is 75° and no chloride is bound having both ligands within 160–180°. My new analysis reveals the coordination angle to be better adhearing to theory. Most models in the subset of models with similar properties as GgpS (explained in Figure 17) have an angle between 145–155°, including GgpS with 149°. The only exceptions, 3v62 and 2qw5, have angles of 168° and 175°, respectively.

Protein	PDBid	Coord. - Sequence
GgpS_cl,suc		AFWPILHSF**P** W QFTYDSSDWD
	3tji	NVNAYWRNG**P** L MNNAISGVDM
D-Galactonate dehydratase	3tw9, 3twa, 3twb	VVSGYWRNG**P** I MNNALSGVDM
	4e4f	YKGAYWRRG**P** V TMSAISAVDM
	4f4r	YRGAYWRRG**P** V TMTAIAAVDM
D-Mannonate dehydratase	3thu	YKGAYWRRG**P** V TMTAIAAVDM
Lectin	3v62	RDGYKSLAKT I FGTYGIDCSF
Xylose isomerase	2qw5	RTFDPSSNY**P** E QRQEALEYLK

Figure 17: Proteins with similar chloride binding properties as GgpS.

The table lists proteins whose chlorides are coordinated by two nitrogens from the amide bonds within helix secondary structures. The nitrogens must belong to two different molecules of a biological unit and the solvent accessible surface area of the chlorides must be less than 5 %. Sequences include the ten positions preceding and following the chloride–binding amino acids. A) Sequence logo for proteins in the table, only unique sequences. B) Sequence logo for proteins with same restrictions except from solvent accessible surface area. Maximum is now 30 %. Sequence logos made with WebLogo (Crooks et al., 2004).

Several hits are sugar binding/modifying proteins, often dehydratases (Figure 17). They share a proline in front of the amino acid coordinating the chloride but, unlike GgpS, the chloride coordinating amino acid is never a tryptophan. Since peptide bond NH groups are usually bonded to carbonyl groups in the helix, the chlorides are all positioned at the N-terminal end of the helices, where also a positive partial charge from the helix dipole may attract the anion. However, upon closer inspection, there is no structural similarity to GgpS, as the helices are all from type α and not of type 3_{10} as in GgpS. In addition, while a couple of sugar-modifying enzymes are found, there are no glucosyltransferases at all.

2.3.3 W110P mutant – a model for chloride-free GgpS

Later sections will discuss the change of GgpS activity while titrating chloride. In order to link apparent changes to the binding of chloride and to understand unspecific ionic strength effects, we designed a binding deficient mutant. In addition, a 3D-structure of this mutation may characterise the conformational effect of

chloride release. Since the anion is bound via the amino subgroup of the protein backbone the only possible mutation was the substitution of the tryptophan to proline.

Such substitutions usually affect protein folding because the proline backbone ring structure restricts the ϕ angle within the protein backbone to the region around -60°. This increases backbone rigidity and can impede folding or enforce a different conformation. With residue 109 already being a proline, the mutation is further restricting due to formation of a diproline unit. Averaging across all four position in $wt_{Br,suc}$ and $wt_{Cl,suc}$ we get the following ϕ and ψ angles for Pro109 and Trp110 respectively: -60°, -22°, and -56°, -30°. Together with Gln111, these amino acids form the 3_{10}-helix η1 with all residues in *trans* conformation. Saha and Shamala have analysed the occurrences of diprolines in respect to their geometry (Saha & Shamala, 2012). Out of 809 analysed diprolines, 749 are in *trans* conformation and seven take part in a 3_{10}-helix. Six of them mark the beginning of that helix and are preceded by phenylalanine or tyrosine as would happen in a W110P mutant (Figure 18A). This setting is therefore unusual but possible. The ϕ and ψ angles with standard deviations for the first and second proline according to Saha and Shamala are (Figure 18A): -49° (9°), -41° (7°) and -57° (5°), -26° (6°). This indicates our ψ angle of the first proline to be lower than expected. Still, since the difference is low and the other angles fit, we anticipated no or just minor structural changes in the W110P mutant.

The W110P mutant behaves like the wild type protein during expression and purification (Figure 5). Moreover, it crystallises in the same space group with virtual identical cell dimensions (Table 12). Also, the monomers and the tetramer are very similar. The RMSD of the C_α-atom pairs between each GgpS wild type and W110P monomer are 0.650 Å over 466 amino acids and 0.721 Å over 442 amino acids for chains A and B, respectively. Nevertheless, there are differences, which are most prominent in chain B (Figure 11) as explained in the following paragraphs.

Model (Chain)	Sequence
1I6L (A)	^{125}YPP L^{128}
1JG1 (A)	^{153}F PPK^{156}
1NZJ (A)	^{48}D PPR^{51}
1RK6 (A)	^{192}Y PPA^{195}
1V9F (A)	^{123}Y PPI^{126}
1VPM (A)	^{19}F PPD^{22}
2INU (A)	^{343}Y PPF^{346}
W110P	^{108}F PPQ^{111}
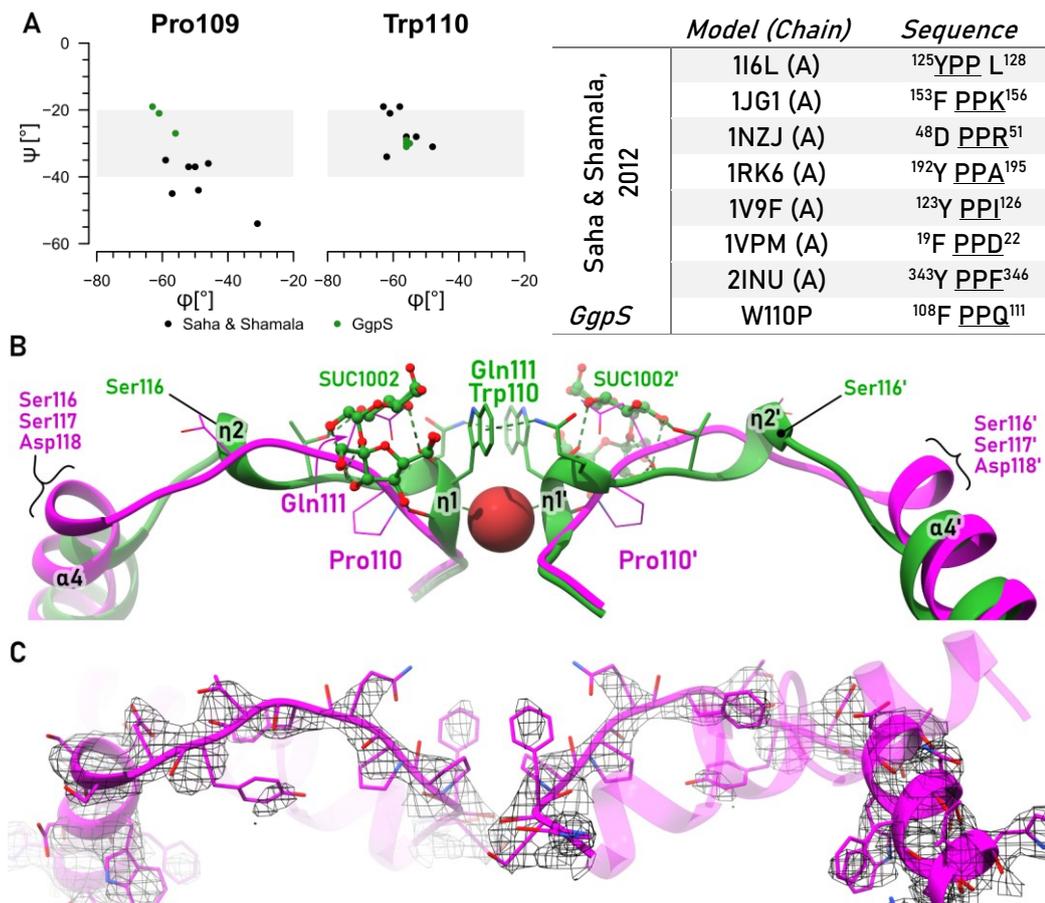
Figure 18: Ramachandran plots of Pro109 and Trp110 and the conformation following position 109 of W110P$_{Cl,SUC}$ superimposed onto wt$_{Br,SUC}$.

A) Ramachandran plots and proteins found by Saha and Shamala with diprolines within a 3$_{10}$–helix, while highlighting the participating amino acids. B) Alignment of W110P$_{Cl,SUC}$ chain B onto wt$_{Br,SUC}$ chain B with annotated features. C) 2mF$_o$-DF$_c$ electron density of the W110P$_{Cl,SUC}$ chain B at 1σ RMSD.

In contrast to our mutation-design hypothesis, substitution of the Trp110 to proline changes the conformation of the whole loop from position 109 to 118 in both chains (Figure 18B). Starting C-terminally, α4 is extended by one turn, now starting with Ser116 instead of Trp119. Thus, the space to Pro109 has to be bridged by three amino acids less and helices η1 and η2 wind up accordingly. In addition, Pro110 moves away from the chloride binding cleft, broadening it in total and thus disrupting both the chloride and the SUC1002 binding site.

I assume this loop with positions 109–118 to be under tension by the helix propensity of positions 116–118 and its counterpart, the chloride bridge with the π-interaction between Trp110 and Gln111. Therefore, I called it the 'tension belt' or short the tb-loop.

By losing the chloride, η1 is weakened considerable and unfolds followed by η2 due to the now up winding α4. This conformational variety is further supported by the models derived from the wild type crystals grown in sulphate. The whole loop downstream of potion 110 becomes often disordered while α4 is extended as in the W110P mutant (Figure 71 and Figure 72).

A tryptophan-proline substitution is usually thought to be very critical. However, a diproline unit in that structural context is possible and neither expression nor purification are affected. The mutation induced a different fold of the tension-belt loop. Also crystals grown in presence of but potentially without bound sulphate exhibit a similar folded sometimes partially disordered loop. Hence, the wild type is not restricted to the chloride-bound loop conformation. Up winding of α4 distorts the binding site which potentially increases flexibility in general but probably stronger in the wild type. Interestingly, B factors increase more strongly only in one subunit of $W110P_{Cl,suc}$ (Figure 25). Nevertheless, I conclude that the W110P is a good chloride binding-deficient mutant and is not suffering under artifical proline induced changes.

2.4 Chloride binding stabilises the active site

No structure of a GgpS tetramer bound to only one choride anion can be presented. A first attempt was the crystallisation in sulphate (and acetate) instead of chloride containing buffer, but the resulting crystals represent possibly the completely chloride-free state. There is sometimes density at the expected anion position but of the two protein chains constituting the binding site, one is always disordered. The supposed sulphate anion would therefore just be bound to one Trp110 without any support on the other side (Figure 71 and Figure 72). However, the missing amino acids there must be somewhere and could give rise to artefacts in locations where they converge across all measured asymmetric units. In addition, the resolution limit is only between 3.0 Å to 4.0 Å (anisotropy), hence any peak can be misleading considerably.
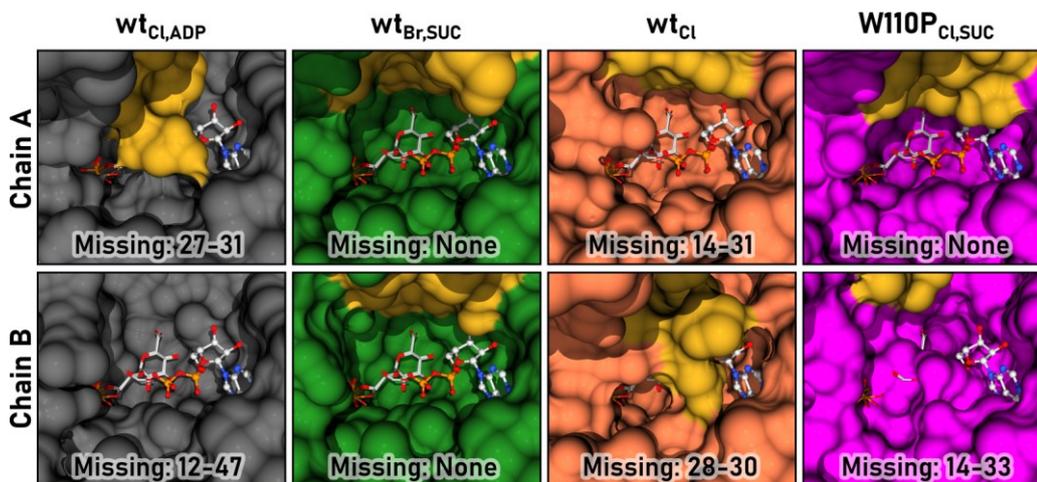
Figure 19: View in the active site pocket.

The perspective is as in Figure 13. Images are a composition of four docked GGPs, ADP from wt$_{Cl,ADP}$ chain A and the respective surface of the given model. The loop (27–33, yellow) between β1/2 (15–26) and α1 (34–44, also yellow) leading into the active site is sometimes disordered and not modelled. Missing amino acids are indicated per image. In chain A of wt$_{Cl,ADP}$, α1 protrudes from the top into the active site blocking the way in/out. However, the active site itself in chain A of wt$_{Cl,ADP}$ is not blocked, it was thought to represent the active state and hence used to dock GGP. α1 is disordered in chain B of wt$_{Cl,ADP}$. That blockage also appears in chain B of wt$_{Cl}$ indicating an inhibition by preventing substrates from entering the active site. However, Sucrose is bound in both active sites of wt$_{Br,SUC}$ (and wt$_{Cl,SUC}$) and hence entry is permitted. Chain A of wt$_{Cl}$ and chain B of W$_{110PCl,SUC}$ have both similar disordered amino acids and no Sucrose bound. However, none was soaked into the first, while in the second, the active site pocket is distorted while the chloride site in unoccupied.

2.4.1 The W110P tetramer is probably partially inactive

The most notable difference between both monomers in the asymmetric unit of the W110P mutant is the missing sucrose in chain B. This raises the question if there is no sucrose bound because of the lost chloride binding capacity or failed soaking attempts.

Soaking GgpS with ligands is usually successful after dropping the crystal into the cryo-protectant and immediate reattachment to the mounting-loop. In all analysed datasets better than 3 Å resolution, electron density peaks for sucrose molecules at positions 1001, 1002 and 1003 can be detected. The refined models have 100 % occupancies and B factors very close to the environment. The W110P$_{Cl,SUC}$ crystal was not treated differently, and peaks are found at the afore mentioned positions 1001 and 1003, but not at 1002. The latter is essentially bound to the now extended tension

belt loop. While the occupancy of SUC1003 is at 100 % in both cases (Figure 7), the sucrose in the other active site is only modelled with 90 % occupancy, indicating a problem as in chain B. But since SUC1003 is at 100 % occupancy and the soaking process is usually very fast, I conclude that sucrose was available but could not properly enter the active site.

A second crystal of the same mutant batch with similar diffraction properties was analysed (W11OP$_{Cl,SUC,2}$). Model refinement stagnates at an R$_{free}$ of 0.2770 compared to 0.2631 of the W11OP$_{Cl,SUC}$. The Wilson B factor is with 112 Å^2 12 % higher and the map is perceived as worse, hence W11OP$_{Cl,SUC}$ is the favoured model. Nevertheless, this is a second crystal soaked in sucrose and the distribution of sucrose molecules and their occupancies is the same as W11OP$_{Cl,SUC}$. In addition, another crystal like wt$_{Cl,SUC}$ but cryo-protected without sucrose yielded a 2.77 Å dataset of a ligand free structure (wt$_{Cl}$). Accordingly, it has no sucrose molecules bound in the active sites but the chloride positions are still occupied. Figure 19 compares the active site pockets and its distortions is only present in the W11OP mutant. Thus, active site alteration is correlated with the occupancy of the chloride site and not with the presence of sucrose in the active site. As a side note, the models of the potentially anion-free structures grown in sulphate have similar alterations in the active site (Figure 73, Figure 74, Figure 75 and Figure 76).

One of the monomers in the asymmetric unit of W11OP$_{Cl,SUC}$ cannot bind sucrose anymore and according to Figure 19 no substrate at all. The activity of the W11OP should hence be about 50 % lower than the wild type in the fully chloride occupied state.

2.4.2   Chloride site affects the active site via the tension belt loop and helix α3

In the following paragraphs, several water molecules become important. They have been automatically modelled at 1.80 Å resolution in the wild type structure by phenix.refine (Afonine et al., 2012). However, at a rather low resolution of 3.35 Å resolution,
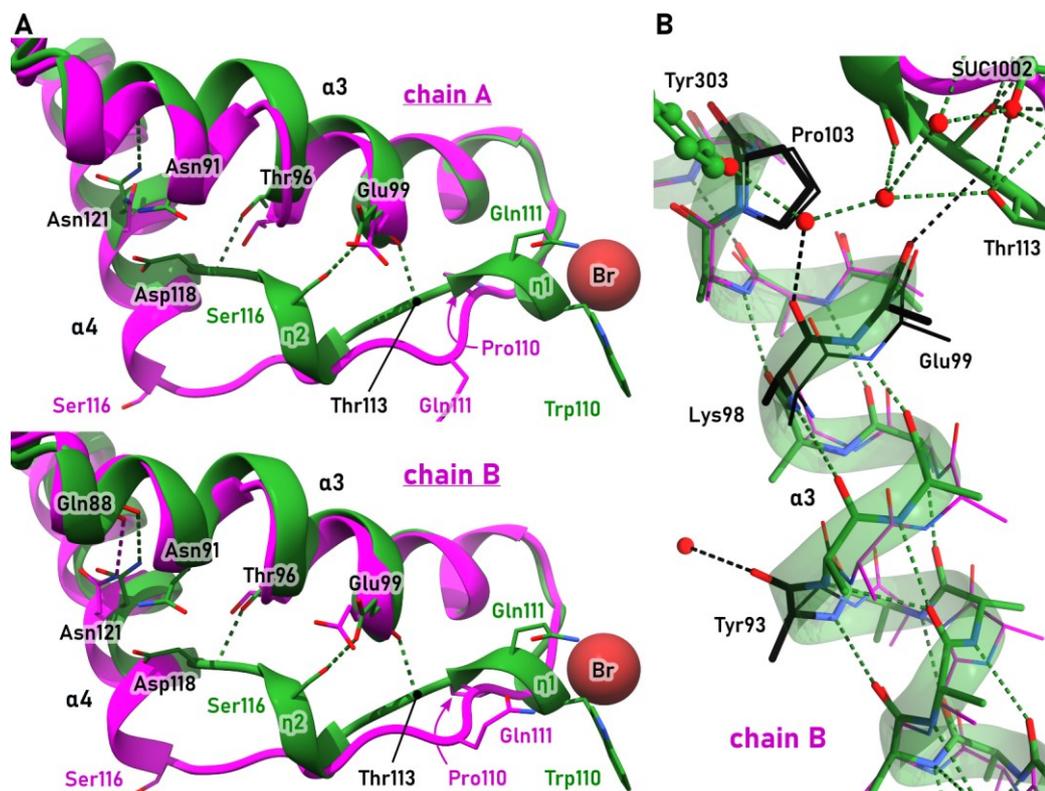
Figure 20: Chloride and bromide induce a conformational change of the 'tension belt' loop and further destabilise the already weakened helix α3.

$wt_{Br,SUC}$ chain B and $W110P_{Cl,SUC}$ are depicted in green and magenta, respectively. The ribbon is forced to go through the C_α–atoms. A) Losing the chloride/bromide induces the extension of α4 and linearization of η1 and η2. Due to this change, three amino–acid to amino–acid hydrogen bonds are lost. Water molecules are excluded for visibility. B) Pro103 leads to disturbances by inducing a π–bulge with an increased diameter. The carbonyl oxygens of Tyr93, Lys98 and Glu99 are not able to form intra helical hydrogen bonds and interact with two waters or Thr113. Polar interactions below 3.5 Å distance are indicated by dashed lines.

no water molecules were placed in the W110P mutant at all, hence their apparent absence is not due to the structural changes.

After releasing the chloride, the tension belt loop changes its conformation: α4 is extended by one turn forcing η1 and η2 to unfold. This change disrupts three hydrogen bonds between the tb-loop and Thr96 and Glu99 in helix α3 (Figure 20A). The figure does not show waters molecules for visibility reasons, however, there are some connecting α3 and the tb-loop additionally (Figure 21A). These changes happen in both copies of the asymmetric unit of $W110P_{Cl,SUC}$. Figure 22 depicts the $2mF_o$-DF_c electron density for molecule B of the asymmetric unit. Features in the map for the other molecule of $W110P_{Cl,SUC}$ are usually more pronounced.
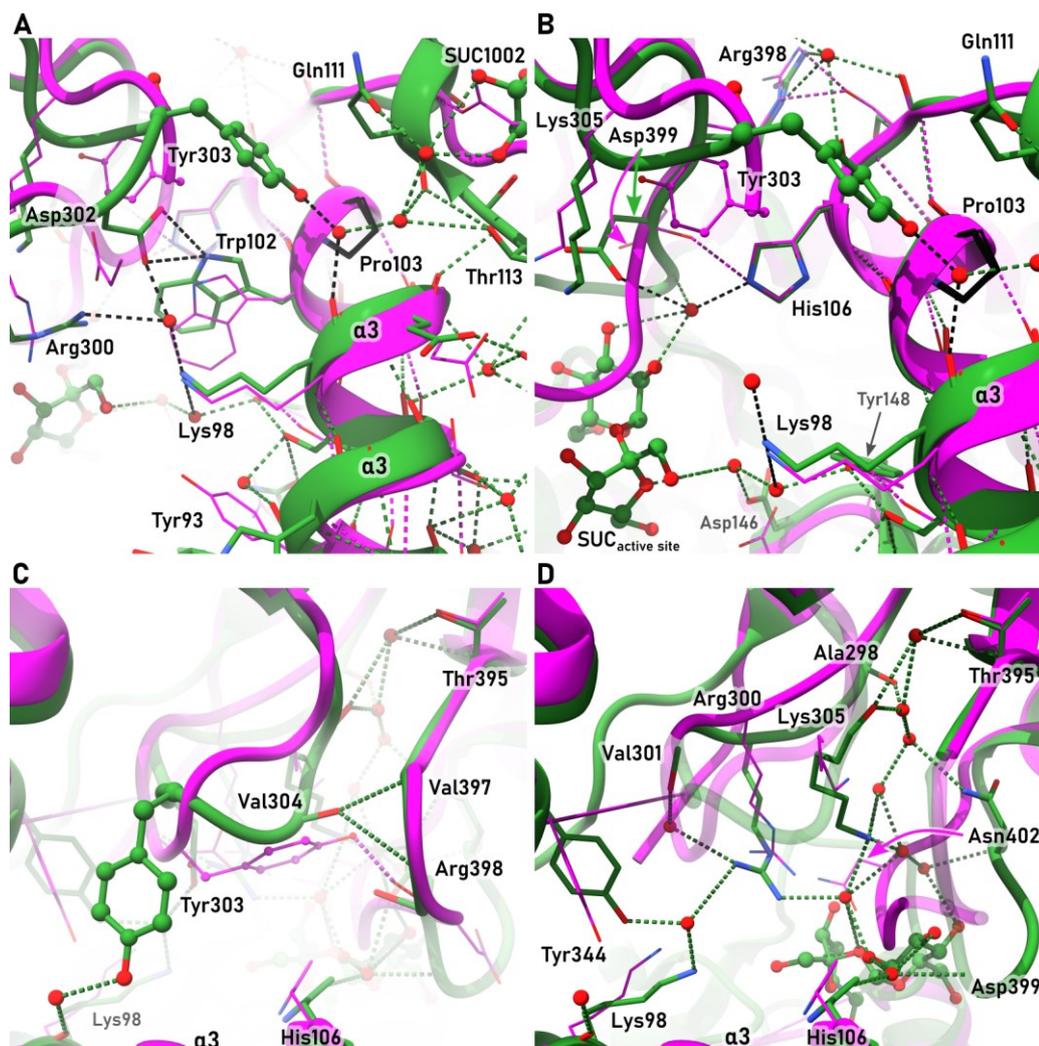
Figure 21: Loss of chloride/bromide affects the active site via α3.

wt$_{Br,SUC}$ and W110P$_{Cl,SUC}$ are depicted in green and magenta, respectively. A) Perspective on α3, the tb-loop on the very top right and the active site in the left. B) Slightly zoomed in view of A) with focus on the background while omitting some residues in the foreground (Trp102, 301–302). An oxygen of the phosphate in the docked GGP replaces the water molecule interacting with and being below Lys98. C) As A) but roughly rotated by 90°, α3 is at the bottom. D) Same as C) but with focus on the background while omitting the foreground. Polar interactions below 3.5 Å distance are indicated by dashed lines. Similar perspectives for the sulphate grown crystals are found in the appendix (Figure 73, Figure 74, Figure 75 and Figure 76).

Helix α3 is not a conventional α-helix as its intra-helical hydrogen bond network is severely disturbed. Amino acid 103 is a proline one turn before the actual C-terminus. This proline has of course no free hydrogen bond donor and requires more space vertically. This allows some carbonyl oxygens to form hydrogen bonds with atoms outside the helix (Figure 20). Glu99 in the Pro103 directly preceding turn interacts with Thr113 in the tb-loop (Figure 20). Likewise, neighbouring Lys98, Tyr93 of the next prior

'Tension belt'-loop & α3 α3

wt_{Br,SUC} **chain B**
W110P_{Cl,SUC} **chain B**
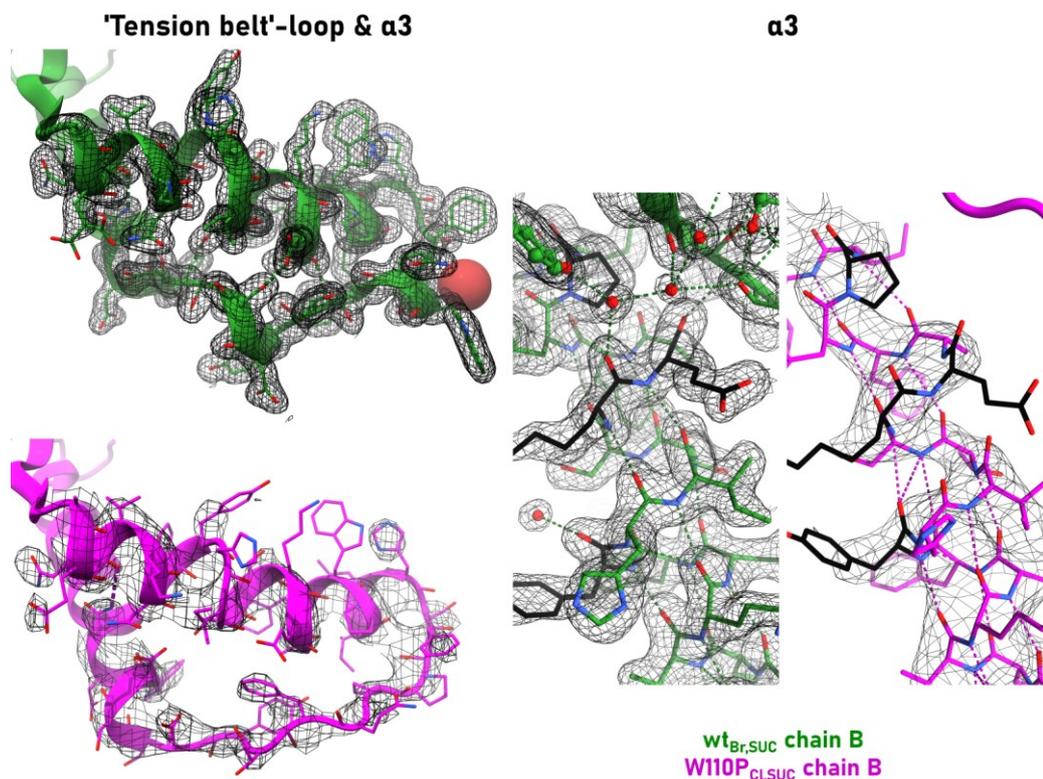
Figure 22: Electron density of the 'tension belt' loop and α3.

2mF_o–DF_c map is depicted at 1σ RMSD contour level.

turn interact with water molecules each. The latter introduces a so called π-bulge, a turn with larger radius including four amino acids and intra-helical hydrogen bonds in $O_i \rightarrow NH_{i+5}$ manner instead of α-helix conforming $O_i \rightarrow NH_{i+4}$. Tyr93 and Ly98 of this helix are of special interest since they are most likely interacting with the G3P-moiety in the active site (section 2.2.4).

A closer analysis reveals Lys98, Trp102, and His106 as the beginning of a hydrogen bond network, mostly via water molecules, towards other parts of the active site. The water molecule close to the backbone carbonyl of Lys98 is interacting with Tyr303 but also with Thr113 of the tb-loop as well as via another water molecule with SUC1002 and Gln111 (Figure 21A). Tyr303 is part of the loop interacting with the phosphates of ADP (Arg300, Lys302) and its base (Gly299) as explained in section 2.2.4.

The automatically placed waters were evaluated by Coot for potential metal ions (Emsley et al., 2010). In the vicinity of the bespoken positions only one potential potassium was found in
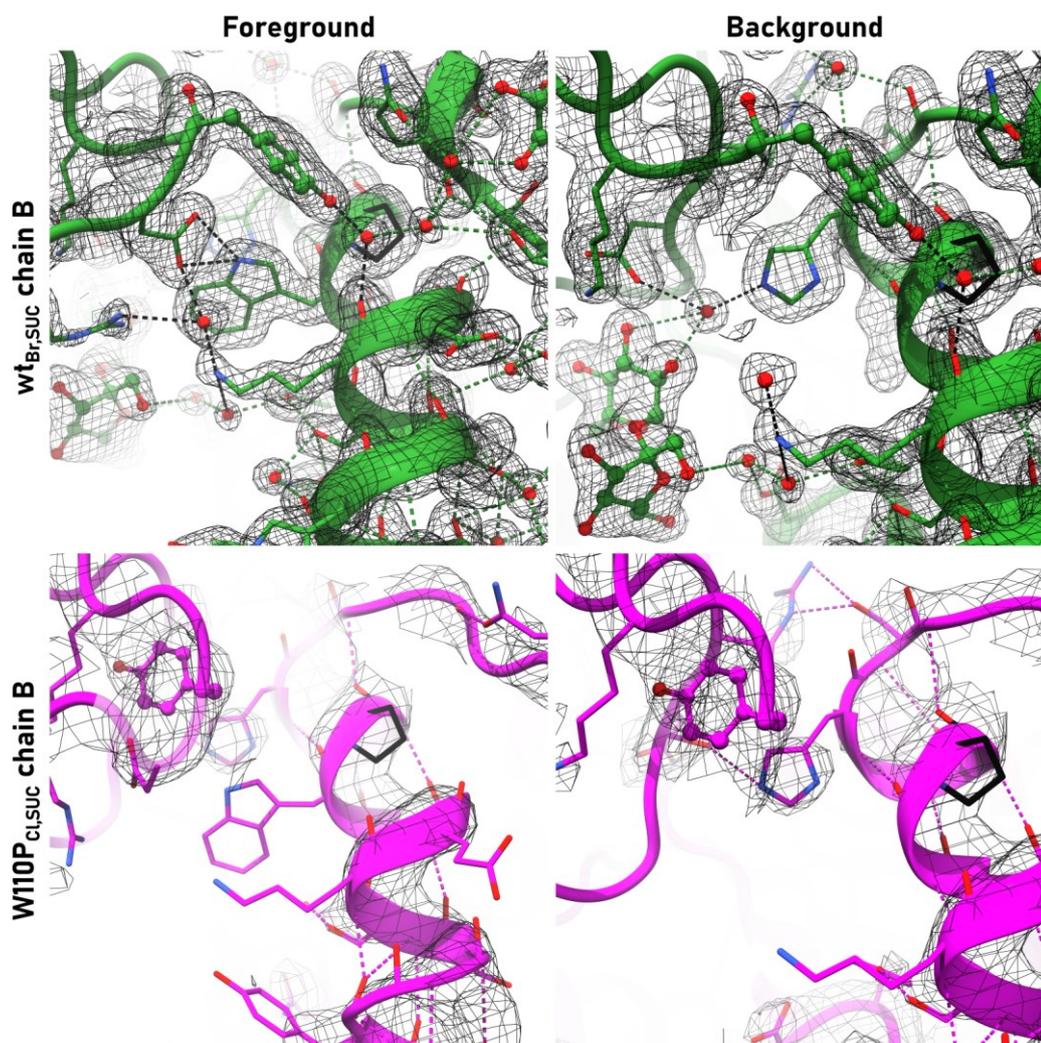
Figure 23: Interaction between α3 and the loop around position 300 with electron density.

2mF$_o$–DF$_c$ map is depicted at 1σ RMSD contour level.

wt$_{Br,SUC}$ chain B but not in chain A or in any subunit of the wt$_{Cl,SUC}$ model. The electron density of that particular water molecule bound by Ala298 is very low and thus not placed in the other cases (as in Figure 21D). The anomalous map of wt$_{Cl,ADP}$ revealed no unexpected signal at that position, although potassium should give a similar if not stronger signal than chloride, sulphate or phosphate at the given wavelength. Hence there are most likely no further ions than the chloride involved.

All changes described in the following paragraphs are only observed in chain B of W110P$_{Cl,SUC}$. The other copy in the asymmetric unit adopts conformations similar to the chloride bound wild type. The mentioned Tyr303 is here rotated by almost 180° including its backbone. This goes in hand with a conformational
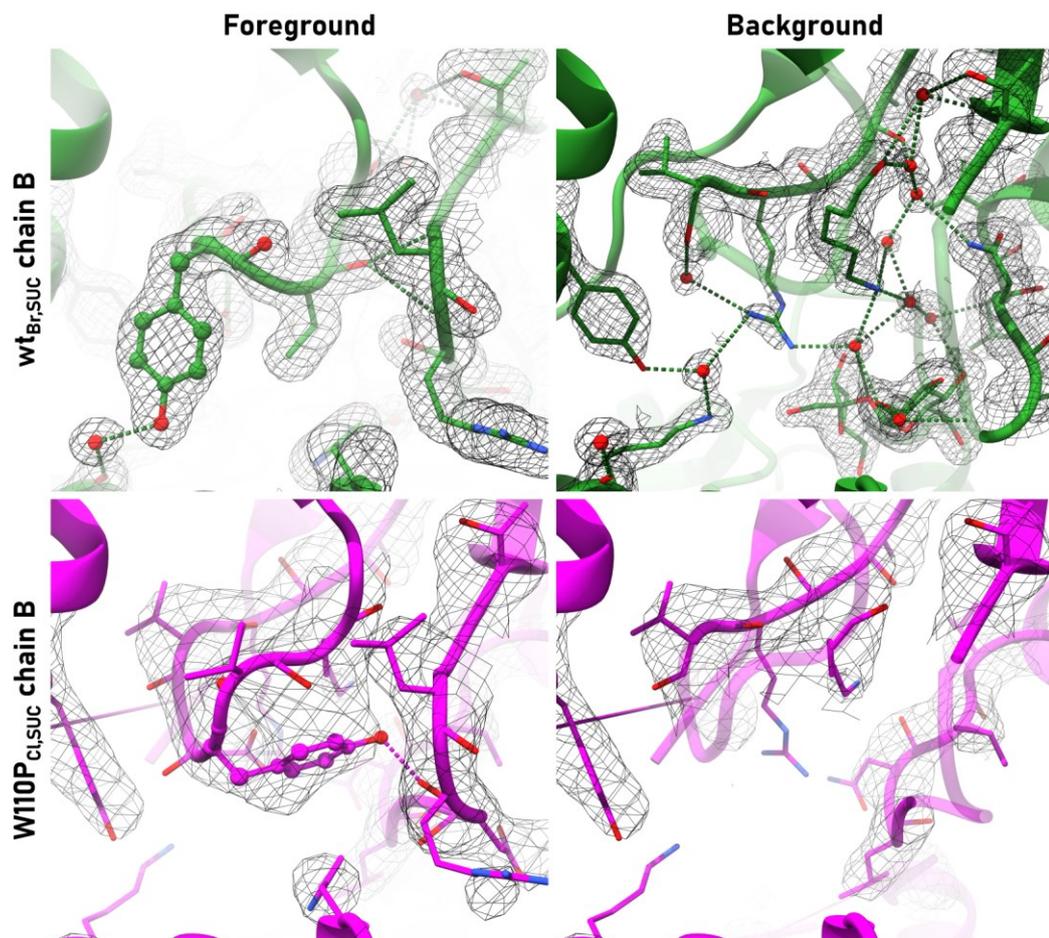
Figure 24: Interaction between the loops around positions 300 and 400 with electron Density.

2mF_o–DF_c map is depicted at 1σ RMSD contour level.

change of the entire loop from 298 to 305 (Figure 21C). Although, this rotation changes the interaction between this loop and positions 397 and 398, it probably does not induce that movement directly. Nevertheless, the important active site loop 399–404 adopts a different conformation and effectively takes the position of SUC1001, rendering this active site inactive (Figure 19). However, explaining this is difficult. The opposite site of this loop (beyond the right border of Figure 21D) does not change, so it cannot cause the movement. His106 stabilises the water network close to the SUC1001 and it is shifted slightly in the W110P mutant (Figure 21B, middle and D, bottom). Another water network exists and is stabilised by Ala298 and Lys305, which, as mentioned before, shift both. In the wild type, the network interacts with Asn402 and in the W110P mutant, Asn402 is one of the most moving amino acids in that loop at all (Figure 21D, right).
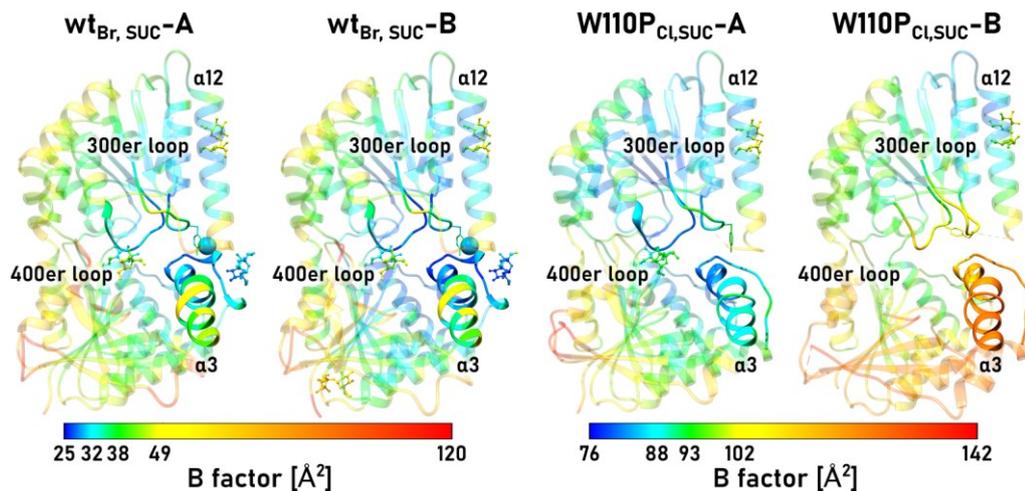
Figure 25: B factors of W110P$_{Cl,SUC}$ and wt$_{Br,SUC}$.

The colours represent the B factor and are scaled to their distribution, i.e. quartiles, in the asymmetric unit. Colours range from blue (minimum) via cyan (25 % quartile), green (median), yellow (75 % quartile) to red (maximum). B factors between both chains of wt$_{Br,SUC}$ are very similar. α3 is special since its middle part, the π–bulge has higher B factors than its termini. While chain A of W110P$_{Cl,SUC}$ is similar to the wild type, chain B has much higher B factors, predominantly at N–terminus (bottom) but also the active site.

A lot of the bespoken interactions between different parts of the protein are via water molecules. Moreover, in the W110P mutant all direct interactions are lost and a larger gap is formed between the tb-loop and α3 (Figure 20A). That helix stability depends on the tb-loop conformation because of the Pro103 and the π-bulge disturb the common intra-helical hydrogen bonds (Figure 20B). This helix is recognisable in chain B of the W110P$_{Cl,SUC}$. But sidechain conformations are uncertain due to low or very low electron density, thus yielding very high B factors (Figure 25 and Figure 23). Therefore, the hypothesis how the chloride binding/release influences the active site is as follows: Due to a disturbed intra-helical hydrogen bond network, helix α3 is severely destabilised. This flexibility impedes directly binding of G3P via Tyr93 and Lys98 (Figure 21A, B). It also propagates to the active site loops around positons 300 and 400 which are important for ADP-glucose binding (Figure 21). This can block the active site (Figure 19). In the anion-free state, the tb-loop is in a retracted conformation not interacting with helix α3 (Figure 20). However, upon anion binding, the tb-loop rigidifies and undergoes direct interactions

with helix α3. This structural support reduces flexibility in the helix and consequently in the active site loops as well (Figure 25).

The flexibility is also demonstrated by the second W110P model, briefly mentioned at the beginning of this section. There, loops 300–306 and 399–403 of chain B appear to be in wild type conformation. However, there is very low electron density at the expected position of the sucrose molecule and the loop between 300 and 306 is surrounded by positive difference density. This indicates that both conformations are present simultaneously in this crystal. Unfortunately, a refinement with alternate conformations failed because the alternate states overlap considerably, the resolution is low with 3.30 Å and the dataset is apparently worse with a higher R_{free} (0.2770 versus 0.2631) than for $W110P_{Cl,SUC}$ at a similar resolution. Nevertheless, there is no density for the sucrose molecule, indicating this subunit being less active due to high flexibility, again.

A missing sucrose molecule in one subunit of the asymmetric unit and correlating conformational changes lead to the hypothesis that activity of a chloride-free GgpS is reduced by up to 50 %. However, the sucrose is also missing in chain B of the second W110P model without showing these pronounced changes. This indicates that the actual structure of the active site is not causative. However, likewise, an incomplete soaking was excluded. The only similarity between both W110P structures is the high flexibility of the discussed parts in their chains B. Hence, sucrose cannot bind and the activity is probably reduced but certainly not as strong as initially thought. On the contrary, an induced fit mechanism by the real substrates, especially ADP-glucose, might stabilises the active site again. Both monomers around one chloride site are crystallographically symmetric. Hence, another question is why only one dimer around one chloride site is affected although both chloride sites have the W110P mutation. It can indicate an interaction between both sites and flexibility of the active site is only allowed in one chloride site-surrounding dimer simultaneously. In the crystal, chain B is favoured because of the lattice but active and inactive dimers may be in an equilibrium in solution.

2.5 Chloride binding modulates GgpS activity

GgpS can bind chloride, and the two-chloride occupied and chloride-free states have differences potentially affecting activity. So far, chloride was only considered as DNA releasing factor due to electrostatic shielding but not as a direct interaction partner. Hence, I went back to the literature in order to understand how our finding might fit published data before conducting my own activity assays.

2.5.1 Previous publications may have missed out a chloride effect in the low millimolar range

The only ion previously found to be directly important for GgpS is Mg^{2+} (Schoor et al., 1999). They tested a 'crude' protein extract later declared to be inhibited with DNA (M. Hagemann et al., 2001). In addition, they also tested the activation of GgpS by different salts, of course in presence of magnesium chloride. Interestingly, independent of the counter cation, chloride is always the best activator, however their conclusion focuses entirely on cations ignoring the anions.

Since Schoor's report, only cations, more specifically mostly Mg^{2+}, were considered to be relevant for GgpS activity and it was usually provided by purifying in and/or adding its chloride salt to activity assays. The only other experiment about Mg^{2+} and purified GgpS is depicted in Figure 1 in the Hagemann et al. publication. It claims to depict the activation of recombinant GgpS from as low as 0 mM magnesium chloride without the presence of any sodium chloride. However, the exact assay conditions are not described and as stated in their method section, the protein was dialysed usually against 5 mM magnesium chloride. Moreover, the caption of Figure 2 states the standard assay contains 10 mM Mg^{2+} but does not detail where that magnesium comes from, i.e. the counter anion is not declared. Hence, the observed effects might come from the magnesium counter ions, chlorides, and other chloride anions not accounted for. Most published experiments are usually carried out with 4–10 mM but sometimes also up to 40 mM magnesium

chloride. Thus, the actual chloride content varies at least between 8–20 mM up to 80 mM.

So far, all publications claim that ions, except Mg^{2+}, are only important for nucleic acid removal (Jens F Novak et al., 2011). However, Benjamin Roenneke reports in his PhD thesis a sodium chloride dependent increase of DNA-free GgpS activity up to around 100 mM. He performed his assay in presence of 20 mM magnesium chloride (Roenneke, 2014, p. 61).

Taken together, this means that the chloride-free GgpS was potentially never sampled correctly and any explanation of the observed effect could have missed out or mixed up with an Mg^{2+} effect. Simultaneously, there is no structural evidence for a Mg^{2+} requirement in the active site of GT20 members to which GgpS belongs to (S. S. Lee et al., 2011). Moreover, there are 'no known' metal dependent glycosyltransferases with the B fold (Moremen & Haltiwanger, 2019). These prior experiments give a rough upper bound in the low millimolar range for the effect of chloride binding and highlights the importance of monitoring and declaring any ion in the sample. With regards to Mg^{2+} being essential for GgpS: I am not entirely convinced that GgpS requires Mg^{2+}. However, while there are very good structural indications, there is no disproof yet and since I am focusing on the chloride effect, I decided to keep 4 mM Mg^{2+} in all activity assays but used the acetate salt instead.

2.5.2   Setting up the activity assay

Investigating the effect of ions on the activity of an enzyme is a challenging venture, since the mere presence of the ion and others may have an effect without a dedicated binding site. Franz Hofmeister described stabilizing and destabilizing effects of different salts with same ionic strength on proteins (Hofmeister, 1888; Lo Nostro & Ninham, 2012). In addition, I explained why a chloride effect could have been missed out (section 2.5.1) and concluded in order to avoid or at least understand any interference, the presence of all ions has to be carefully monitored during reaction, including substrates and pH, but especially for the presence of anions (Figure 26).

Ion	Concentration [mM]	Ionic Strength [mM]
Na^+	18.1	9.1
K^+	16.7	8.4
Mg^{2+}	4.0	8.0
Li^+	4.0	2.0
Cl^-	0.5	0.3
$HEPES^-$	16.7	8.4
Ac^-	8.0	4.0
PEP^{3-}	2.0	9.0
$NADH^{2-}$	1.8	3.6
$ADPG^{2-}$	4.0	8.0
$G3P^{2-}$	4.0	8.0
Total	79.8	68.6

- Benchmark: 1x PK/LDH; No GgpS, ADPG, G3P but 4 mM ADP
- Full assay: 10x PK/LDH; 30 nM GgpS and substrates as in table
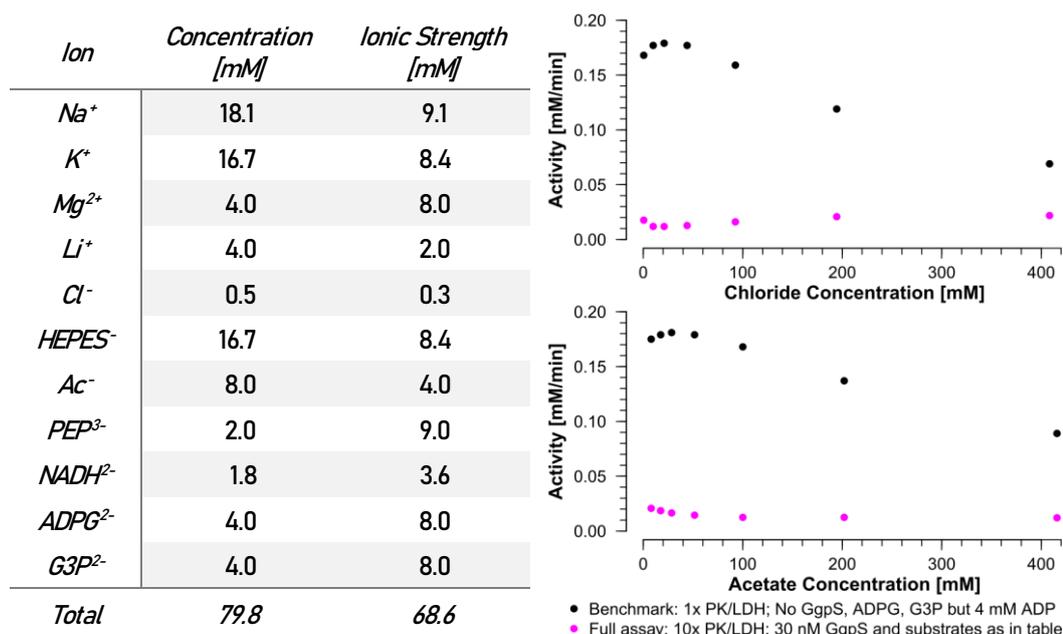
Figure 26: Assay composition and benchmarks of pyruvate kinase (PK) and lactate dehydrogenase (LDH) with sodium chloride and sodium acetate.

Sodium, potassium and lithium are counterions of the substrates. HEPES/KOH dissociation was calculated by the pH at start of the reaction and the zwitterionic form was assumed to have no ionic strength. Sodium chloride and acetate have similar inhibiting effects on the both assay enzyme PK and LDH (black dots, no GgpS, ADP-glucose (ADPG) and glycerol–3–phosphate (G3P) but 4 mM ADP). At all times, the 1x coupled enzymes are faster than GgpS (magenta dots). To further ensure that GgpS is limiting, all measurements (including magenta reference) were performed in presence of ten times more coupled enzymes than in this benchmark demonstration without GgpS. PEP: Phosphoenolpyruvate.

Chloride was only used for the purification of the protein and not part of any other assay ingredient. The protein was always diluted and chloride content adjusted so that the assay eventually contained 0.5 mM chloride. Except from 8 mM acetate (added as magnesium acetate), there was no other anion than the substrates required for GgpS and the pyruvate kinase (PK) and lactate dehydrogenase (LDH) themselves. Both enzymes, react differently on cations. Especially potassium is an activator of pyruvate kinase while sodium competes and activates less efficiently (Kachmar & Boyer, 1953). Raising the sodium ion concentration thus acts like a net inhibition. The activity of both enzymes was analysed under typical sample conditions excluding ADP-glucose and G3P, the substrates of GgpS. Tested were sodium chloride and acetate. To ensure GgpS to be limiting, eventually ten times more pyruvate

kinase and lactate dehydrogenase were used than the amounts utilised in these benchmark experiments (Figure 26).

In accordance to previous publications and mentioned before, I decided to add 4 mM Mg^{2+} by supplementing the acetate salt. Regardless of any GgpS requirements the pyruvate kinase is Mg^{2+} dependent (Baek & Nowak, 1982).

 The activity assays were run in true biological triplicates, i.e. with freshly extracted proteins from independent expressions. Besides sodium chloride, also bromide, acetate and sulphate were titrated. Concentrations have been chosen based on the equivalent ionic strength, thus a smaller concentration range was evaluated in case of sulphate. The sodium chloride titration was repeated with a narrower sampling in the low millimolar range.

A sample without addition of any extra salt serves as common standard throughout all five data series. Figure 27 depicts all data from all replicates and Figure 28 the averages. Each replicate series can vary in activity by 10–30 %, while the relation between the samples within this series remains similar. Figure 29 focuses on the relation between samples by normalising each sample of a series to the sample without extra added salt before averaging the replicates. This highlights the differences induced by the employed salt concentrations.

2.5.3   Chloride influences the activity of GgpS in a biphasic manner in the low millimolar range

Under the assay conditions (0.5 mM chloride and 8 mM acetate, 68.62 mM ionic strength) without any extra salts, the GgpS wild type has a k_{cat} of 10.52 s^{-1} with a standard deviation (SD) of 1.26 s^{-1} across all 15 measurements (5 dataseries, 3 triplicates). The addition of sodium chloride decreases activity to 70 % (SD = 4 percent points) at 11.2 mM. In general, activity is below 72 % between 3.85 mM and 24.3 mM chloride. It eventually rises again to 97 % (SD = 2 points) at 84.2 mM and stagnates around 118 % (SD = 5 points) at 194 mM and 124 % (SD = 2 points) at 408 mM total chloride.
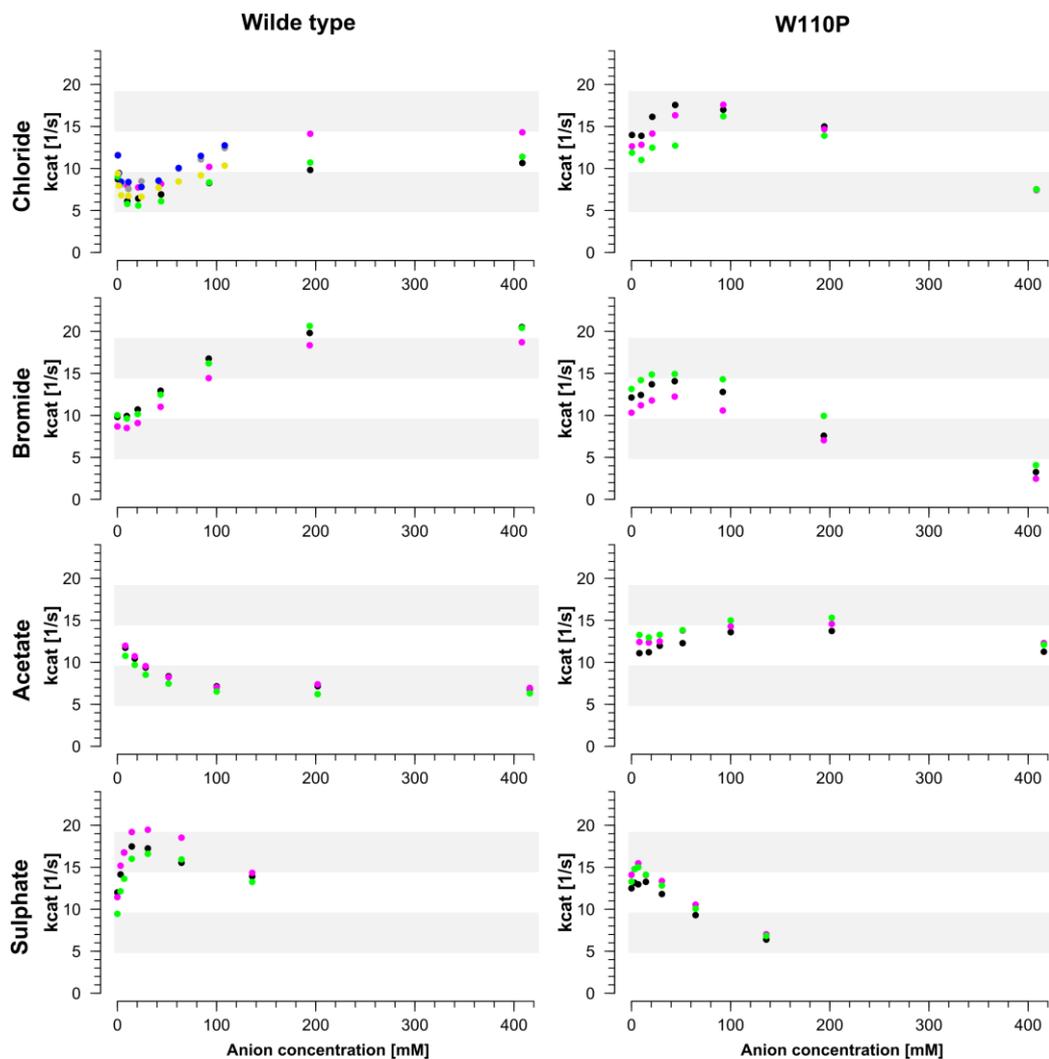
Figure 27: GgpS wild type and W110P mutant activity in presence of different anions, measured in triplicates.

The wild type was titrated twice with chloride. The second time, samples were chosen specifically around the minimum at 20 mM. Each replicate data series differs in its total activity. Better comparability is achieved by normalising each series to the 0 mM extra added salt sample prior to averaging (Figure 29).

In section 2.3.1, the model of GgpS bound to bromide instead of chloride was analysed. As described, bromide binds identical as chloride and no differences between the structures are detecable. However, bromide affects GgpS differently than chloride. It does not inhibit as strong as chloride at low concentrations, in fact up to 20 mM bromide has apparently no effect at all on the activity. Eventually, like chloride, the activity rises and reaches a maximum at 194 mM. This maximum, though, lies with 206 % (SD = 5 points) much higher than the 118 % with chloride.

As explained, due to the addition of magnesium acetate, there is at all times 8 mM acetate present in every sample and as acetate
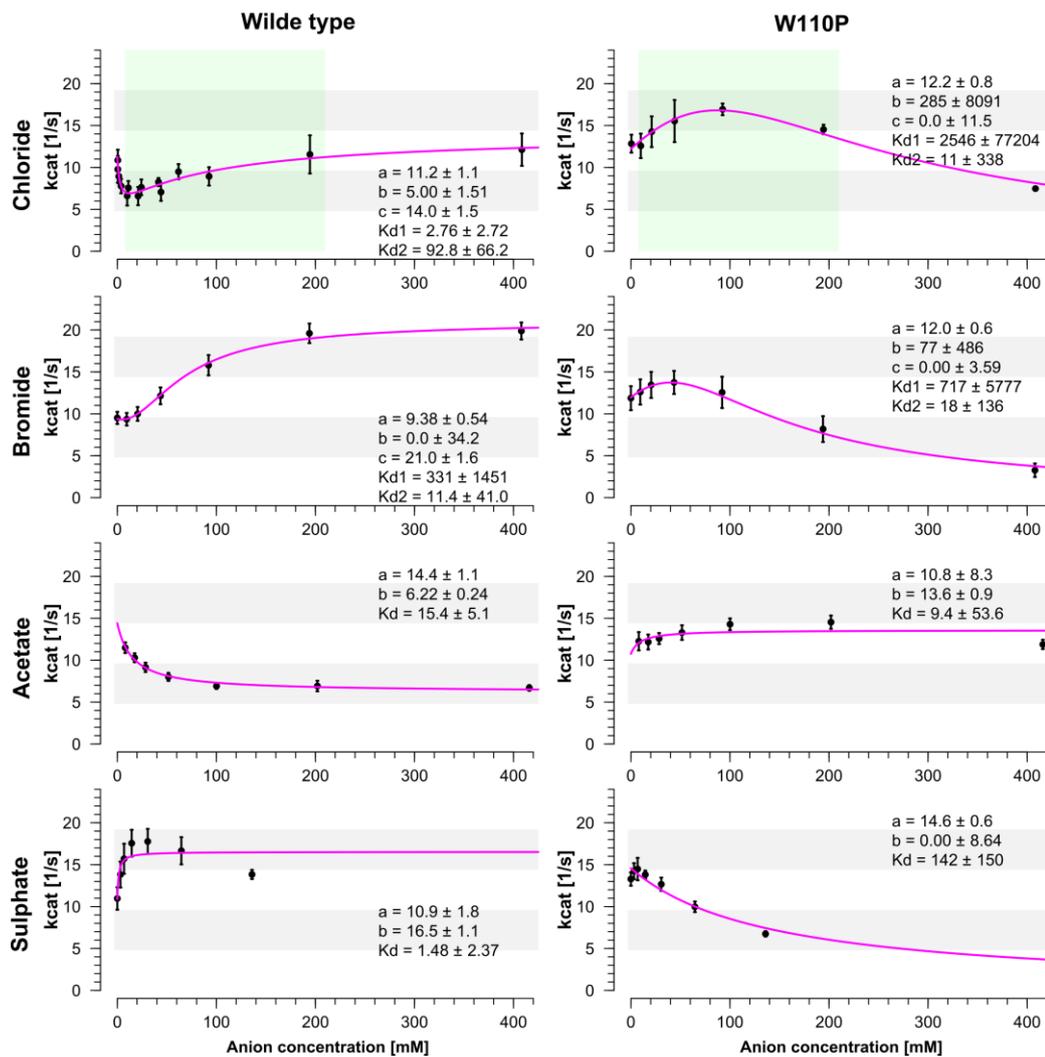
Figure 28: GgpS wild type and W110P mutant activity in presence of different anions, measured in triplicates.

While all salts show a pronounced and dissimilar effect on the activity of the wild type, they have a similar effect on the W110P mutant: First activity increases and then decreases. Error bars: standard deviation; Green highlighted area: Minimum physiological range of intracellular chloride concentration in *Synechocystis* PCC6714 (Reed et al., 1985). Fit explained in section 2.5.5.

was thought to have no effect on GgpS, it was chosen as negative control. However, acetate indeed affects the enzyme, reducing its activity down to 60 % (SD = 1 point) at and beyond 100 mM.

Sulphate was chosen as an alternative to acetate but again this divalent anion has a strong effect on GgpS. Like chloride and bromide, it exhibits two phases. Whereas both monovalent anions first inhibit and then activate, sulphate switches those phases. More sulphate activates proportionally up to a maximum of about 160 % between 15 mM and 31 mM sulphate. Then, activity drops proportionally to 127 % at 136 mM (last data point, ionic strength

equivalent to 408 mM chloride sample). However, as depicted in Figure 29, the biphasic nature might be an artefact, since the last four triplicates have a high variance.

The k_{cat} of the W110P mutant without added extra salt is with 12.56 s⁻¹ (SD = 1.11 s⁻¹) 19 % higher than the wild type and thus similar to the value of the wild type at 194 mM. In the following paragraphs I describe again the activity based on the scaled then averaged replicates. As such the no-salt sample is again the origin with an activity of 100 %. Activity increases to 132 % or to 16.9 s⁻¹ (SD = 10 points or 0.69 s⁻¹) at 93 mM and decreases afterwards to 58 % or 7.47 s⁻¹ (SD = 5 points or 0.04 s⁻¹) at 408 mM chloride. Bromide acts similar but earlier and with lower activity; Increase to 116 % (SD = 2 points) at 44 mM and decreases to 27 % (SD = 4 points) at 408 mM.

The effect of acetate and sulphate changes as well. Similar to chloride but a bit later, the maximum activity is around 150 mM with a maximal value of 119 % (SD = 4 points) and a minimal one of 97 % (SD = 5 points) at 416 mM. Sulphate is also following this trend, having its maximum at 109 % (SD = 5 points) at 6.83 mM (21 mM ionic strength) and its minimal activity at 51 % (SD = 1 point) at 136 mM sulphate (408 mM ionic strength).

While all anions have a different effect on the k_{cat} of the wild type, they act very similarly on the mutant: An increase of activity in the mid-range of ionic strength (90–160 mM) followed by a more or less strong decrease. This becomes clear when plotting the samples against ionic strength instead of concentration (Figure 30). This indicates that the W110P mutant and probably the wild type as well react on ionic strength alone. However, the anion species is still of some importance. Bromide for example inhibits much stronger (down to 33 %) than chloride and sulphate (70 %, 61 %) at 476 mM ionic strength. The comparison of both proteins also indicates that all tested anions are able to occupy at least to some extent the 'chloride' binding site and can modulate the activity this way.
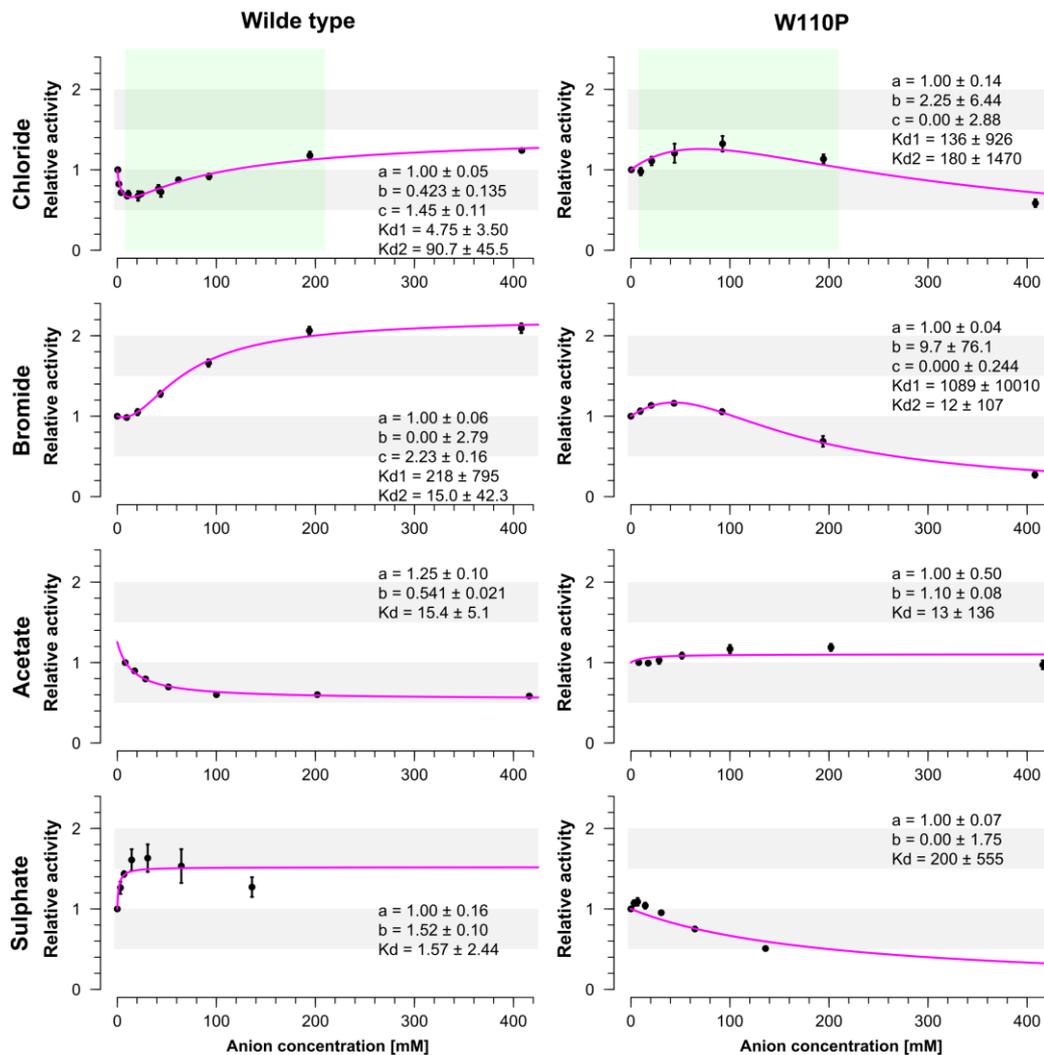
Figure 29: GgpS wild type and W110P mutant scaled activity in presence of different anions, measured in triplicates.

Similar to Figure 28 but better comparability is achieved by normalising each series to the first sample (no extra salt added) prior to averaging. Thereby, systematic differences per protein batch are bypassed. Wild type and W110P mutant are not directly comparable, as they have different activities without adding extra salt (Figure 28). The error bars (standard deviation) highlights where the relation within one replication series differs from the others. Green highlighted area: Minimum physiological range of intracellular chloride concentration in *Synechocystis* PCC6714 (Reed et al., 1985). Fit explained in section 2.5.5.

There is also another important observation. Acetate inhibits and all samples were measured in presence of 8 mM acetate. This indicates that in the beginning all wild type samples are actually to some degree inhibited and the titrated anion is competing against acetate. While this certainly changes the activity at a low anion concentrations, it does not change the biphasic behaviour of chloride and bromide since increasing their content up to 10–20 mM inhibits even stronger.
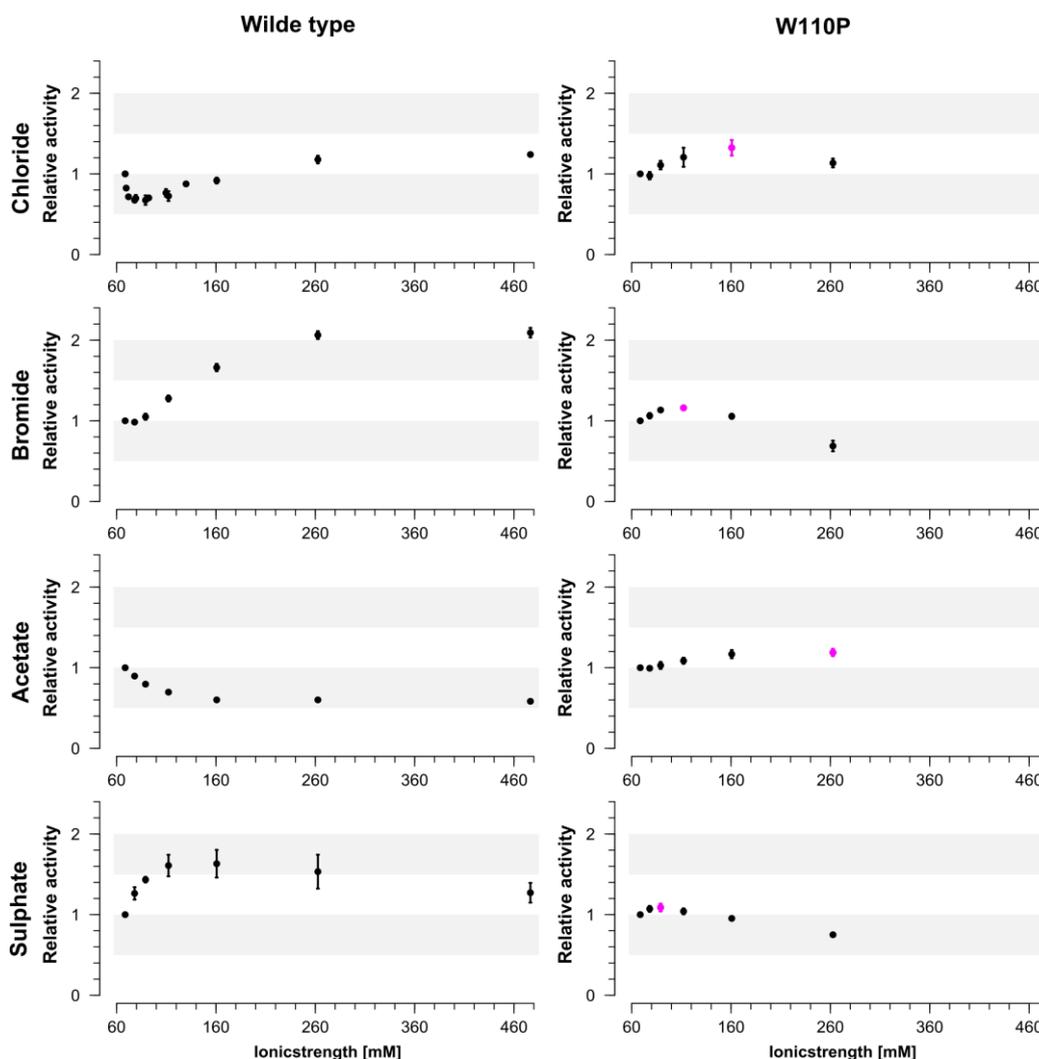
Figure 30: GgpS wild type and W110P mutant scaled activity in presence of different anions, measured in triplicates plotted against ionic strength.

Activity per sample as in Figure 29. The sample salt concentration was chosen based on same ionic strength. Magenta points highlight maximum activity in the W110P mutant. Highest activity occurs around 110 mM. Error bars: standard deviation.

A linear extrapolation of the acetate series reveals the k_{cat} to be at least 8.8 % (11.44 s⁻¹) higher than anticipated. Thus, the 19 % higher k_{cat} of 12.56 s⁻¹ (SD = 1.11 s⁻¹) of the W110P mutant is very close to that of the wild type in the absence of anions. However, this means that the activity of chloride-free and fully chloride occupied GgpS with 12.19 s⁻¹ (1.93 s⁻¹) are very similar, although the prediction based on the structures was, that the latter will be less active. As described, high ionic strength reduces the activity of W110P mutant and potentially affects the wild type similarly, explaining this discrepancy.

All four tested chloride salts have a considerable and different effect on GgpS wild type activity. Especially chloride, but probably bromide as well, provoke a biphasic response of activity: Up to 10–20 mM, activity decreases by 30 % (chloride) or stagnates (bromide) but increases with further added salt, until it stagnates after 200 mM at around 120 % (chloride) or twice the stating activity (bromide). Thus, chloride mostly affects GgpS below 100 mM which is a similar range not test correctly in earlier publications because of the addition of magnesium chloride. In addition, GgpS is slightly ionic strength dependent, which is another fact explaining why the complicated influence of chloride has not been discovered earlier. Activity is also biphasic with sulphate but reversed, while acetate induces only an inhibition by 50 %. The activity unifies for the W110P mutant, there is only a slightly increased activity followed by a more or less strong decrease. Therefore, I conclude that the observed effects on the wild type are genuine anion effects. However, before further analysing or refining the assay, a binding assay was developed which should confirm the observed activity as a function of one or two binding events of the corresponding anions.

2.5.4   Chloride K_ds are in the low to mid millimolar range

The proteins were purified as described and labelled with Sulfo-Cy5 NHS-ester overnight. With 1.81 and 1.91 fluorophores per wild type or W110P tetramers, respectively, the label efficiency was similar in both cases. The binding of different anions to the tetramers was analysed in 30 mM HEPES/KOH pH 7.6 and 0.05 % Tween-20 in capillaries using a Monolith NT.115. In addition to sodium chloride and bromide, labelled GgpS was also titrated with sodium sulphate and acetate.

The presented fluorescence data are the first experiments and each data point represents one measurement only. Due to the preliminary nature of this assay, the protein concentration was changed/optimised between individual experiments: 8.2 nM (wild type with chloride), 9.2 nM (wild type with all other anions) or 8.0
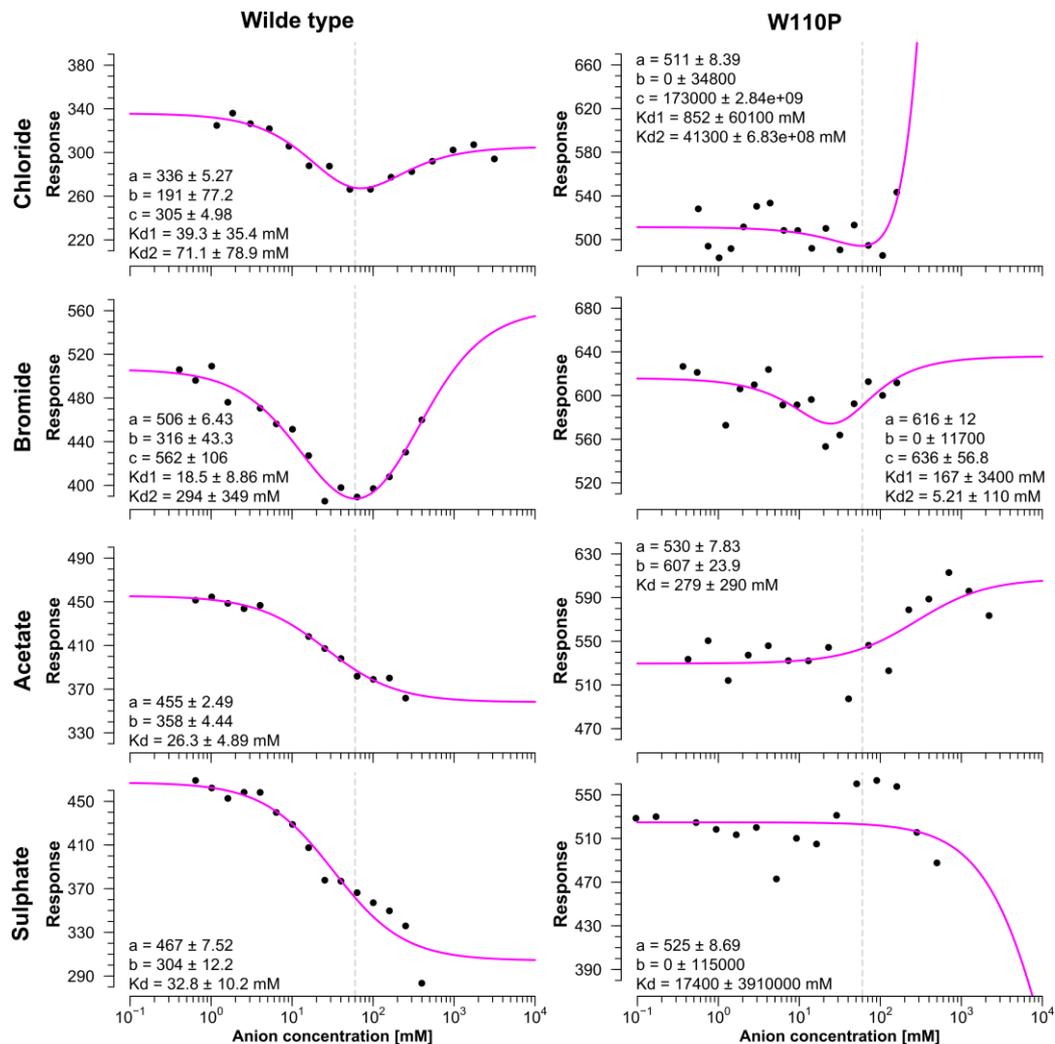
Figure 31: Fluorescence of Cy5 coupled GgpS titrated against chloride, bromide, acetate and sulphate.

8.0 nM W110P mutant and 9.2 nM wild type except for chloride (8.2 nM). Vertical dashed line is at 60 mM. The response axis includes always a range of 180 but the starting point was chosen to show all samples.

nM (all W110P). In addition, the tested concentration range it not yet optimal.

The fluorescence of the wild type is dependent on the present anion and its concentration (Figure 31). Chloride and bromide exhibit both a biphasic binding behaviour. With increasing anion concentrations, the signals decrease to a minimum around 60 mM and eventually rise again to around approximately starting levels. As a rather crude estimate, the inflection points are below 10 mM and between 200–300 mM. Sulphate and acetate, however, cause a monophasic behaviour with the inflection points around 20–30 mM. The W110P mutant does not show any specific response about species and quantity of the anion, in fact they show a random

distribution apparently. The only exception is acetate which see-mingly responds in a dose-dependent manner, however, the spread of the samples is much higher than in the wild type and the pattern may appear as saturation curve by chance. Again it must be stressed that all data points are only single measurements and that these are preliminary results. Besides the the W110P mutant, a subsequent denaturation of the wild type diminished any effects (Figure 77), hence further indicating a structure-dependent effect. This also explains why a quenching and enhancement of fluorescence occurs in the first place. Structural changes shield or expose the fluorophores from/to the solvent, altering the quantum efficiency.

By mutating Trp110 to proline we disrupted the anion binding site and I concluded that the W110P mutant is a good model for an anion binding deficient GgpS (2.3.3). Accordingly, the response of the mutant is independent from the salt/anion concentrations. The binding of two chloride or bromide anions can be explained by different responses of the three possible states of GgpS: (i) empty, (ii) one site occupied, (iii) both sites occupied. A model explaining the observed responses must accommodate the fraction of each state present at a certain point in this three-state equilibrium. I developed a simple model based on the law of mass action assuming each degree of occupation coincides with only one state/conformation and transition only occurs obligatory upon binding or release (section 8.4.2). This model comes naturally with one dissociation constant (K_d) per transition, or one per binding site. Chloride and bromide were fit against equation (14) representing said three-state equilibrium and acetate and sulphate against equation (17) representing only a simple two-state equilib-rium with one K_d.

This yields K_d s in the range of 20–40 mM for the first chloride/bromide binding event. Fitting the binding of the second anion becomes more unreliable with a K_d of 71.1 mM (standard error of 78.9 mM) and 294 mM (349 mM), respectively. Fitting the W110P data usually yields high K_ds with very high standard errors

both in the molar range. This corresponds to the seemingly random distribution of the data.

For both, chloride and bromide, the first k_d is smaller than the second. This indicates a negative cooperative effect and thus an interaction between both sites. A crosstalk between both chloride sites has already been hypothesised in section 2.4.2 of this thesis for the W110P mutant and the chloride-free state. However, response saturation was not achieved and the higher second K_ds go in hand with higher standard errors. First and second K_d may indeed be equal eventually.

Sulphate has a 28 % bigger thermochemical radius (2.31 Å) than chloride (1.81 Å) and apparently hinders binding at the second site at all. Acetate on the other hand has a thermochemical radius of only 1.86 Å, just 3 % bigger (Simoes et al., 2017). However, it has similar effects as sulphate. This may be explained by its shape and atomic composition. The thermochemical radius of an ion is more comparable to the radius of gyration of an object. Both are derived from an inherent property across an assumed spherical shape. This may be a good assumption for a tetrahedral ion like sulphate, but is questionable in case of an anisotropic anion like acetate. Here the charge is located to one side of the ion while the other has an apolar methyl group. Thus acetate may have a different effect as the thermochemical radius imply.

The K_ds in the millimolar range are rather high for typical protein:ligand interactions. However, these concentrations are of physiological relevance as detailed later and can be explained by structural data. First, the chloride is only bound by two coordination partners. This occurs only in less than 10 % of all PDB structures (section 2.3.2; Carugo et al., 2014). The expectation, from the data in the PDB, is a coordination by four (35 %) or three (27 %) partners. Second, the unoccupied binding site is potentially flexible (wt$_{SO4,SUC}$, wt$_{SO4,GOL}$) or even adopts a different fold (W110P$_{Cl,SUC}$; Figure 18), thus for successful binding, one chloride and both parts of the binding site must come together. These two features would result in a low affinity binding site.

Comparing the results with the activity we can notice common elements. Chloride and bromide anions show both a biphasic character in the interaction and activity assays and acetate provokes always a monophasic response. Only sulphate does not fit since it is now monophasic, though, the important last samples of the activity assay have higher variances and thus might be wrongly interpreted. The first more reliable K_d is always estimated to be around 20–30 mM with about 10 mM standard error. This coincides roughly with the turning points of the activities. However, minimum activity of chloride would be expected between both K_ds i.e. at around 55 mM but the errors of the fits are currently as high as the values itself. Hence, this assay is appropriate for the characterisation of the anion interaction with GgpS but has to be further refined.

2.5.5 Fitting the activity against the two-K_d model

In order to fit the activity data and get a better understanding about the activity of each GgpS version, I applied the previously developed two-K_d model (section 8.4.2). The response factors described the expected fluorescence, now they correspond to the k_{cat} values or the relative activity. In regards to fitting constraints, the relative activity of unbound GgpS is assumed constant since the data are normalised to their first data point, representing almost that state. Furthermore, the effect of half as well as fully occupied tetramers and the K_ds must be larger than zero. Technically, the K_ds from the interaction assay should be used here, however, as their estimation is preliminary I decided to fit them as well.

Because the acetate series was originally normalised to its 8 mM sample, the effect factor for unbound GgpS was not constrained except from being positive. In addition, assuming all anions bind at the same site, having 8 mM acetate in each sample creates a competitive situation. This was not modelled and is therefore especially for lower estimated K_ds an important error source.

In a first attempt, all averages and normalised values were fitted (Figure 28 and Figure 29). The activity factors for chloride

indicate the fully occupied tetramer to be more active than the chloride-free GgpS, as hypothesised previously based on the crystal structures (section 2.4.2). The difference is low with an activity of unity and 1.45 (standard error: 0.11), especially if the unoccupied state is assumed to be slightly inhibited because of the actual presence of acetate. The first K_d is much smaller with 4.75 mM (3.50 mM) compared to 39.3 mM (35.4 mM). On the other hand the second is slightly higher with 90.7 mM (45.5 mM) compared to 71.1 mM (78.9 mM; Table 2). In good agreement with the structure-based hypothesis are the results of bromide. The fully occupied state is more than twice as active with 2.24 (0.16), while the half-occupied state is inactive, although with a huge error (2.71). The first K_d is at 214 mM with a high error (764 mM), which originates most likely from the fact that the first phase is shown in only two samples. The second K_d is at 15.6 mM (42.7 mM). Therefore, the hypothesis that only different K_ds explain the strong difference between the chloride and bromide response cannot be validated due to high standard errors. Interestingly, the K_d s of the interaction assay are quite similar but reversed. The K_ds for acetate and sulphate are in the very low millimolar range with 15.4 mM (5.1 mM) and 1.57 mM (2.44 mM) compared to around 20–30 mM from the interaction assay.

As mentioned, also the W110P mutant exhibits an ionic strength dependence. This can also be present in the wild type overlapping with the specific anion binding effect and impede its estimation. A simple correction would be based on a multiplicative model, wt_{md} in equation (1), where ionic strength either enhances or attenuates the binding site specific activity. However, the exact effect is unknown and the stabilisation of α3 is most likely the critical chloride-induced change. It can be achieved by either binding site occupation or the mere presence of a stabilising agent, i.e. kosmotrope. Hence, the effect might be additive and thus model wt_{ad} (2) would be more suitable. In that case, although GgpS should by totally inactive under certain conditions, α3 may be partially de- or stabilised chloride-site independent by surrounding ions including sodium.
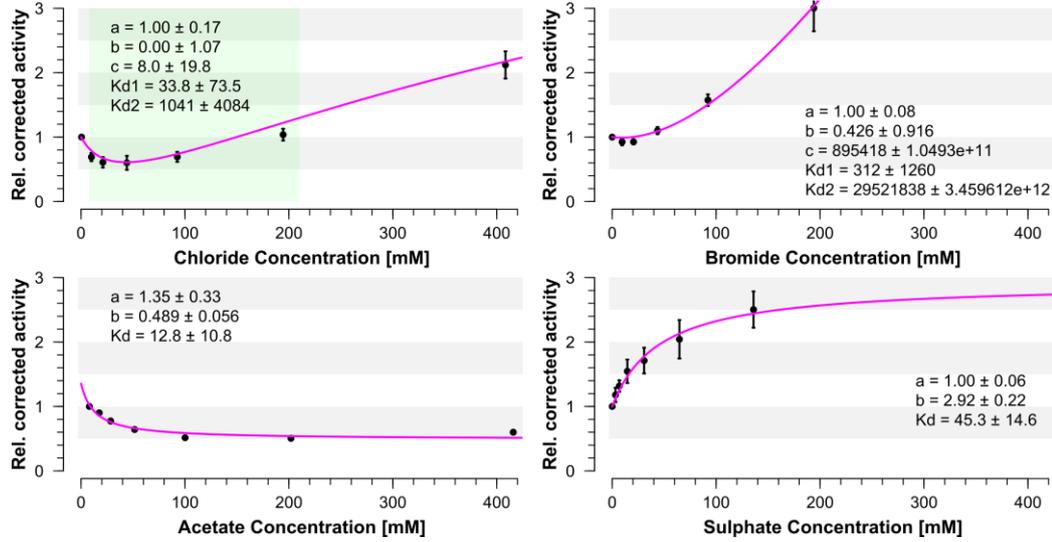
Figure 32: GgpS activity, multiplicative ionic strength deconvoluted (wt_{md}).

Green highlighted area: Minimum physiological range of intracellular chloride concentration in *Synechocystis* PCC6714 (Reed et al., 1985).

$$wt_{md,i} = \frac{wt_i}{W110P_i / W110P_0} \qquad (1)$$

$$wt_{ad,i} = wt_i - (W110P_i - W110P_0) \qquad (2)$$

Both types of deconvolutions lead to monophasic sulphate responses within the measured range (wt_{md}: Figure 32, wt_{ad}: Figure 33). The K_d of the multiplicative correction fits the fluorescence quenching data within errors but the activity increases to 2.92 (0.22) while the fit with the additive model yields very low errors with an activity of only 1.96 (0.03) and a lower K_d of 11.6 mM (1.61 mM; Table 2). The estimations for acetate is similar with activities of 1.35 (0.33) and 1.43 (0.64) for the actual anion-free GgpS and 0.4–0.5 for GgpS bound to one acetate. In addition, the K_d s fit the fluorescence quenching data (section 2.5.4) within errors. The fit for bromide of the multiplicative deconvoluted data fails. The W110P mutant is strongly inhibited at high salt concentration and thereby requires the wild type to be extremely more active. The dip in activity at low bromide concentrations eventually becomes insignificant like noise and thus the activity is best described by just one K_d, a clear contradiction to the fluorescence quenching data. The additive deconvolution, however, leads to K_d s of 58.6 mM (standard error: 59.5 mM) and 95.6 mM
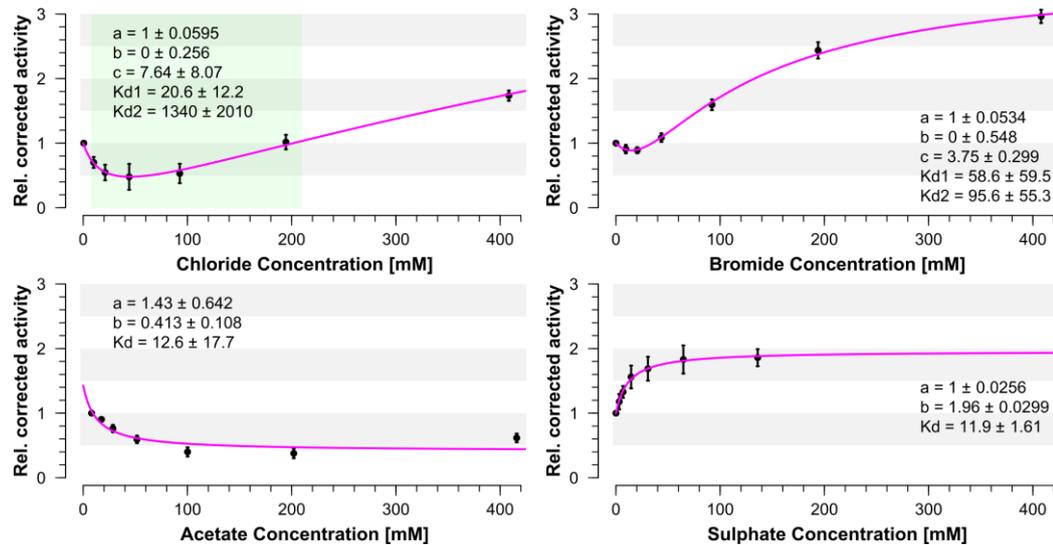
Figure 33: GgpS activity, additive ionic strength deconvoluted (wt_{ad}).

Green highlighted area: Minimum physiological range of intracellular chloride concentration in *Synechocystis* PCC6714 (Reed et al., 1985).

(95.3 mM). They are more similar to the fluorescence quenching with the same apparent type of cooperativity but high errors. For chloride both deconvolutions are similar again. The activity of the one-chloride bound tetramer becomes 0 with standard errors of 1.07 (wt_{md}) and 0.256 (wt_{ad}). This agrees with bromide convoluted and additive deconvoluted estimates. On the other hand, the fully occupied state becomes highly speculative, probably because K_{d_2} is in the 200 mM range and the assay lacks samples there.

Salts are known to impact protein stability (Hofmeister, 1888; Lo Nostro & Ninham, 2012) and, as described in section 2.4.2, the chloride sites regulates the activity by modulating the stability/flexibility of the active site. According to this, bromide should have the most adverse effect on protein stability of the tested anions and indeed, bromide is inhibiting the activity of the W110P mutant most. However, the effect may be counteracted by specific binding at the 'chloride' site. Thus, a deconvolution may not be necessary at all or more likely overestimates true activity. Deconvoluting via wt_{ad} is milder than via wt_{md} and thus seemingly better. The same might be true for sulphate but in a twisted way. According to Hofmeister, sulphate should stabilise thus increase activity but it may also compete with (ADP) phosphate group binding in the active site, decreasing activity. The latter effect

Table 2: Overview about all fit-parameters from the activity and interaction assays.

Chloride

	Activity	wt_{md}	wt_{ad}	Interaction
a	1.00 ± 0.05	1.00 ± 0.17	1.00 ± 0.06	336 ± 5.27
b	0.423 ± 0.135	0.00 ± 1.07	0.000 ± 0.256	191 ± 77.2
c	1.45 ± 0.11	8.0 ± 19.8	7.64 ± 8.07	305 ± 4.98
K_{d_1}	4.75 ± 3.50	33.8 ± 73.5	20.6 ± 12.2	39.3 ± 35.4
K_{d_2}	90.7 ± 45.5	1041 ± 4084	1339 ± 2009	71.1 ± 78.9

Bromide

	Activity	wt_{md}	wt_{ad}	Interaction
a	1.00 ± 0.06	1.00 ± 0.08	1.00 ± 0.05	506 ± 6.43
b	0.00 ± 2.79	0.426 ± 0.916	0.000 ± 0.548	316 ± 43.3
c	2.23 ± 0.16	———	3.75 ± 0.30	562 ± 106
K_{d_1}	218 ± 795	312 ± 1260	58.6 ± 59.5	18.5 ± 8.86
K_{d_2}	15.0 ± 42.3	———	95.6 ± 55.3	294 ± 349

Acetate

	Activity	wt_{md}	wt_{ad}	Interaction
a	1.25 ± 0.10	1.35 ± 0.33	1.43 ± 0.64	455 ± 2.49
b	0.541 ± 0.021	0.489 ± 0.056	0.413 ± 0.108	358 ± 4.44
K_d	15.4 ± 5.1	12.8 ± 10.8	12.6 ± 17.7	26.3 ± 4.89

Sulphate

	Activity	wt_{md}	wt_{ad}	Interaction
a	1.00 ± 0.16	1.00 ± 0.06	1.00 ± 0.03	467 ± 7.52
b	1.52 ± 0.10	2.92 ± 0.22	1.96 ± 0.03	304 ± 12.2
K_d	1.57 ± 2.44	45.3 ± 14.6	11.9 ± 1.6	32.8 ± 10.2

would be similar in both proteins but the stabilisation might by be compensated by sulphate binding in the 'chloride' site partially. Thus a deconvolution should only correct at higher concentrations to as smaller degree, hence the milder additive model is better again.

According to my structure-based prediction, the chloride-free GgpS should be more flexible and hence less active than the fully chloride occupied one. The activity assay indeed supports 15–25 % less activity. While the structures of these states do not differ strongly, fluorescence quenching of the one chloride/bromide occupied state suggests conformational changes which coincides

with strong if not total activity loss. Qualitatively spoken, the similar sized acetate inhibits as the first bound chloride or bromide does, while the much larger sulphate solely (hyper-) activates like the binding of the second chloride or bromide. Nonetheless, both prevent binding at the second site and have therefore a distinct different effect from chloride and bromide. Measuring the true effect of chloride site specific binding alone is challenging. The chloride binding site is able to bind all tested ions including acetate and the overall activity is probably affected by Hofmeister or other ion dependent effects.

2.6 Chloride binding modulates DNA inhibition

Previously it was revealed that GgpS is inhibited by DNA, and further analysis generalised this to poly-anions like heparin and the phosphate backbone of nucleic acids (Jens F Novak et al., 2011). Hence, the sequence of the nucleic acid or even the direct shape of the molecule is less important. The inhibition is eventually relieved by electrostatically shielding GgpS from the inhibitor with small ions like sodium and chloride. However, *Synechocystis sp.* PCC6714 has usually an ionic strength of around 100 mM of small ions, predominantly potassium, in its cytoplasm (Reed et al., 1985). Moreover, the DNA release mechanism overlaps with the DNA independent chloride binding effect. Hence the question arises if the DNA release itself is more complex and modulated by the chloride binding sites or even not present at all. The introduced system of wild type and W11OP mutant can help answering this.

While titrating GgpS with different anions, two further samples without and with 194 mM of additional ionic strength in presence of 1.25 mg/ml sheared (Figure 78) herring sperm DNA were analysed (Figure 34). From previous studies I expect that DNA in absence of any additional salt/anion inhibits strongly, while it has no effect if additional ions are present. Likewise, inhibition of the W11OP mutant by DNA is expected, however, the activity should not recover if an occupied chloride binding site modulates DNA inhibition. Independent of the influence exerted by the binding site, the previous mode of action, i.e. electrostatically shielding
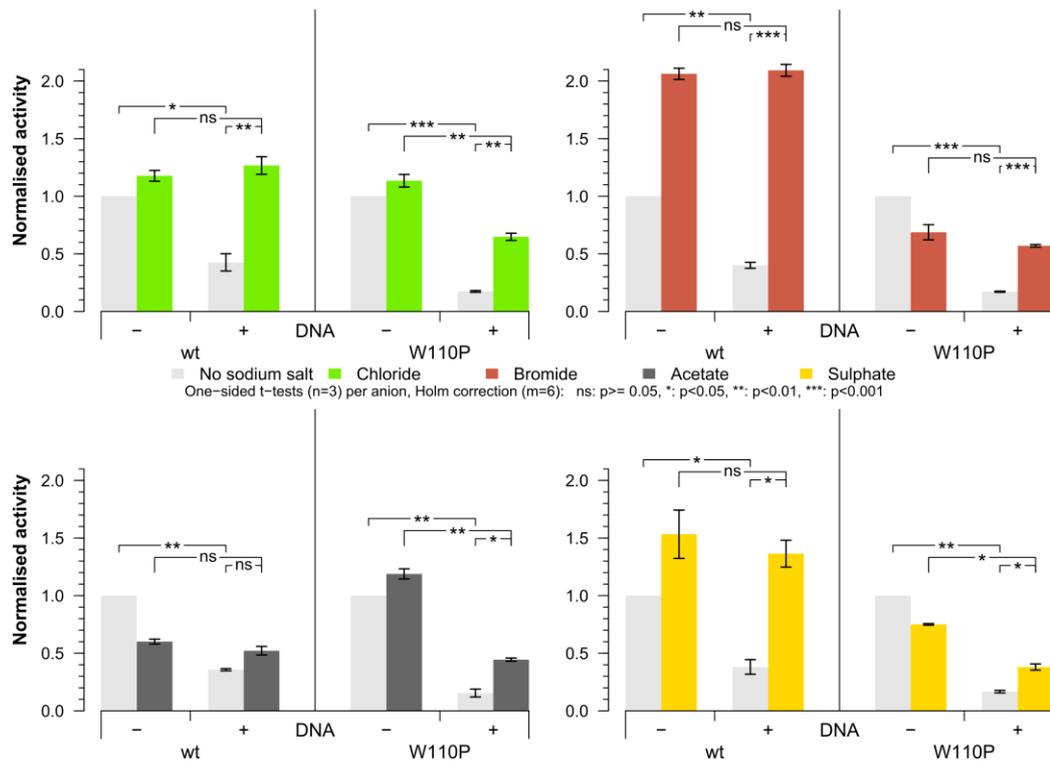
Figure 34: Influence of DNA in combination with different sodium salts on GgpS activity.

Herring sperm DNA was sheared by sonication to 250–750 bp length (Figure 78). Activity was measured in presence of 1.25 μg/μl of that DNA and 194 mM ionic strength of the sodium salts of chloride, bromide, acetate and sulphate. All samples were measured in biological triplicates including their control without any salt. Error bars represent standard deviation and stars indicate p-value (legend) by one-sided t-tests ($n = 3$) with Holm correction ($m = 6$) within groups of salt species.

GgpS from the DNA, might still be (partially) true, so again two effects can overlap leading to partial recovery only if the chloride-binding sites are missing. Accordingly, GgpS was tested in absence and presence of DNA and sodium chloride or the three other sodium salts. Differences between key conditions according to expectation were evaluated by Holm corrected (Holm, 1979) one-sided t-tests with individual variances. The used data are the averages of the raw data without normalisation per batch.

As expected, DNA inhibits significantly (p<0.05) down to 42 % (standard deviation: 7 percent points). Further, supplementation with chloride significantly increases (p<0.01) activity again to 127 % (8 points), which is similar to the observed activity without DNA of 118 % (5 points). The W110P mutant is also significantly (p<0.001) inhibited by DNA down to 17 % (1 point). However,

while the activity increases significantly (p<0.01) if chloride is added, the activity does not recover completely and stays at 65 % (3 points). This is 49 percent points lower (p<0.01) than in the absence of DNA but in the presence of chloride (113 %, 5 points). Summarised, wild type and mutant are both inhibited by DNA. However, while the activity is fully recovered if the wild type is supplemented with additional 194 mM sodium chloride, the mutant is still only about half as active. This indicates a combination of both, a specific chloride binding effect and the electrostatic shielding of GgpS from the DNA.

I have shown so far, that the effects of bromide on GgpS alone are very similar to that of chloride. The only difference is a stronger activity increase if the second bromide is incorporated into the tetramer. Hence, I expect full recovery of DNA inhibition. This is indeed true for the wild type. DNA had no effect on activity if about 200 mM sodium bromide was present. However, the W110P mutant exhibited apparently a full recovery as well. W110P's activity decreased (p<0.001) in absence of bromide due to DNA to 17 % (0 points) and recovered (p<0.001) 'partially' back to 57 % (1 point) similarly as with chloride (65 %; 3 points). Though, in contrast to the case with chloride, this is almost the maximum activity at high bromide concentration (69 %; 7 points) and thus not significantly different. I explained in section 2.5.3 that bromide has a strong binding site independent inhibiting effect on the mutant. How this inhibition occurs is unknown and most likely interferes here, explaining the abnormal behaviour of the mutant. I established with chloride that, the binding site independent effect i.e. the electro static shielding effect is not sufficient, hence the full recovery of the wild type by bromide is only explained by a specific bromide binding effect.

The effect of acetate on the other hand is inconclusive. There is no significant recovery nor a significant difference to the DNA-free condition determinable with the wild type. DNA inhibited to 36 % (1 point), acetate alone to 60 % (2 points) and the combination led to an activity of 52 % (4 points). The W110P mutant responded on acetate similarly as on chloride, DNA inhibited to 15 % (3 points,

p<0.01), the addition of acetate recovered (p<0.05) activity partially to 44 % (1 point) while full recovery would be achieved at higher 119 % (4 points, p<0.01). This indicates the binding site independent DNA recovery effect to be present as well. However, whether DNA release is facilitated by acetate binding is indistinguishable. The reason is most likely the inhibition by acetate itself which interferes with the DNA inhibition. On the other hand, a specific acetate binding DNA release can explain why GgpS is not fully inhibited in absence of any extra salt. As mentioned earlier, 4 mM magnesium acetate are present ubiquitously, including the 'no extra salt' samples. Acetate has an K_d of around 20 mM, thus 8 mM could already lift the DNA inhibition slightly.

In contrast to bromide and acetate, sulphate can fully replace chloride.

The activity of the wild type in presence of DNA at the tested ionic strength of 194 mM is only completely rescued if the chloride sites are occupied by chloride, bromide and sulphate while, the binding site deficient W110P mutant only recovers partially. Therefore, I conclude that chloride binding accelerates DNA release, i.e. it enables full activity at lower ionic strengths. Acetate and sulphate can apparently only occupy one binding site simultaneously. The mechanism is elusive so far, but in the case of sulphate, this singular binding event is as effective as binding of two chloride/bromide anions. The response to acetate indicates but does not proof, that binding of one acetate facilitates DNA release as well. In the last section, I concluded that the effect of one sulphate bound to the tetramer mimics to some degree the effect of two chloride/bromide anions and the effect of one acetate that of only one chloride/bromide. Hence, DNA release might be facilitated by the first binding event already. However, binding to one chloride/bromide is simultaneously inhibiting thus no activation would occur.

The results explain why GgpS can be inhibited intracellularly by DNA. The dominantly present potassium can most likely, in

contrast to anions, not bind in the 'chloride' site and has therefore just the binding site independent ionic strength effect. However, 194 mM sodium chloride i.e. 194 mM ionic strength can only rescue partially that way. Hence the 90 mM ionic strength of potassium (plus 10 mM due to chloride and sodium) will not have a strong activating effect in the cyanobacterium and GgpS would remain mostly inhibited.

2.7 Two predicted DNA binding modes can explain inhibition and recovery

2.7.1 Sucrose binding coincides with positive coulombic potential in the pore

The coulombic surface potential of GgpS was analysed for patches of positive charge, since the inhibiting poly-anions are most likely bound by oppositely charged motifs. The outside surface of the toroidal GgpS tetramer is predominantly negatively charged (Figure 35). Positive patches are only found around the active site pockets and within the central pore of the tetramer. Coincidentally, besides a sucrose molecule in the active site, two more sucrose molecules per subunit (1002 and 1003) were found to be bound in the pore at each monomer. These sucrose molecules, in total eight molecules in the complete tetramer, can be divided into two groups of which each lie on a straight line. Together they form a right-handed double helix-like appearance with approximately 21–26 Å inner and 36–37 Å outer diameter. While the handedness fits the commonly found A and B forms of RNA and DNA, their outer diameter is only around 20 Å.

This brief analysis reveals two distinct possible binding modes. The first is on top of the active site pocket, thereby probably blocking substrate entrance and explaining DNA inhibition. However, why chloride binding affects the DNA interaction is not obvious. In the second binding mode, the nucleic acid double helix is encompassed by the tetramer and it assumes the positions of the sucrose molecules bound in the pore. One of these sucrose molecules is the one close to Trp110 and as mentioned briefly, the
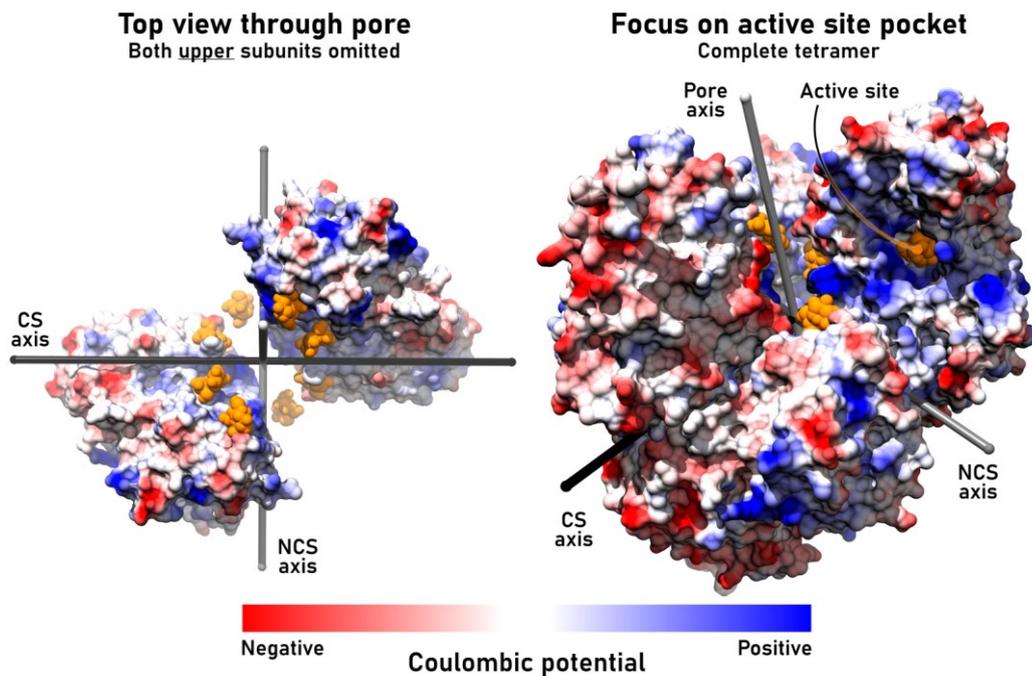
Figure 35: Coulombic surface potential of GgpS (wt$_{Br,SUC}$).

Positive patches are found in the pore and around the active site entrance. Eight sucrose molecules (SUC1002 and SUC1003 from each subunit, orange spheres) form a right-handed double helix in the pore. For better visualisation, the two upper subunits have been omitted in the top view (left). CS axis: crystallographic axis through chloride binding site, NCS axis: non-crystallographic symmetry axis.

chloride site is located in the pore side of the torus. Thus, the mechanism how chloride binding enforces DNA release becomes trivial and distortions of the tb-loop due to DNA interaction may explain the inhibition.

2.7.2   Early crystallisation attempts of GgpS with DNA were unsuccessful

Long overhanging DNA usually interferes with crystallisation, hence the shortest still binding single stranded DNA (ssDNA) is preferred for setting up crystallisation experiments. According to Novak et al., ssDNA homo-oligomers with six nucleotides (either adenine, thymine or cytosine) inhibit GgpS at least partially (Jens F Novak et al., 2011). Preliminary binding experiments with fluorescently labelled ssDNA of this length but with randomised sequences were performed via size exclusion chromatography. They confirmed binding of hexamers, octamers and decamers in an approximate 1:1.3 ratio of GgpS tetramers to ssDNA (data not shown, Seyma Bozkus). Crystals grown in the presence of these oligonucleotides were of very low quality and not mounted or

analysed. Eventually, lab priority was shifted towards finalising the activity assay. However, a docking analysis of DNA/RNA with ClusPro was performed.

2.7.3   Docking: Introduction to ClusPro and preparation of results

For docking, the tetrameric assembly of $W110P_{Cl,suc}$ with removed sucrose molecules was chosen. It should bind nucleic acids more likely according to the activity assays. Nevertheless, initial docking was performed with the wild type yielding similar results but only those from W110P are presented in detail. As binding partner, ideal DNA or RNA molecules generated by Coot were submitted to the ClusPro server (Kozakov et al., 2017). All nucleic acids are based on the same 30 nucleotide long random sequence which was truncated or modified appropriately. The server supports officially only RNA but also DNA is accepted and is correctly processed. A question about the usage of DNA was unanswered by the developers. Hence, the results with DNA should be critically judged. However, according to publication about ClusPro, nucleic acids are submitted as 'receptors' and are un-changed during processing and only the 'ligand' is energy minimised anyway.

For average sized proteins, ClusPro generates $10^9–10^{10}$ starting combinations of receptor and ligand. These combinations are energy minimised under consideration of four energy models differently weighting electrostatic, hydrophobic and DARS poten-tials against each other. Simplified, the DARS potential favours fitting shapes (Chuang et al., 2008). Eventually, the 1,000 lowest energy assemblies are clustered and ordered by population size, while the structure with the highest number of neighbours within a cluster becomes the cluster centre.

ClusPro outputs the top 30 cluster centre and usually only a few densely populated clusters should indicate a good docking result. However, since the tetramer is D2 symmetric, at least two symmetric equivalent assemblies can be expected with GgpS. Moreover, a frame shift like phenomenon led sometimes to differ-ent clusters, although they show effectively the same binding

mode. Therefore, a three-dimensional histogram was calculated: The space was evenly distributed in voxels of 1.5 Å edge length and the number of nucleic acid atoms across all clusters, weighted linearly by cluster size, were counted per voxel. This histogram was visualised with each voxel, i.e. bin, represented by an individual atom whose depicted radius scales with bin frequency. Thus, overlapping binding modes get highlighted while symmetrically equivalent or competing modes are depicted in similar strength. On the other hand, less likely binding modes may appear but will be less noticeable.

2.7.4 Docking: Length of double stranded nucleic acids determines predictions

Since GgpS is inhibited by poly-aniones, the 'electrostatic-favoured' model was chosen. From all analysed nucleic acids, only the short ones docked successfully. Figure 36 depicts the 3D histograms in top view, since in any case the nucleic acids dock into the pore. Interestingly, they tend to cluster at the left dimer pair of the W110P mutant. This pair is in its apparent inactive conformation, while the right dimer pair is in its active wild type-like conformation (2.4.1). On the other hand, the active sites of the active pair attracts the termini of $ssRNA_{10}$ and to some extend $ssDNA_8$.

By default ClusPro docks with four different energy minimising models, the electrostatic-favoured is one of them. A brief analysis of the failed docking attempts reveals often successful dockings using the energy model ignoring the DARS potential, i.e. relying entirely on modelled physical interactions (Chuang et al., 2008). Figure 37 depicts the 3D histograms, but this time in side view along the chloride binding sites connecting axes. The single stranded nucleic acids dock very similar to the electrostatic-favoured model in the pore. Double stranded nucleic acids on the other hand dock above and below the pore, parallel to the previously defined NCS axis (Figure 35). dsDNA interacts with the extensions above the pore from similar facing subunits of the D2 tetramer (compare Figure 6), thereby bridging the pore itself
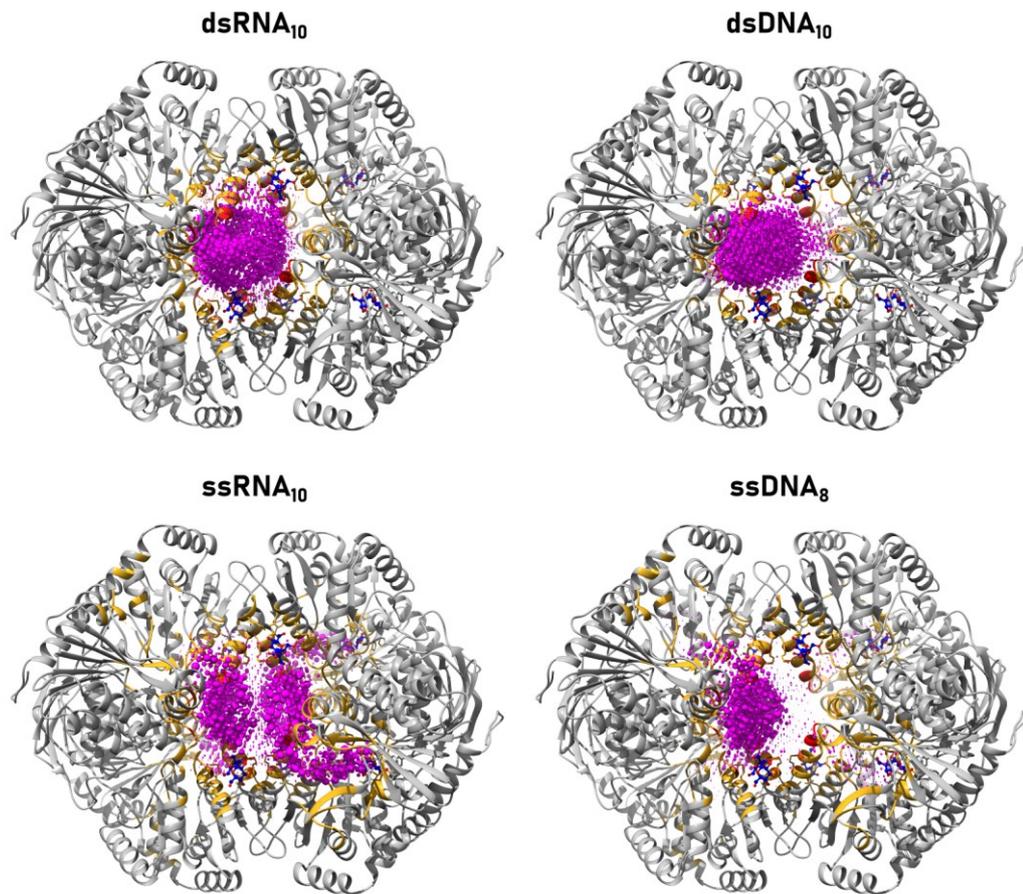
Figure 36: All results from ClusPro docking of W110P$_{Cl,SUC}$ with the 'electrostatic-favoured' model as 3D histogram.

Top view through pore. dsRNA and dsDNA with 10, 20 and 30 base (pair) length as well as ssRNA with 10 and ssDNA with 8 bases were submitted. RNA was supplied as A form and DNA as B form. However, in addition, dsRNA$_{30}$ in B form was submitted as well. All not depicted docking combinations failed. All nucleic acids dock into the pore. Radius of the spheres represent cluster-size weighted bins of the histogram. Blue ball and sticks: Sucrose of W110P$_{Cl,SUC}$ not present during docking.

covering the active sites. dsRNA however is still pore centred. It cannot enter but one end docks directly above the pore and the other points around 30°–70° away from the pore axes (Figure 37, bottom).

Two different binding modes were determined. Longer nucleic acids apparently favour the binding across and smaller ones into the pore. ClusPro does not modify the nucleic acids, thus a somehow bent nucleic acid is not modelled. This potentially explains why longer DNA/RNA are not docked to or through the pore as further detailed in the next section.
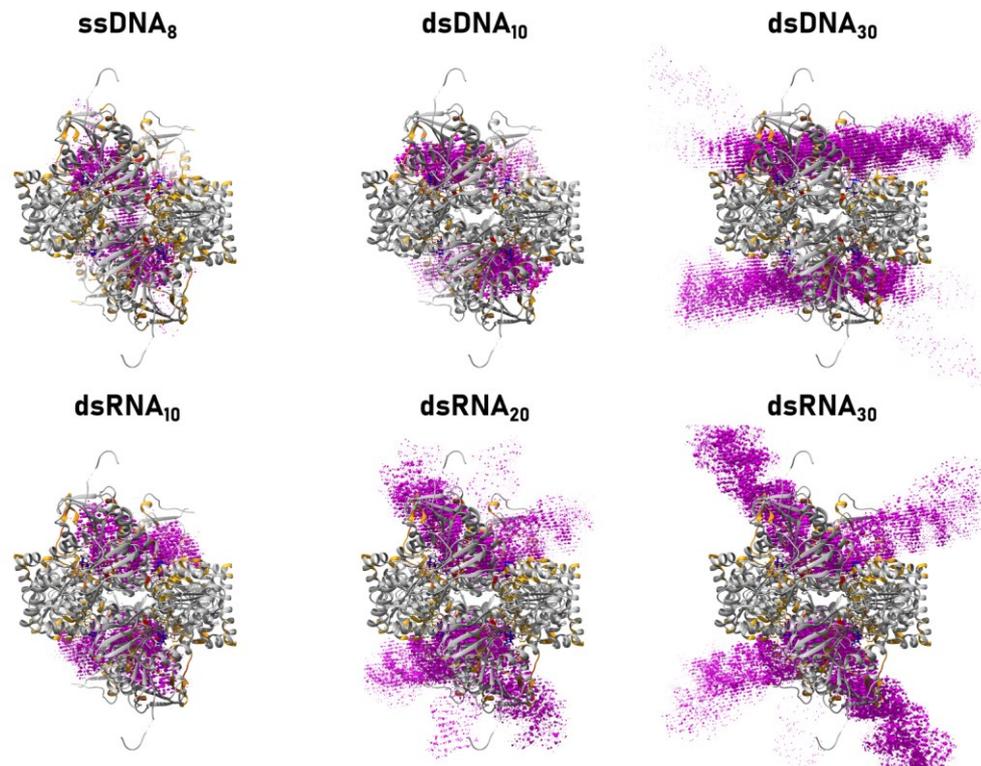
Figure 37: All results from ClusPro docking of W110P$_{Cl,SUC}$ without DARS potential as 3D histogram.

Side view along axis connecting chloride-binding sites. dsRNA and dsDNA with 10, 20 and 30 base (pair) length as well as ssRNA with 10 and ssDNA with 8 bases were submitted. RNA was supplied as A from and DNA as B form. However, in addition, dsRNA$_{30}$ in B form was submitted as well. All not depicted docking combinations failed. Only ssDNA$_8$ docks into the pore, all other nucleic acids prefer the cleft above and below the actual pore. dsDNA lies parallel to the previously defined NCS axes bridging the pore while dsRNA terminates above/below the pore and molecules from different docking results create an angle of 115°. Radius of the spheres represent cluster-size weighted bins of the histogram. Blue ball and sticks: Sucrose of W110P$_{Cl,SUC}$ not present during docking.

2.7.5   Docking: Pore and active pocket binding mode in detail

dsDNA$_{10}$ in the electrostatic-favoured docking target is located inside the pore. The best two clusters show very similarly docked dsDNA molecules, which are only rotated with respect to each other by a few degrees. As mentioned, the double helix is closer to the inactive dimer. One end of a dsDNA$_{10}$ molecule extends out of the pore, the other does not because it would clash with protein parts on the other exit of the pore. Nevertheless, due to the symmetry of GgpS, the DNA can be placed in an opposite mode protruding the
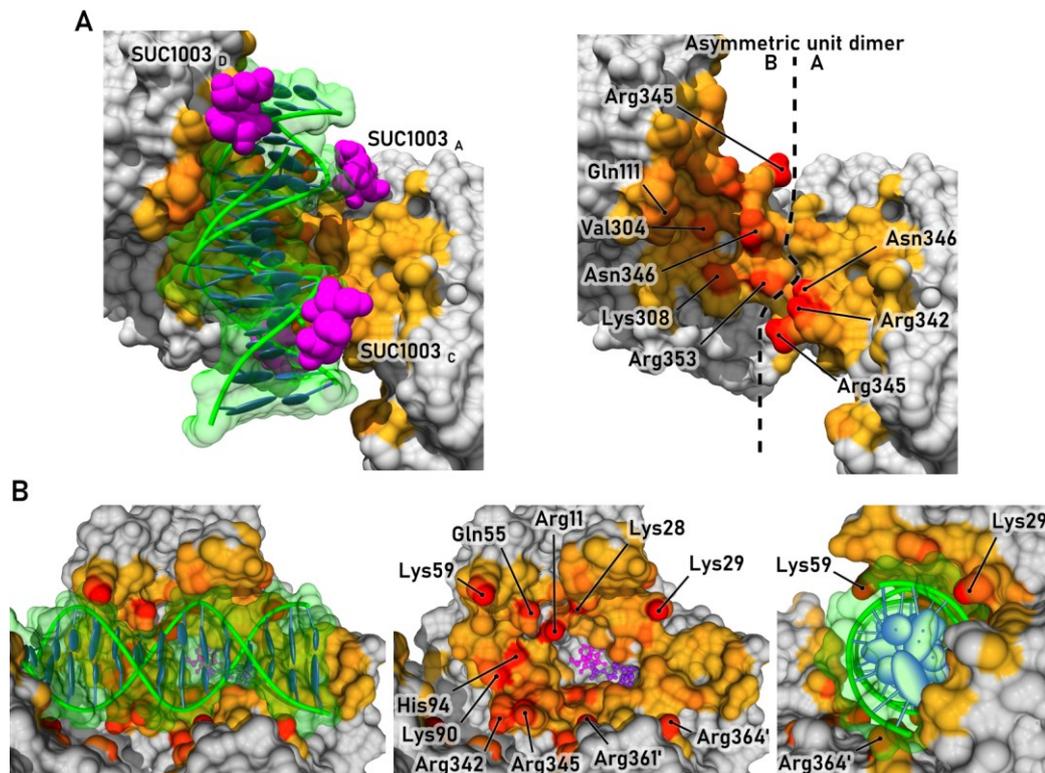
Figure 38: Top ClusPro results for dsDNA$_{10}$ (electrostatic-favoured) and dsDNA$_{30}$ (without DARS potential) with W110P$_{Cl,SUC}$.

Docked nucleic acids are depicted in green with transparent surface. Grey surface is from docking partner with a yellow to red gradient indicating the atoms highly interacting with nucleic acid molecules atoms weighted linearly by cluster size (truncated to 0–99 % quantile of final score). A) Pore binding mode. dsDNA$_{10}$ (clusters one and three) are visualised on the left, while omitted on the right. Predominantly positions 342–353 (α12) are involved. The double helix follows approximately the sucrose molecules. Sucrose was not part of the docking and is only provided for reference. Also position of Gln111 is indicated. B) Pocket binding mode. dsDNA$_{30}$ visualised on the left, while omitted in the middle. On the very right: Rotated by 90°. As depicted, the nucleic acid covers and blocks the active site pocket (with sucrose and ADP as reference) but is not long enough to block the other simultaneously. Important amino acids are part of the protruding β-sheet β1/2 (11, 28, 29), β4/α2 (55–59), α3 (90–94), β14/α12 (342, 345) and α12 (361, 364) of the adjacent subunit. α3 runs parallel to the pore axes on the left side of the active site pocket (Lys90 and His94, terminates below Lys59.

opposite pore exit as realised in cluster three. Figure 38A depicts the W110P$_{Cl,SUC}$ with the first and third docked clusters. The double helices overlap by 5 bp and their backbones align very well. Again, owing to the symmetry, the overlap occurs exactly in the middle of the pore, i.e. close to the chloride binding site. Most interacting amino acids are part of helix α12 (342–353), but also Gln111 is under the top ten interactors (Table 3). The DNA binds indeed parallel to α3 and the tension belt loop and it is almost parallel to the pore axis (Figure 39). Hence, inhibition and chloride site
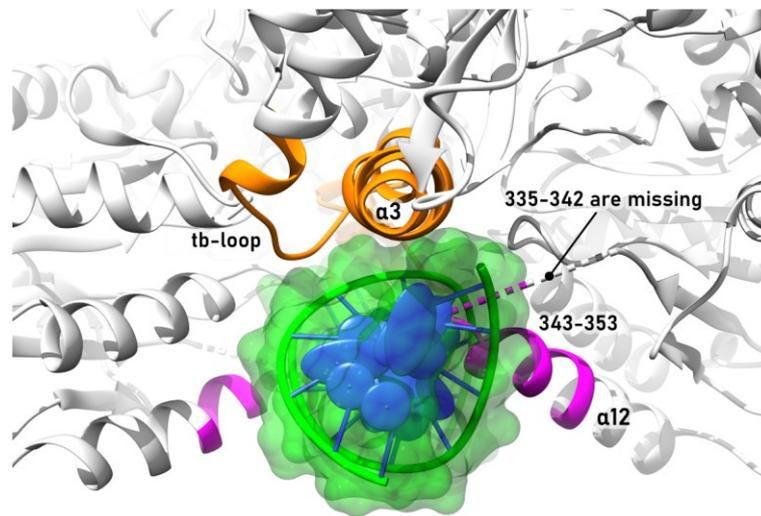
Figure 39: First cluster of dsDNA$_{10}$ (electrostatic–favoured model) in top view.

The perspective is aligned with the longitudinal axis of the nucleic acid. It is docked parallel to α3 (orange), whose N-terminus is at the top. The tb-loop (orange) is downstream of α3. The N-terminus of α12 (magenta) is inserted into the major groove of the DNA. Amino acids 342–335 were not built in that subunit. The docked nucleic acid is depicted in green with transparent surface.

dependent DNA release may be accomplished both via remodelling of the tension belt loop and α3. The N-terminus of α12 (343–353) is inserted into the major groove of the DNA (Figure 39). Amino acids 342–335 were not built in that subunit and could potentially prevent this binding mode. However, they were not build because there is no electron density due to high flexibility. Moreover, they are partially build and very flexible in the assumed anion-free structures (Figure 75, Figure 76) and probably adjustable for DNA binding. Interestingly, Tyr344 can interact via a water molecule with Lys90 and Arg300 in the wild type (Figure 21D).

Without DARS potential, dsDNA$_{30}$ is posed in a way that it blocks the active site of at least one subunit (Figure 38B) and it extends across the pore close to the active site of the adjacent subunit. Unfortunately, this subunit is missing amino acids 14–33, i.e. an essential part of the active site pocket. Hence, this pocket appears as being not blocked. Moreover, the DNA is not placed symmetrically, as is visible in Figure 37. The active site blocking terminus of the DNA attaches closely to the protein, while the other one does not. Hence, the closely bound end can probably not be extended without clashes while the other does not completely block the active site pocket. However, docking with the wild type yielded
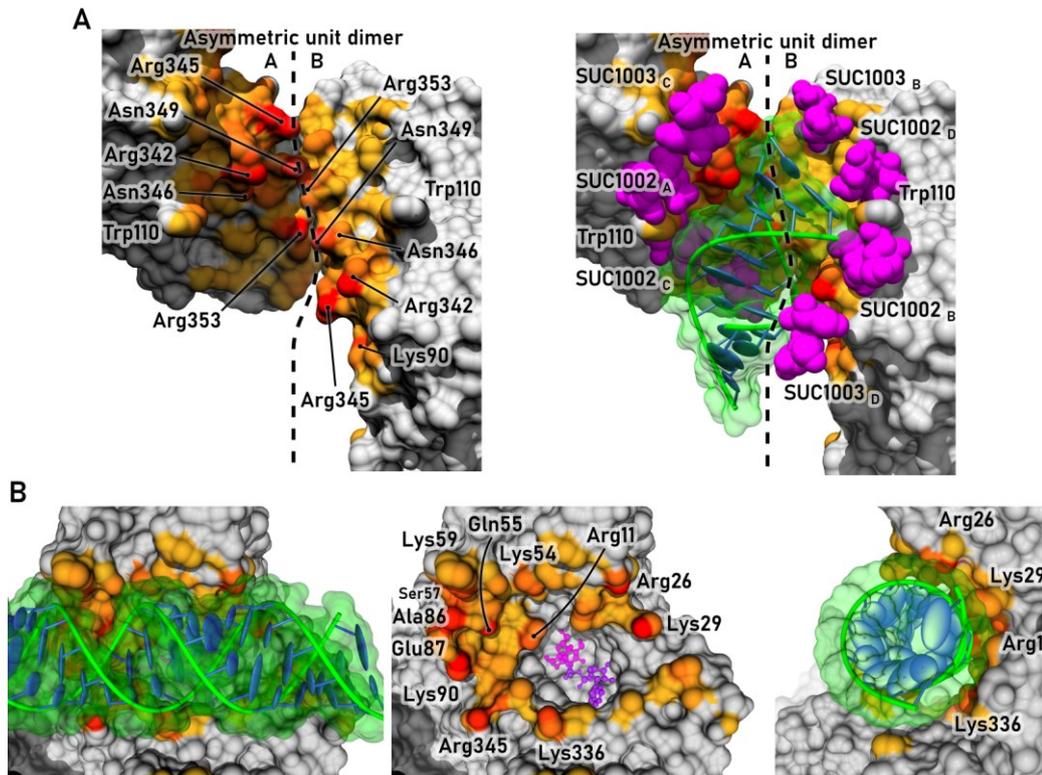
Figure 40: Top ClusPro result for dsRNA$_{10}$ and dsRNA$_{30}$ (A form in contrast to usual B form DNA) with wt$_{Br,SUC}$.

Colours as Figure 38. A) Pore binding mode. dsRNA$_{10}$ visualised on the left, while omitted on the right. Predominantly positions 342–353 (α12) are involved. The double helix follows approximately the sucrose molecules. Sucrose was not part of the docking and is only provided for reference. Also position of Trp110 is indicated. B) Pocket binding mode. dsRNA$_{30}$ visualised on the left, while omitted in the middle. On the very right: Rotated by 90°. As depicted, the nucleic acid covers and blocks the active site pocket (with sucrose and ADP as reference). Due to symmetry a similar double helix can also block the related active site and properly overlaps with the first. Thus a longer one double helix should be able to block two active sites, however, docking failed. Important amino acids are part of the protruding β-sheet β1/2 (11, 26, 29), β4/α2 (54–59), α3 (86–90) and β14/α12 (336, 345). α3 runs parallel to the pore axes on the left side of the active site pocket (Ala86 to Lys90).

a similar binding mode with dsRNA$_{30}$ instead of dsDNA$_{30}$ (Figure 40). There, RNA binds symmetrically, blocking both sites and the RNA can be extended on either sides. Aside β-sheet β1/2 with linker to α1 (11, 26, 28, 29), also β4/α2 (59), α3 (90, 94) and α12 (340, 342, 345) are under the top ten interacting elements/amino acids (Table 3). The mechanism of inhibition, is obvious, the active site entrance is blocked. Chloride site dependent DNA releases may be induced via helix α3. Interestingly, β-sheet β1/2 was previously assumed to be of unknown regulative purpose (section 2.2.4) as it

Table 3: Important residues in both DNA binding modes and of reported assays (B. M. E. Roenneke, 2014).

Prevents biotinylation	Less inhibition	Same inhibition	Pore ($W110P_{Cl,suc}$)	Pore ($wt_{Br,suc}$)	Pocket ($W110P_{Cl,suc}$)	Pocket ($wt_{Br,suc}$)
59	80	281	304	90	11	11
61	270	281/286*	308	342	28	26
286	281/286*	336	342	345	29	29
308	353*	345	345	346	55	54
336	485	353*	346	349	59	55
		364	353	353	90	59
					94	86
					342	87
					345	90
					361	336
					364	345

*: Partial inhibition; Green: Found by docking and at least one assay; Yellow: Prevents biotinylation but not inhibition; Red: Clear contradiction; Blue: Activity not tested

is highly flexible if comparing the different crystals and it connects directly to α1 which probably locks the substrates in the active site.

During his PhD in the laboratory of Professor Reinhard Krämer, Benjamin Roenneke worked on GgpS and its inhibition by DNA (B. M. E. Roenneke, 2014). He performed two experiments in order to determine the DNA binding site and found nine lysine or arginine residues which influenced DNA activity inhibition or altered biotinylation pattern upon DNA binding (lysines only). Of those nine residues, only three are found amongst the more important ones by the binding prediction inferred from the work presented here (Lys59, Lys308, Arg353; Table 3). The latter two are part of the pore binding mode and Lys59 is exclusive to the pocket mode. However, this lysine is positioned above the pore, thus its biotinylation might be prevented by the binding of a longer dsDNA through the pore. Roenneke's results indicate that the mutation of Arg345 does not alter DNA inhibition, but it is present in all my predictions. The other positions found by Benjamin Roenneke are all on the outer side of GgpS and thus not close to the pore or the active site pocket.

Matthias Mörgelin (Colzyx AB, Lund, Sweden) imaged in collaboration with us GgpS bound to 20 bp long dsDNA labelled
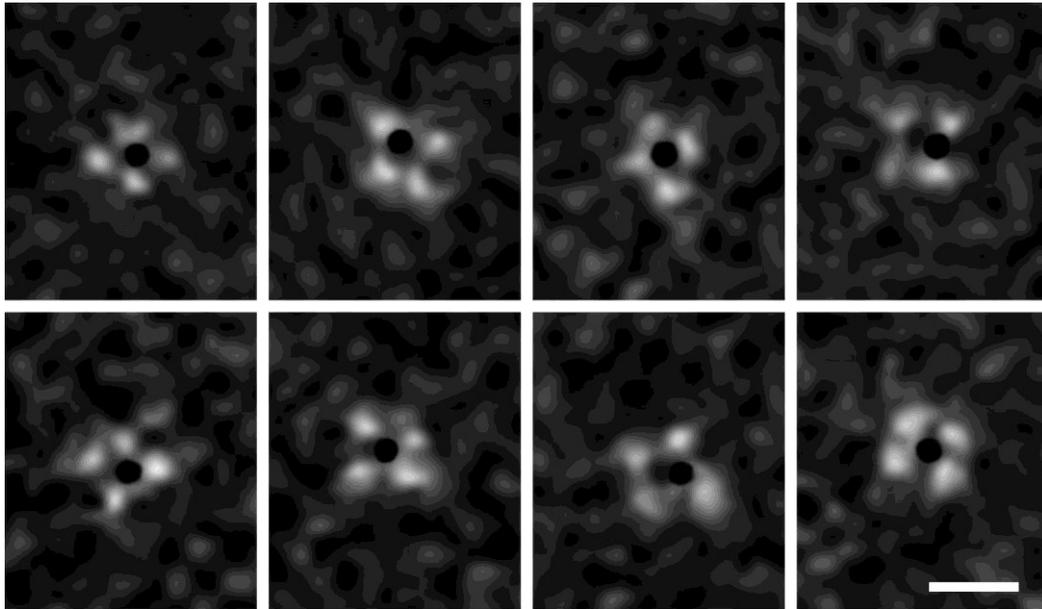
Figure 41: Negative stain electron micrographs of GgpS wild type under low salt conditions in presence of 20 bp dsDNA labelled with 30 Å gold particles.

The particle is apparently predominantly in or above the pore, however, no statistic is available yet. Imaged by Matthias Mörgelin (Colzyx AB, Lund, Sweden).

with 30 Å diameter gold particles via negative stain electron microscopy. Samples were sent to him 1–2 years before the docking calculations were made and the calculations are therefore not based on these predictions results but on the assumptions that dsDNA binds in the pore or to the outer side. Finally, the EM data were acquired only recently without further statistics so far. He found that the gold particles are usually above or close to the pore indicating the pore binding mode is the correct one (Figure 41). However, the exact DNA position is hard to determine. The DNA is not directly visible and with 20 bp rather short. Around 30–40 base pairs are required to bridge the pore and block both active sites with one dsDNA helix. A 20 bp dsDNA could potentially bind only above one pocket with its gold particle over the pore. However, if bound above one pocket only, the gold particle is expected to be at the rim of the tetramer at least in half of the cases but this does not happen.

ClusPro does not modify the 'receptor' during docking and since the nucleic acids have to be submitted as the 'receptor', they remain as perfectly linear molecules. Longer nucleic acids are therefore not bend during the prediction process, may clash and

are therefore excluded from the result. $dsDNA_{10}$ demonstrates this very well: Two different clusters overlap, thus indicating a bent 15 bp long dsDNA would be able to bind into the pore, so it passes through it completely without clashes. Even longer (flexible) dsDNA may bind in this mode, but assemblies with lower energies are achieved by binding across the pore blocking the active sites. Negative stain electron micrographs and prior biochemical assays favour the binding mode through the pore. Here, it would potentially interact with the chloride site, the tension belt loop and α3. How the nucleic acid gets into the pore is so far unknown. A helicase/AAA protein like threading process seems unlikely as it requires extra energy. However, a simple mechanism would be the breakup of the tetramer at low chloride concentrations. Preliminary electron micrographs by Matthias Mörgelin demonstrated this with the W110P mutant for the very first time, but not for the wild type at low sodium chloride concentrations (section 2.2.3). Nevertheless, there is no strict evidence against the nucleic acid binding across the pore thereby blocking the active sites directly.

3 GɢᴘS – Sᴜᴍᴍᴀʀʏ, ᴄᴏɴᴄʟᴜꜱɪᴏɴ ᴀɴᴅ ʜʏᴘᴏ-ᴛʜᴇꜱᴇꜱ

3.1 Physiological implications of the chloride binding sites

Synechocystis sp. PCC6714 is closely related to PCC6803 (Figure 69), the organism of which the here tested GgpS originates from. In minimal medium with almost no salts, intracellular chloride concentration drops to 6.8 mM in PCC6714 cells (Reed et al., 1985). Hence, the minimal physiological chloride concentration coincides approximately with the minimal measured GgpS activity between 5 mM and 25 mM (Figure 28 and Figure 29). If the cyanobacterium is subsequently shocked with 500 mM sodium chloride (approximately the concentration present in sea water), the intracellular chloride concentration increases to around 200 mM, coinciding with the maximal activity of GgpS. *Synechocystis sp.* usually tolerates more than 1 M sodium chloride externally, hence, internal concentrations may increases further without the capacity to increases activity. However, the expression of GgpS is also triggered and probably more important in low salt adapted cells (Marin et al., 2002, 2006). Eventually, enough glucosylglycerol has been produced and the intracellular chloride concentration decreases. This inactivates all GgpS although the intracellular protein concentration remains high.

In vivo studies suggest almost complete inactivation of GgpS at low ionic strengths (M. Hagemann & Erdmann, 1994; Martin Hagemann et al., 1996; Marin et al., 2002). Chloride alone cannot inactivate GgpS as powerful, even if the one-chloride bound tetramer is indeed fully inactive (section 2.5.3). Assuming both chloride sites have the same k_d, then only one third off of all GgpS tetramers would be in the one-chloride bound state in presence of a chloride concentration equal to the k_d (according to model in section 8.4.2). Another third would be chloride-free and the second would have already bound two chlorides. Full inactivation is therefore, as proposed earlier, most likely achieved by DNA (poly-anion) binding. This reduces the activity of the chloride-free state

potentially to zero (section 2.6, Jens F Novak et al., 2011). As such, with a chloride concentration at the k_d, two thirds of all GgpS will now have a severely reduced activity. This fraction will increase, if the chloride concentration is further below the k_d and if the second chloride k_d is higher in general.

DNA inactivation was thought to be an unspecific effect: the ions shield DNA and GgpS electrostatically from each other without specific binding. In my assay, the DNA is effectively released by 200 mM sodium chloride and around 70 mM additional ionic strength (due to fixed assay ingredients). Roenneke et al. report that 100–150 mM sodium chloride are actually already enough (Jens F Novak et al., 2011). However, potassium alone accounts for 90 mM ionic strength (180 mM cations) in *Synechocystis sp.* PCC6714 adapted to a low salt medium (Reed et al., 1985). If the DNA is released by just small ions in the environment, i.e. by increasing ionic strength, then DNA would not be able to inhibit *in vivo* at all (Jens F Novak et al., 2011). The W110P mutant demonstrates, that binding of chloride or similar anions is essential for full DNA-inhibition recovery of GgpS, as it is still inhibited in an environment containing as much as 270 mM ionic strength in total. This means, the affinity of DNA to the two-chloride bound GgpS is lower than to the chloride-free (W110P) GgpS. And it explains why the natural potassium dominated cytosol does not interfere with DNA inhibition: Potassium ions cannot be incorporated into the chloride/anion binding site and they are not sufficient to release the DNA just due to ionic strength effects as sodium chloride is not able to do so with the W110P mutant.

Benjamin Roenneke analysed the GgpS from *Synechococcus sp.* PCC7002 and WH8102 briefly (B. M. E. Roenneke, 2014). He found that GgpS from PCC7002 is strongly induced by sodium chloride; that it interacts with DNA and is inhibited to some degree by DNA as well. The GgpS from stain WH8102 is different. It is neither induced by sodium chloride nor does it bind DNA. The sequence alignment with GgpS from PCC6803 reveals accordingly a substitution of Trp110 and Gln111 to threonine and histidine, respectively (Figure 69). Moreover, Lys308, Arg345 and Arg353,

which are important for the binding mode of DNA through the pore, are all substituted by glutamate. Hence, several of the important positions found in this thesis are substituted in the GgpS of WH8102 – a GgpS which is apparently also not regulated by the mechanism described in this dissertation. PCC7002 has 76 % sequence identity with 99 % coverage, while WH8102 has only 52 % identity with 96 % coverage, this indicated other potential changes.

Chloride regulates GgpS directly and modulates DNA inhibition. These findings have been missed by other researchers because it is hard to determine, and it is affected by unspecific ion effects. Moreover, Mg^{2+} was declared to be critical (M. Hagemann et al., 2001) and as of that, chloride was always present in all assay due to the addition of magnesium chloride. Therefore, with a K_d for chloride in the low millimolar range, the chloride-free state has also been missed.

3.2 A model explaining the effects of chloride and DNA on GgpS

Based on the presented results, a model for GgpS regulation by chloride and DNA can be derived (Figure 42). Important for this model is the single-chloride occupied state. I was able to estimate its activity roughly and the fluorescence quenching suggests stronger structural changes compared to the chloride-free and the two-chloride occupied state. However, further structural data supporting this are not available. This model is therefore an early working hypothesis and no final conclusion.

GgpS is tetrameric but the crystallographic symmetry and the presence of only two chloride sites suggests the separation into two mechanistically units. These are the dimers forming one chloride site.

According to PDBePISA, the tetramer is destabilised in the absence of chloride, and at least the W110P mutant was very recently found by negative stain electron microscopy to break apart into dimers, at least transiently (Figure 10). As demonstrated, the

W110P mutant is a good model for the chloride-free state (section 2.3.3). This tetramer is asymmetric because one dimer is considerably destabilised and becomes slightly less active (2.4). The crystal lattice probably selects a specific conformation, hence there are no alternate conformation obvious in the same subunit of the asymmetric unit in the W110P crystal. In solution, no such discriminating factor is present and both dimers probably switch conformations mutually. This apparent asymmetry is therefore potentially just flexibility. The hypothesis is supported by the assumed chloride-free structures of the crystal grown in sulphate, as the active site and the tension belt loop are always highly flexible.

The activity of the two-chloride/bromide occupied state was initially expected to be twice as high as of the anion-free state (2.4.1). This is indeed the case for the two-bromide but not the two-chloride-bound GgpS (section 2.5.3). The latter is probably only 20–40 % more active, depending on the acetate interference in the assay. This difference may be explained by an induced fit mechanism of the real substrates. W110P$_{Cl,SUC}$ structure does not have the large ADP moiety in its active site, thus missing several hydrogens bond. W110P$_{Cl,ADP}$ and may be W110P$_{Cl,ADP-glucose}$ crystals should be obtainable via soaking. The W110P mutant also demonstrated, that higher ion concentrations affect activity independent of the binding site. Chloride is 8 % smaller than bromide (Simoes et al., 2017), thus the latter might stabilise more and counter the adverse effects leading to a higher activity eventually (or in general). Another explanation for the strong effect of bromide, can partly be explained by different K_ds for the anion species. If the second K_d is for example lower for bromide than for chloride, the maximum would be reached sooner. However, preliminary interaction experiments cannot confirm different affinities yet (section 2.5.4).

Only a symmetric tetramer i.e. with both chloride sites *occupied or unoccupied* is (highly) active. The one-chloride bound i.e. asymmetrically occupied tetramer is on the other hand inhibited or potentially completely inactive. The activity assay only gives the

combined activity of all three states present at a given chloride concentration. The maximum fraction of GgpS tetramers in the one-chloride bound state is present at a concentration equal to the k_ds (assuming $k_{d_1} = k_{d_2}$). It would only be one third. In addition, the constant acetate background and the binding site unspecific ionic strength effect further influence the estimation of the activity of the asymmetrically occupied tetramer. Hence, it could be as inactive as suggested by the bromide data and the explored deconvolutions.

The interaction assay indicates real structural asymmetry of the one-chloride bound tetramer due to stronger quenching of the fluorophore (Figure 31). Occupation of both chloride sites might put the whole tetramer under tension and confers stability. However, if only one site is occupied, there is no countering force and the whole tetramer is distorted. α12 will potentially be involved as it is the dominant element in the interface between the two chloride dimers. The asymmetry could also mean, that the second binding event is impeded, as it has to work against the other chloride-stabilised dimer. However, the structural change can simultaneously facilitate the second binding event by pre-consti-tuting the binding site. Hence, the second K_d could be all: lower, equal, or higher. So far it is apparently higher but there is no significant difference due the preliminary nature of the interaction assay (Table 2).

A structural asymmetry has been observed in the model of the W110P mutant and the crystals of the wild type grown in presence of sulphate (Figure 12). I believe they are the chloride-free or are at least comparable to the chloride-free state. The observed structural asymmetry would go in hand with the initial assumption of reduced activity of this state. However, as mentioned before this could be just flexibility mitigated if educts occupy the active site. At least soaking with ADP should be possible.

As presented in section 2.7.5, inhibition by DNA may occur by either blocking the active site pockets or by binding in the pore close to the chloride site. Interestingly, the pore is bigger than the
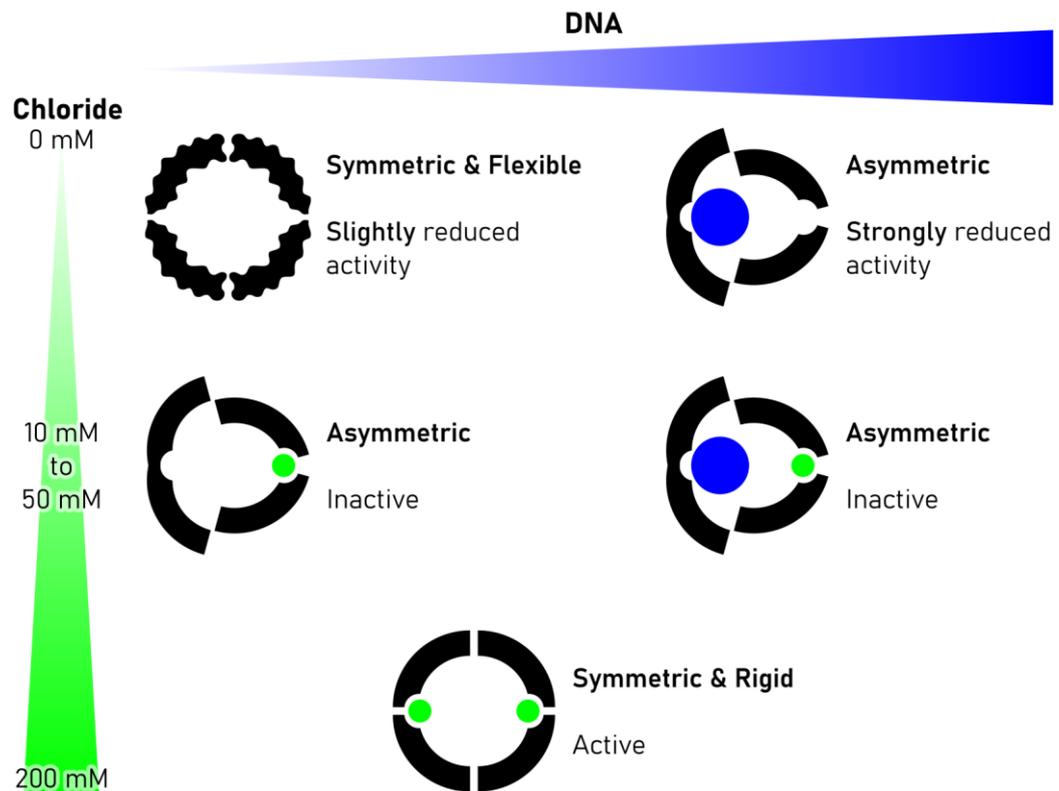
Figure 42: Hypothetical model of regulation of GgpS by chloride and DNA. The *apo* tetramer, i.e. chloride–free without DNA might be might be more pseudo–symmetric due to high flexibility.

DNA and neither can one dsDNA interact with both sites nor can two dsDNA helices occupy both sites simultaneously. Inhibition through the pore can be explained by the described model: The flexibility of the chloride-free tetramer allows DNA binding. This binding event could induce a similarly asymmetric tetramer as the one-chloride bound state. This may not inhibit completely as demonstrated in the assays (section 2.6), but subsequent binding of one chloride certainly does explaining previous results (Jens F Novak et al., 2011). If both sites become later occupied, the stability of both chloride-dimers strengthens and remodels the DNA-accommodating sites. This stops the interaction with the DNA in pore.

3.3 SUC1002 and SUC1003 – Allosteric modulation sites?

The sucrose molecule 1002 bound close to the chloride site implicates an allosteric inhibition site (Figure 8). Other compatible solutes (with sucrose not being tested) can inhibit GgpS (Jens Florian Novak, 2010). Interacting with the activating chloride

element would be an elegant approached to control glucosyl-glycerol synthesis under a sudden salt shock, if compatible solutes are already present. However, besides sucrose also G3P can apparently bind there, and sucrose-bound chloride-occupied models are equal to the ligand-free and ADP bound structures. The only exceptions are movements of the first β-sheet and helix α1 in/on the outer side of the tetramer, induced by sucrose binding in the active site. Thus, glucose-like compatible solutes probably inhibit simply by a competitive mechanism. There is also no evidence for allosteric activation. Moreover, as initial response, *Synechocystis* produces glucosylglycerol and sucrose, hence a sucrose inhibition would be counterproductive (Reed et al., 1985).

While the SUC1002 sites are in the middle of the pore, the SUC1003 sites are closer to the upper or lower end of the pore (Figure 7). Its interaction partners are on helices α11 and α12. α11 is of special interest since on top of its N-terminus is the active site loop around position 300. However, there is no hint of any effect.

The purpose of SUC1002 an SUC1003 site remain elusive. As discussed earlier, they are positioned similar to a nucleic acid double helix in the pore. If this binding mode is true, these sites could be able to interact with sucrose and similar molecules just because they are bulky and polar—a side-effect.

3.4 Required experimental evidences and improvements

3.4.1 Octamerisation

The last purification step of GgpS was a size exclusion chromatography in 20 mM HEPES/KOH and 150 mM NaCl. After dilution of the highly concentrated protein, the sodium chloride concentration was adjusted to 0.5 mM. If the chromatography was performed without sodium chloride, the peak usually shifted to a larger particle size approximately corresponding to an octamer (Figure 5). Preliminary experiments indicate an ionic strength dependence, but it was deprioritised, since the biphasic effect was unknown at the time and 0.5 mM NaCl was deemed as irrelevant. In addition, the activity assay was performed with at least 68 mM

ionic strength, which probably suppresses octamerisation as indicated by dynamic light scattering (Figure 10).

A considerable strong effect of the potential octamerisation on activity and fluorescence quenching is unlikely. If the activity is reduced by octamerisation, increasing chloride concentrations should activate, however, an inhibition is observed. If the activity is enhanced by octamerisation, the activity of the W110P mutant should decrease directly with increasing ionic strength, however, it does not. Assuming on top that the W110P mutant does not octamerise, then the W110P mutant at low ionic strength would have to be less active than the wild type but it is more active. Accordingly, fluorophores like Cy5 are usually quenched in presence of more protein (Marmé et al., 2003), hence converting octamers to tetramers would increase fluorescence but it decreases.

Nevertheless, increasing the ionic strength, perhaps by using 100 mM HEPES/KOH in purification and assay buffers, would rule out any effect thereby improving the assay accuracy. Moreover, all assays could be performed without any actively added chloride. *In vivo* the octamerisation probably does not occur owing to the presence of ions at 100 mM ionic strength (Reed et al., 1985).

3.4.2   Tetramer breakdown

The W110P mutant apparently forms dimers independent of ionic strength according to the very recent negative stain electron micrographs (Figure 10). This is the first time a tetramer breakdown has been observed and can explain how DNA can bind through the pore even if the breakdown is only of transient nature. As of now it is unclear which kind of dimer this is. An evaluation by PDBePISA judged the dimer around a chloride anion (A_2) as not stable, while the dimer found in the asymmetric unit (AB) is 'maybe' stable. The structural dynamics of GgpS have to be further researched by for example thermal shift assays, EM or cross linking experiments.

3.4.3 Structural determination of the single-chloride occupied state

As detailed in the hypothetical model, the one-chloride occupied state is key in understanding the mechanisms behind GgpS regulation. Therefore, it is highly recommended to determine its structure. Sulphate and acetate bind apparently just once and would therefore be perfect for crystallisation. However, grown crystals were of low quality. The 'sulphate crystals' diffracted anisotropically with resolution limits of 3.0 Å to 4.0 Å and datasets of the acetate crystals were not even recorded. The analysis revealed similar structures to the W110P mutant but the presence or absence of sulphate could not be confirmed. Crystallisation with chloride/bromide might not be successful because the K_ds of both sites are close to each other, if not equal. If the latter is true, only 50 % of all tetramers would be in the desired state at most, hence crystals must grow from a heterogenic population. Another approach would be a desalting of two-chloride crystals. While an easy experiment, a larger structural change is assumed and may destroy the crystals.

An applicable method for direct structural insights is cryo-electron microscopy. It is in theory capable of selecting the desired states from a heterogeneous population and grid preparation should not be extremely challenging because the protein is soluble and not membrane bound. This gives full control over sample properties like surface tension. Cryo- or just negative stain electron microscopy can be further utilised to understand the dimer/tetramer dynamics.

3.4.4 Activity

The assays were all performed in presence of 0.5 mM chloride. Therefore, activity and fluorescence of the chloride-free GgpS could deviate from the presented values. The difference would be the bigger, the lower the lowest K_d is. This highlights the importance of accurately determined K_ds. However, according to the interaction assay (Figure 31), saturation is usually almost achieved around 1.0 mM indicating that the 0.5 mM is not distorting strongly.

Mg^{2+} is reported to be critically for GgpS activity as demonstrated (M. Hagemann et al., 2001; Schoor et al., 1999). On the other hand, there is no requirement for this cation in the $S_N i$-like reaction of sterochemistry-retaining glucosyltransferases with GT-B fold like GgpS (S. S. Lee et al., 2011). As concluded in section 2.5.1, I am not convinced that Mg^{2+} is required or has a bigger significant activating effect. I have not conducted an experiment further supporting my claim since it seemed to be of no importance in the beginning and the used pyruvate kinase of the assay is Mg^{2+} dependent anyway (Larsen et al., 1997). However, the addition of magnesium also adds anions which seemingly all interact with the chloride site. Chloride and DNA inhibition can describe several observations which may have been believed to display the Mg^{2+} effect. Hence, I recommend to check the Mg^{2+} requirement of GgpS again. This should finally proof or disproof the hypothesis on the requirement of magnesium ions. However, total removal from the assay might not be possible because pyruvate kinase requires Mg^{2+}. Potassium for higher pyruvate kinase activity was added via HEPES/KOH titration. As noted above, more HEPES buffer should be used to rule out any octamerisation effect. This HEPES could be titrated with KOH and $Mg(OH)_2$, thus eliminating interfering anions.

GgpS was inhibited down to only 40 % by DNA in my assays (section 2.6). It is possible that only the one-chloride occupied state in presence of DNA is fully inactive. This is the state usually tested in previous publications (Jens F Novak et al., 2011). Hence, I recommend a chloride titration of GgpS in presence of DNA. This would also give a better understanding how the chloride site affects DNA binding/inhibition.

3.4.5   Interaction – Fluorescence quenching

A better range of salts than currently sampled must be measured. Lower bounds are currently imposed by the presence of 0.5 mM chloride as discussed above. However, lower concentrations are not necessarily required since saturation is almost achieved with the current set up (Figure 31). This is not the case

for the higher concentration regime. Microscale thermophoresis was first used to determine affinities and after protein denaturation, the signal started to correlate with salt concentrations above 100 mM. This indicates an MST response of the fluorophore itself. It did not occur with the fluorescence itself (Figure 77), hence, the assay was evaluated as reported. Nevertheless, higher salt concentrations, could interfere and might impose upper boundaries.

3.4.6 Further mutants – W110A and Q110A

Trp110 is interacting with the bound chloride anion but also via a π-interaction with Gln111 in the adjacent subunit (Figure 18). The importance of that interaction could be characterised by mutating the tryptophan or the glutamine to alanine. Both mutants might lose the chloride binding capacity as the π-interaction supports the binding site. However, backbone flexibility would be as in the wild type and thus any ionic strength effect more comparable between chloride binding and binding deficient GgpS.

FtsH

4 FTSH – INTRODUCTION

4.1 Protein degradation – A keystone to cell homeostasis

Of the four major biopolymers in living cells only proteins usually assume an active role and so are involved in or at least close to many chemical processes. These processes are not perfect and can directly damage or form by-products like harmful reactive oxygen species (Dasuri et al., 2013). These damaged and malfunctioning proteins have to be removed and replaced before they accumulate and distort normal cellular activities (Anisimova et al., 2018). Likewise, a cell adapts its proteome to new environments, conditions and the cell cycle by replacing old proteins with others of different function (Alber & Suter, 2019). Hence, proteogenesis and proteolysis are constantly renewing and remodelling the proteome of the cell. The former is executed by the ribosome machinery and its numerous constituents, interactors and regulators. Proteolysis on the other hand is more intricate. Besides complete degradation also, maturating and regulating cleavages are performed (Kühnle et al., 2019). Moreover, proteins can fold into rather tight three-dimensional knots or even big aggregates inaccessible for simple proteases. With AAA$^+$ proteases and lysosomes, two different independent locations and pathways for protein degradation have evolved.

Lysosomes are the degradative endpoints of endocytosis and autophagy but can also be used for storage as realised in the well-known plant vacuoles (Xu & Ren, 2015; Yim & Mizushima, 2020). Endocytosis describes the uptake of extracellular nutrients, smaller preys or pathogens into endosomes while autophagy is the packaging of proteins, other intracellular molecules or polymers (microautophagy) into lysosomes but also large protein aggregates or complete organelles (macroautophagy) into autophagosomes (Jackson & Hewitt, 2016). Autophagosomes subsequently fuse with endosomes and maturate into lysosomes if degradation is desired. Key properties of lysosomes are the high acidity of pH 4.1–4.2 and the concentration of acid dependent hydrolases degrading the

content (Ponsford et al., 2020; Schröder et al., 2010). Proteases localised here are predominantly cathepsins, small simple endo- or exo-serine, aspartate and cysteine proteases (Cocchiaro et al., 2017).

While degradation is well understood on lysosomes, protein unfolding and aggregate dissociation are not. The acidic environment of course changes the net charge and thus the solubility of proteins. This can increase but usually lowers solubility and facilitates aggregation and apart from the chaperones Hsc70 and Hsp90AA1 involved in chaperone-mediated autophagy there are no other chaperones localised in the lysosomal lumen (Santos et al., 2020; GO:0043202, 22.11.2020). Hsc70 and Hsp90AA1 are both critical for chaperone-mediated autophagy. Here no volume is engulfed by membranes but proteins are directly translocated through the lysosomal membrane (Tekirdag & Cuervo, 2018). First, the chaperone Hsc70 recognises proteins with an exposed KFERQ-motif due to damage or misfolding. This motif is found in about 40 % of all mammalian proteins. Hsc70 recruits other chaperones including Hsp90AA1 and hands the unfolded proteins over to LAMP2A which translocates them through the lysosomal membrane. There, they are subsequently received by lysosomal Hsc70 and eventually degraded. Hsc70 is also involved in microautophagy.

Lysosomes are part of the cytomembrane system and thus limited to eukaryotes with the usual exceptions (Lonhienne et al., 2010; Tekirdag & Cuervo, 2018). The chaperone mediated autophagy on the other hand is only found in mammals with some homologs also in *Drosophila* and *Danio* (Dacks & Field, 2007).

A much older degradation pathway are AAA[+] proteases. They evolved into different directions after the three kingdoms. Bacteria developed LonA, ClpXP, HslUV and the membrane bound FtsH. Archaea also have a derived, now membrane bound version of Lon (LonB) besides the PAN/20 S holoprotease. Last but not least, eukaryotes rely on the related but strongly further evolved proteasome (Sauer & Baker, 2011).

AAA$^+$ proteases contain an AAA$^+$ ring of six individual ATP consuming modules. Each module consists of two domains required for conversion of chemical to kinetic energy. Unfolding is achieved by translocation through the axial pore and in contrast to chaperones, substrates are not allowed to refold afterwards but are directed to a ring of six or seven protease modules with their active sites in the inner lumen. Thus, only translocated and unfolded proteins reach the proteolytic sites.

The simplest AAA$^+$ proteases are Lon and FtsH because they are hexamers of just one protein only. The others can consist of numerous proteins and have in addition one or two further N-terminal modules required for substrate specificity usually. While all AAA$^+$ proteases including bacterial LonA are cytoplasmic, archaea LonB and bacterial FtsH have N-terminal transmembrane and periplasmic domains/modules. The Clp family and HslUV have their homohexameric AAA$^+$ (ClpA, ClpC, ClpE, ClpX, HslU) and proteolytic active homoheptameric (ClpP, ClpQ) or homohexameric (HslV) rings on different proteins. The final complexes are constituted by two protease rings (ClpP and ClpQ do not mix) with one AAA$^+$ ring on one side. However, ClpAP and ClpCP have one AAA$^+$ on both sides of the proteases. The proteases of the PAN/20 S holoprotease form an octaicosamer of two different proteins in four rings ($\alpha_7\beta_7\beta_7\alpha_7$). A homohexameric PAN (AAA$^+$) ring completes this machine on both sides. Finally, there is the only AAA$^+$ protease in eukaryotes, the 26 S proteasome. It has a similar 20 S core particle but each α or β protein is a derived unique one. Also, the 19 S regulatory particle on either side is more complex. It consists of nine *lid* proteins as well as a heterohexameric AAA$^+$ ring und three further proteins in the *base*. One further protein interacts with both subcomplexes (Sauer & Baker, 2011; S. Zhang & Mao, 2020).

Proteolysis is highly regulated as only damaged, misfolded, aggregated or unnecessary but no functional and required proteins need to be degraded (J. Hanna et al., 2019). A well-known mechanism is the ubiquitin proteasome system (Nandi et al., 2006; Sokratous et al., 2014). A plethora of ubiquitinating enzymes and

other proteins recruiting them recognises a potential substrate, decides if it shall be degraded and eventually labels it covalently. The ubiquitin label is then recognised by the proteasome and removed before the substrate is unfolded and degraded. While covalent linkers are not used in bacteria, common adaptors like SspB and ClpS recognise substrates and hand them over to ClpXP (Bouchnak & van Wijk, 2019; Mahmoud & Chien, 2018; Sauer & Baker, 2011).

The initial substrate recognition is achieved by degrons and their exposition (Mahmoud & Chien, 2018; Sauer & Baker, 2011). A lysosomal degron is for example the aforementioned KFERQ-motif (Tekirdag & Cuervo, 2018). Stalled ribosomes are rescued via SmpB and *tm*RNA which add the SsrA tag to the stalling protein. It is the interaction site for SspB which directs the stalling protein efficiently to ClpXP. Nonetheless, all bacterial AAA$^+$ proteases can directly interact with the SsrA tag and degrade the protein subsequently (Fritze et al., 2020). A basic degron in all three kingdoms is the 'N-end rule' in which predominantly the first but also the second and the third amino acid with modifications (cleavage of N-Terminus, deamination, acetylation) are critical (Dougan et al., 2012). Less specific are degrons with just hydrophobic exposed residues, probably due to damage or misfolding (Bittner et al., 2017; Bouchnak & van Wijk, 2019).

4.2 FtsH – A unique AAA$^+$ protease

4.2.1   The only membrane bound AAA$^+$ protease in bacteria and eukaryotes

FtsH is the only membrane-bound AAA$^+$ protease in bacteria and faces the cytoplasm (Figure 43). Eukaryotes have often several FtsH homologs in their bacteria-derived organelles. Nine different FtsH homologs exist in *Aradidopsis thaliana* chloroplasts and three more are in their or other eukaryotic mitochondria (Nishimura et al., 2016; Opalińska & Jańska, 2018). Archaea evolved probably before FtsH with its two transmembrane helices and periplasmic part because they lack any homolog. They feature LonB as their membrane AAA$^+$ protease while the bacterial counterpart LonA is cytoplasmic only (Rotanova et al., 2004). Also eukaryotes do not
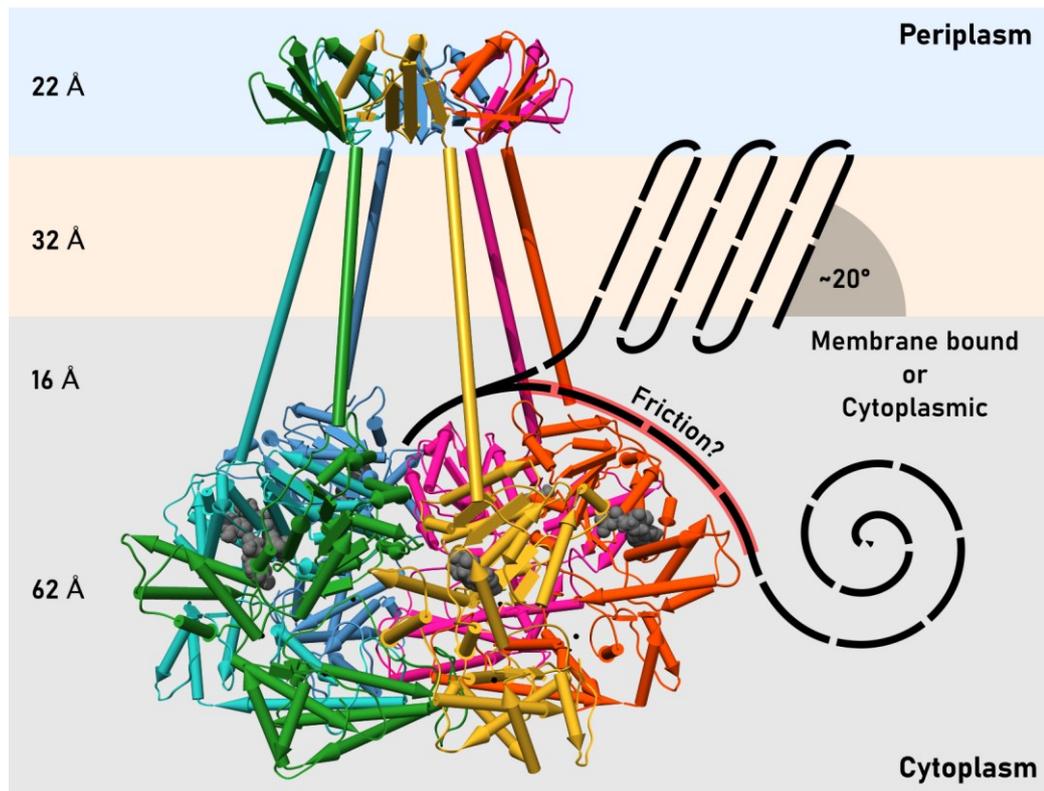
Figure 43: Collage of full–length FtsH. Depicted is a combination of the sixfold symmetric periplasmic domain of *E. coli* (4v0b, Scharfenberg et al., 2015) and the threefold symmetric cytoplasmic part from *Aquifex aeolicus* (6gcn, Uthoff & Baumann, 2018) one of my published structures.

The placement is based on the most recent cryo–EM model (Figure 67). Black dots in the cytoplasmic part are the zinc ions and the grey spheres are ADP. The second transmembrane helix and the linker between membrane and cytoplasmic part of each subunit is indicated by a long tube. Membrane and cytoplasmic proteins can be substrates. Transmembrane helices are usually inclined by 20° and will sit beside FtsH. Translocation of cytoplasmic substrates could be impaired by friction with the AAA ring or the membrane.

have FtsH or a homolog except from their organelles. They indeed have no dedicated membrane bound AAA⁺ at all.

FtsH homologs in mitochondria are found in the inner membrane and are divided in both topologies. i-AAA faces the inter-membrane space and is a homohexameric complex of Yme1 (yeast) or YME1L (Mammalia) with only one transmembrane helix instead of two (Figure 46, Weber et al., 1996). m-AAA faces the matrix with two transmembrane helices and is hence more FtsH-like. However, one version only existing in mammals, called AFG3L2, is truly homohexameric. An additional heterohexamer with additional paraplegin subunits in alternate constitution is present as well (Koppen et al., 2007). Only this second version exists in yeasts with

its subunits called AFG3 (Yta10) and Yta12 (Arlt et al., 1996). The mitochondrial versions are longer with about 130 additional amino acids at the N-terminus (Figure 46).

The plethora of chloroplast FtsH hexamers is not further discussed here. However, homo- and heterohexamers exists as in mitochondria but in different stoichiometric ratios and at least FtsH11 is believed to be present in both organelles. Nishimura published a great review about all proteases in chloroplast (Nishimura et al., 2016). As the great number in chloroplasts suggests, also cyanobacteria have a more complex FtsH system. *Synechocystis sp.* PCC6803 has four FtsHs which localise differently under low or high light conditions (Sacharz et al., 2015). This indicates additional homomers to the previously known 3:3 heteromers of FtsH1/FtsH3 and FtsH2/FtsH4 (Boehm et al., 2012). GgpS is a substrate of FtsH2 (Stirnberg et al., 2007).

4.2.2   A lack of unfoldase activity confers a role as simple waste collector?

As a membrane bound AAA$^+$ protease, FtsH degrades of course membrane proteins. However, also soluble cytoplasmic proteins are recognised and degraded (Bittner et al., 2017). Unlike other AAA$^+$ proteases, FtsH does not recognise a specific and unique degron or a universal adaptor/label like ubiquitin. However, several studies found that patches of 10–20 unstructured and often hydrophobic amino acids are required for degradation. The position of that rather unspecific degron was only sometimes of importance: N- and C-terminal as well as internal ones can be recognised and led to degradation (Chiba et al., 2000, 2002; Okuno et al., 2006). Interestingly, this very diffuse specification also agrees with the SsrA tag which is indeed recognised (C. Herman et al., 1998).

FtsH fails to degrade cytoplasmic globular and properly folded proteins and is therefore believed to prefer damaged and misfolded ones in general (Christophe Herman et al., 2003; Koodathingal et al., 2009). This is in agreement with the unspecific degron, which could indicate a misfolded protein. Among the first found specific substrates are the unmatched SecY subunit of the SecYEG

translocon and the unmatched $F_O a$ subunit of the F_O ATPase membrane complex (Bittner et al., 2017). Both are believed to be less stable if not complexed. FtsH also interacts with the protein-membrane insertase YidC and it potentially directs problematic proteins to FtsH for degradation (van Bloois et al., 2008).

Damaged but also native subunits of unassembled complexes are substrates in mitochondria and chloroplasts. The prominent ones are for example several subunits of the respiratory chain and the photosystem II (Janska et al., 2013; Opalińska & Jańska, 2018). In contrasts to protein degradation, FtsH and its homologs also control protein maturation and complex formation (Arlt et al., 1998; Marta et al., 2007). However, the latter can potentially be caused by failed maturation by the large mitochondrial ribosomal subunit MrpL32 by m-AAA (Bonn et al., 2011; Nolden et al., 2005; Woellhaf et al., 2014). Partial degradation by FtsH homologs is also involved in mitophagy, mitochondrial fission and intermembrane space import of cytochrome c oxidase (Anand et al., 2014; Tatsuta et al., 2007; K. Wang et al., 2013).

4.2.3 FtsH degrades healthy proteins involved in membrane stress response

A careful reader may has noticed that FtsH fails to degrade stable 'cytoplasmic' proteins. This sounds profound but also confusing considering that FtsH, as membrane protein, is probably more important for other membrane proteins. Recent studies used the integral membrane protein GlpG as substrate (Yang et al., 2018, 2019). They demonstrate full activity of GlpG in bicelles up to 85 °C, the whole range of the employed thermostability assay. Their solubilised protein substrate is therefore very stable. However, FtsH, also in bicelles, was able to degrade that protein. The authors concluded that FtsH in proper membranes is a powerful unfoldase if the substrate is another membrane protein. They hypothesise that the special topology of FtsH disfavours cytoplasmic proteins—the entrance pore faces the membrane or more precisely its own transmembrane domain. A substrate must either be besides that domain or the AAA ring if cytoplasmic (Figure 43). This leads to different angles of attack of the unfolding force.

FtsH potentially pulls directly parallel to the longitudinal axis of transmembrane helices, as they usually lay inclined by 20° in the membrane (Bowie, 1997). Cytoplasmic proteins however may interact with the side of the AAA ring during the unfolding process due to their position (Figure 43). This could lead to dissipation of the force by friction, yielding the weak observed unfoldase activity for soluble proteins. As such, FtsH is presumably able to process all membrane proteins if necessary and recognised. Some membrane stress regulators have been found which are degraded even if they are apparently in a native proper folded state (Figure 44).

FtsH is the only essential AAA[+] protease in *E. coli* because it controls the level of LpxC. This enzyme is important for lipopoly-saccharide (LPS) synthesis and shares the same educt with FabZ for production of phospholipids. The latter is obviously important for the membrane maintenance and production but also LPS are integrated into the outer membrane of gram-negative bacteria. They enhance the imposed barrier to the environment, for example against hydrophobic antibiotics (Delcour, 2009). Strict regulation of LPS synthesis is required as it either leads to LPS accumulation and depletes the phospholipid pool or leads to LPS unprotected membranes—both cases are lethal (Erwin, 2016; Ogura et al., 1999). LpxC has an FtsH 'specific' degron which is required for degradation but not sufficient (Führer et al., 2007). The degradation of LpxC is mediated by membrane proteins YciM (LapB) and YejM and correlates with levels of the alarmone (p)ppGpp (Schäkermann et al., 2013). The interactions between those components are not fully understood yet. The cytoplasm facing YciM co-purifies with FtsH and is required for LpxC degradation by FtsH (Klein et al., 2014; Mahalakshmi et al., 2014). It is probably a specific adaptor for LpxC and degradation by FtsH. The involvement of the periplasm facing YejM has only been recently discovered, it potentially senses somehow high LPS levels and inhibits YciM/FtsH dependent degradation (Guest et al., 2020).

An adaptor mediated degradation is also assumed for YfgM but none was found yet (Bittner et al., 2017). The N-terminus of YfgM is essential for processing, but 5 out of 14 residues are acidic, thus
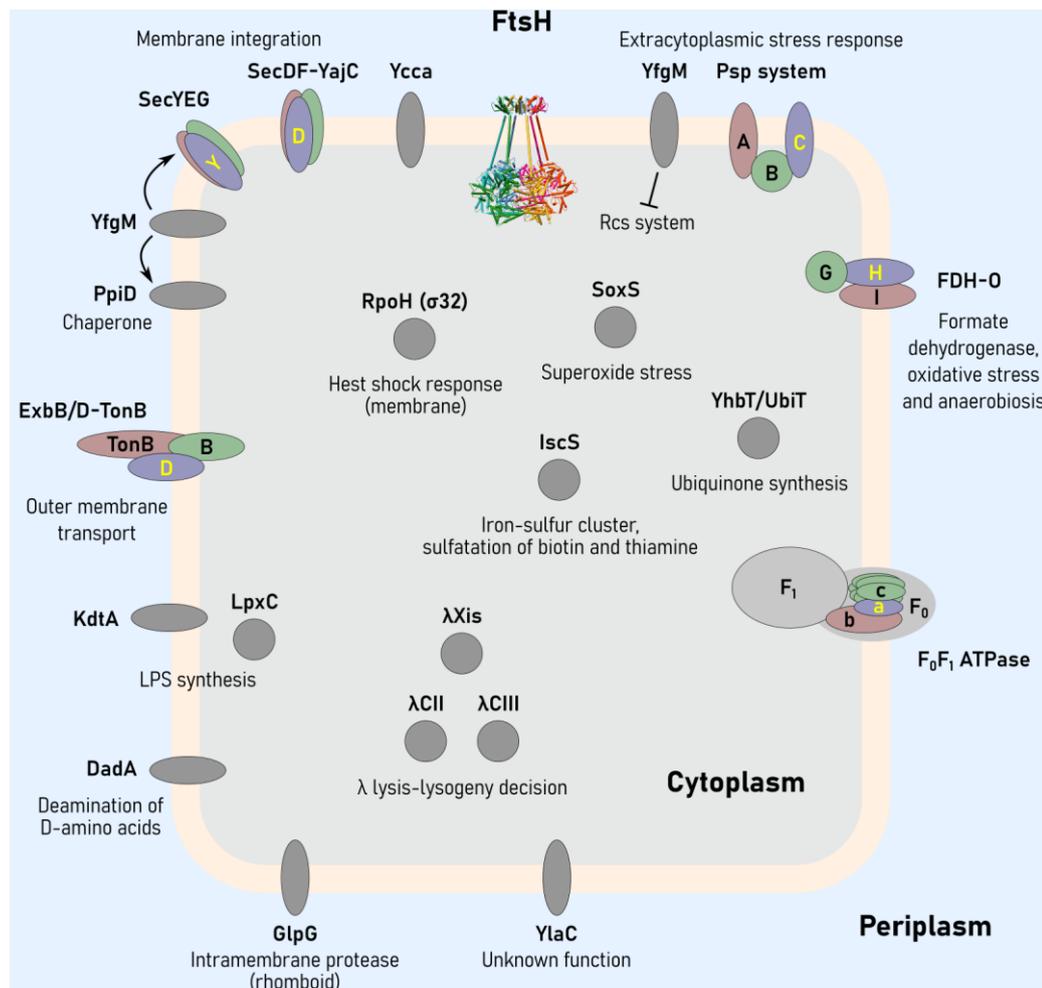
Figure 44: Regulative degradation by FtsH in bacteria.

Image based on Bittner et al., 2017, other publications linked in the main text and corresponding UniProt entries from *E. coli*. Only yellow/blue marked proteins in complexes are degraded, often only if solitary.

not following the hydrophobic rule established for FtsH. Fusing this degron to another membrane protein leads to its FtsH dependent degradation under similar conditions, thus clearly indicating its importance (Bittner et al., 2015). However, the direct trigger for YfgM degradation is unknown but cannot be that specific because of the mentioned fusion experiment with a different protein. YfgM itself is a chaperone or at least strongly related to chaperones in the membrane. It is an ancillary subunit of the SecYEG translocon and can interact with the membrane bound chaperone PpiD (Götzke et al., 2014, 2015). YfgM also suppresses signal transduction via the Rcs system (regulation of capsule synthesis, Wall et al., 2018). Here, RcsF senses damage to the outer membrane or the peptido-glycan and RcsB is eventually activated and induces expression of several anit-stress genes (Laloux & Collet, 2017). These genes are

important for cell division, salt tolerance, adaptation to the stationary phase and the biosynthesis of colanic acid. A building block for a cell encapsulating polymer (A. Hanna et al., 2003). YfgM is degraded by FtsH for example under osmotic shock and stress related to stationary growth (Bittner et al., 2015; Westphal et al., 2012). It thereby enhances Rcs response and most likely the chaperone activity in the membrane. Interestingly, YfgM is not degraded if uncomplexed (Bittner et al., 2015).

Further membrane and cell envelope controlling proteins were found to be FtsH substrates (Figure 44), amongst others: λCII and λCIII (λ phage), RpoH (σ32; heat shock), SoxS (superoxide stress), PspC (phage shock protein system) (Griffith et al., 2004; C Herman et al., 1997; Kobiler et al., 2002; Singh & Darwin, 2011; Yura & Nakahigashi, 1999). Bittner et al. compiled a moderately recent list of specific FtsH substrates (Bittner et al., 2017).

FtsH and its homologs are critical for membrane and organelle quality control. It definitely degrades damaged and misfolded proteins but has also important regulative purposes.

4.3 Membrane and periplasmic modules of FtsH

FtsH is anchored by two transmembrane helices to the membrane and they flank the periplasmic domain of roughly 70 amino acids (Figure 46). They are believed to be of regulative importance.

The proteins HflK and HflC are membrane bound with membrane protruding parts facing the periplasm. Both can form a heterodimer (HflK/C) and co-purify with FtsH (Yoshinori Akiyama et al., 1998; Bandyopadhyay et al., 2010; Kihara et al., 1996; Saikawa et al., 2004). They presumably interact via FtsH's periplasmic part as its deletion prevents co-purification. However, the authors also present simultaneously that FtsH does not oligomerise without the periplasmic part although it is still active. The HflC/K complex but also the individual proteins inhibit the degradation of SecY and λCII by FtsH *in vitro* (Bandyopadhyay et al., 2010; Kihara et al., 1995). Their overexpression *in vivo* leads

accordingly to the stabilisation of both substrates and their deletion to the destabilisation of SecY. However, λCII is again stabilised. The authors concluded that both proteins potentially compete about their degradation (Kihara et al., 1997).

HflK and HflC contain a prohibitin homology domain (PHB) and hence belong to the family of prohibitins/flotillins (Chiba et al., 2006). Another membrane bound PHB protein (QmcA) faces the cytoplasmic side and potentially interacts with FtsH as well. Eukaryotic prohibitins also inhibit m-AAA and are involved in numerous pathways of mitochondrial quality control such as oxidative phosphorylation, apoptosis, unfolded protein response, mitophagy or mitochondrial dynamics (Signorile et al., 2019).

The periplasmic part of FtsH and its homologs have been structurally characterised several times (An et al., 2017; S. Lee et al., 2011; Ramelot et al., 2013; Scharfenberg et al., 2015). Well before the cryo-EM 'resolution revolution', Lee et al. reported a cryo-EM model of the whole heterohexameric AFG3/Yta12 m-AAA at around 12 Å. The periplasmic modules form here a crown-like structure with the detergent micelle as base (Figure 45). The other publications report about the crystal structures of the periplasmic modules alone. A flat hexamer is formed after applying crystal symmetries and thus contradicting the crown-like appearance of Lee et al. (Figure 43). However, this could be a feature. An et al. reports, that ionic interactions in the interface are apparently pH dependent. They consequently propose a pH dependent switch from the flat hexamer to the dispersed crown-like shape. However, further decreasing the pH steadily increases the degree of oligomerisation and the high acidity (pI 4.9) of their *Thermotoga maritima* periplasmic domain is only found in closely related bacteria. Nevertheless, this mechanism might go in hand with the observation that FtsH activity is membrane potential dependent (Y. Akiyama, 2002). In contrast to the acidic residues, more conserved hydrophobic residues in the interfaces are proposed as potential unfolded protein receptor sites if in the crown-like shape. NMR structures of all other periplasmic domains are monomeric,
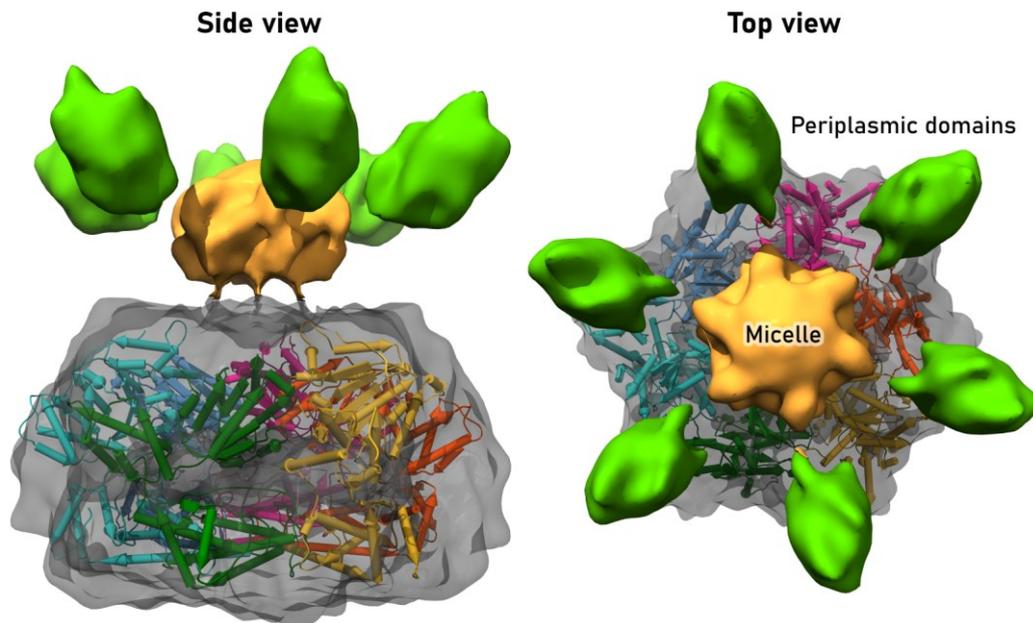
Figure 45: Cryo-EM model of the heterohexamer AFG3/Yta12 m-AAA (*Saccharomyces cerevisiae*) with crown-like periplasmic domain arrangement.

The C3 symmetric FtsH hexamer from *A. aeolicus* was fitted into the cytoplasmic part (6gcn, Uthoff & Baumann, 2018). The darker region corresponds to lower coulomb potential, i.e. the proteolytic chamber (EMD-1712, S. Lee et al., 2011). *S. cerevisiae* AFG3/Yta12 have 761 and 825 amino acids while FtsH has usually around 600–650 (Figure 46).

however, as the cytoplasmic part from *A. aeolicus* FtsH is monomeric in solution so may be the periplasmic as well (Vostrukhina et al., 2015).

Due to the topologies of HflK, HflC and FtsH, it is believed that they interact via the periplasmic part of FtsH. This means, a signal is somehow passed through the membrane to the cytoplasmic side.

4.4 AAA⁺ family substrate translocators

Most NTP binding proteins are P-loop NTPases which are divided into two divisions: <u>K</u>inases and <u>G</u>TPases are usually found in division KG, while the others including the AAA⁺ family belong to the ASCE (<u>a</u>dditional <u>s</u>trand, <u>c</u>atalytic <u>E</u>) division. Subfamilies share motifs and function and are often grouped to clades defined by common major motifs (Table 4). The following annotations are based on the publication by Millers and Enemarks. Moreover, most informations are compiled of their review correlated with others from Hansen and Whiteheart and Iyer et al. (Hanson & Whiteheart,

2005; Iyer et al., 2004; Miller & Enemark, 2016). The alignment in Figure 46 is annotated with all features discussed here.

4.4.1 Critical and conserved sequence motifs

The Walker motifs are essential for NTP binding and define the P-loop itself and hence the group of NTPases (Walker et al., 1982). The canonical sequence for the Walker A motif of the P-loop is GXXXXGK[T/S] in which X can be any amino acid. The P-loop links $\beta 1$ and $\alpha 1$ and is required for ATP binding. ATP hydrolysis is Mg^{2+} dependent. The cation neutralises the negatively charged phosphates and is coordinated by the β-phosphate and the Walker A threonine in addition to four water molecules. The sequence of the Walker B motif, characteristic for P-loop hydrolases as well, is hhhhD[D/E]. h can be any hydrophobic amino acid. It is found on $\beta 3$ and the ASCE division name is given by the glutamate of this motif that activates a water molecule for the nucleophilic attack on the γ-phosphate (Story & Steitz, 1992).

Prominent AAA+ family features are the sensor 1 at the beginning and the arginine finger at the end of the 'second region of homology' (SRH) between and including $\beta 4$ to $\beta 5$. However, the SRH is strictly speaking only conserved in the classic AAA subfamily including FtsH. Nevertheless, sensor 1 and the arginine finger are present in every AAA+ protein. The stronger similarity in the SRH of the AAA subfamily coincides with a missing sensor 2 (described below). Sensor 1 helps the Walker B motif to correctly order the water molecule that carries out the nucleophilic attack on the γ-phosphate (Story & Steitz, 1992). It is usually an asparagine (FtsH), threonine, serine or aspartate. The classification as 'sensor' comes from the fact that it senses or links ATP hydrolysis to other functions, like proteolysis in FtsH (Karata et al., 1999). The arginine finger is critical for AAA+ proteins because it is a *trans*-acting feature as in GTPases (Scheffzek et al., 1997). AAA+ proteins usually constitute a stable hexamer in which the arginine finger stimulates hydrolysis in the neighbouring subunit by interaction with the γ-phosphate (Ogura et al., 2004). Interestingly, the close N-terminal sensor 1 is *cis*-acting in contrast. The arginine

finger can sometimes and especially in FtsH be split in two arginines with similar importance. They are three positions apart on helix α4, thus pointing in similar directions (Karata et al., 1999).

A common element in all AAA$^+$ proteins is an additional α-helix at the N-terminus often called α0 (Smith et al., 2004). Directly in front of this helix is a conserved dipeptide called the N-linker, usually a small hydrophobic amino acid or glycine followed by glycine or a hydrophobic amino acid (FtsH: Ala-Gly). The first position interacts with the base of the nucleotide and the linker interacts in general with the helical domain. The region between the preceding module and the N-linker crosses the Rossman fold core and transmits signals in HslU (J. Wang et al., 2001).

An obvious difference of the AAA$^+$ family to other P-loop NTPases is the helical bundle at the C-terminus. It usually consists of four α-helices almost located 'on the other side of the nucleotide' and can be seen as an individual domain. It is therefore also referred to as lid (covering the active site pocket partially), small, or helical domain in contrast to the wedge, α-β, or big domain.

Another difference is sensor 2. It is located at the N-terminus of α7 in all AAA$^+$ subfamilies except from the classical AAA subfamily. It is either an arginine or lysine interacting with the γ-phosphate. Due to the loose bonding of both domains, ATP hydrolysis can change the interdomain angle most likely through sensor 2 (Ogura et al., 2004). However, sensor 2 may not be conserved in AAA proteins like FtsH, although this energy conversion certainly happens there as well. Sensor 2 is usually *cis*-acting, but especially clades with deviating helical domains can also have a *trans*-acting sensor 2.

Another third sensor is located at the C-terminus of α3. This arginine or histidine interacts with the water activating Walker B sidechain in a *trans*-acting manner in papillomavirus E1 and MCM (Enemark & Joshua-Tor, 2006; Li et al., 2015). No such sensor is known for FtsH but I hypothesise an aspartate at this position is important for pore-1 loop remodelling as published in our paper (section 5; Uthoff & Baumann, 2018).
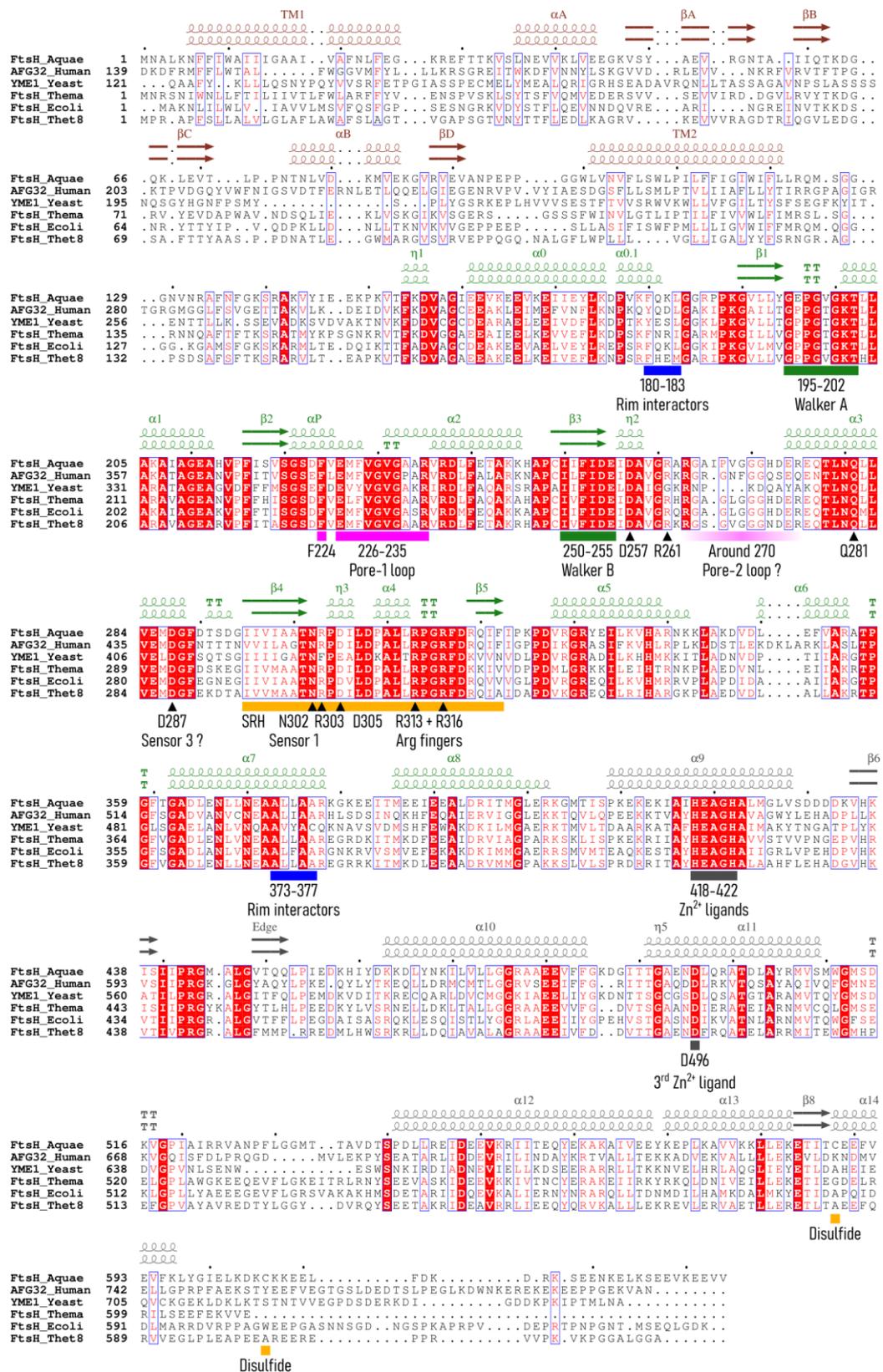
Figure 46: Alignment of FtsH with AFG3L2, Yme1 and other FtsHs.

The names correspond to the UniProt names. Secondary structures are based on chains A (top) and B (bottom) from 6gcn or 4v0b for the periplasmic module (Scharfenberg et al., 2015; Uthoff & Baumann, 2018). They are coloured according to module affiliation. Transmembrane and periplasmic: brown; AAA: green; M41 protease: grey. The transmembrane helices are annotated as given by UniProt.

4.4.2   AAA$^+$ proteins have diverse functions but all are polymer translocators

The overview presented in Table 4 lists the clades and the purpose of their assigned proteins. Usually, six AAA$^+$ modules of six identical or different proteins constitute a ring and translocate their substrates, proteins or nucleic acids, through the formed pore. Additional domains/modules/proteins with a variety of functions determine the final role of the particular AAA$^+$ protein. Hexamers with two rings in which two AAA$^+$ modules are on one chain do also exist. Translocation always leads to the removal of any 3D structure that may be present in the substrate. As such, a protein is unfolded, so that it could refold correctly (AAA$^+$ chaperones) or be degraded (AAA$^+$ proteases). A nucleic acid can be used as rail to either split the double helix continuously (helicases), to search for special sites (replication origin recognition), and/or to serve as a less mobile assembly platform for other proteins (replication clamp and helicase loader). While most nucleic acid AAA$^+$ proteins are perceived as moving along the strand, especially viral helicases move or pack DNA into viral capsids under very high pressure (Fuller et al., 2007).

Only one often mentioned AAA$^+$ member deviates form that general rule of thumb: dynein. Together with the non-AAA$^+$ kinesin, dynein is responsible for the transport along the microtubule. This sounds similar to the movement along DNA, but microtubules are much thicker and although dynein contains a hexameric ring, this complex does not translocate anything. The ring is indeed only a single protein with six AAA$^+$ modules and three appendages. A homodimer with several additional proteins including cargo adapters complete the microtubule walking machine. The movement requires only ATP hydrolysis in one of the AAA$^+$ modules in each protein while three others are completely inactive. The remaining two have probably just regulative purposes (Bhabha et al., 2016).

Table 4: Distinctive sequence features, functions and examples of all clades of AAA⁺ proteins conserved in all three kingdoms including virus-only helicases (Miller & Enemark, 2016).

Name	Changes	Function	Examples
Clamp loader	None	Loading of replication clamp	RFC, γ/δ of DNA polymerase III
Initiator	α-Helix insertion after α2	Replication origin recognition Helicase loading	DnaA/DnaC, Orc/Cdc6
Classic	No Sensor 2, additional α-helix after β2: 'Pore-1 helix'	Chaperone/protease	FtsH, proteasome, ClpABC, katanin, TIP49, AFG1, NSF/Cdc48/Pex
Superfamily III Helicases	ps1βh, no C-terminal AAA+ lid, additional helical bundle, Sensor 2 is trans-acting	Viral DNA helicase	papillomavirus E1 SV40 large T-antigen helicase, adeno-associated virus Rep40
HCLR	Pre-sensor 1 β-hairpin (ps1βh)	Chaperone/protease Helicase	HslU/ClpX, CTD of ClpABC, Lon, RuvB
Helix-2-Insert	ps1βh, β-hairpin after in α2	σ54 RNA polymerase activator endonuclease	NtrC, PspF, McrB
Pre-Sensor 2 Insert	ps1βh, β-hairpin after in α2, additional α-helix after α5, Sensor 2 is trans-acting	Chaperone Motility	MCM, MoxR, YifB, dynein

4.4.3   The hand-over-hand translocation mechanism

The actual translocation mechanism of the usual AAA⁺ protein has been studied for a long time. Random and concerted mechanisms with and without symmetry have been proposed a lot (Sauer & Baker, 2011). The advent of cryo-electron microscopy (cryo-EM) unifies this highly debated field and clarifies differences. 3D models of FtsH-like proteases such as Yme1 and AFG3L2 as well as the 26S proteasome (Dong et al., 2019; Puchades et al., 2017, 2019), of chaperones like Hsp100 and Hsp104 (Deville et al., 2017; S. Lee et al., 2019), of the ESCRT-III polymer disassembler Vps4 (Monroe et al., 2017) and also of a helicase have been published (Gao et al., 2019).

Translocation occurs in general by ATP hydrolysis in one subunit and its subsequent repositioning further downstream of the substrate polymer. This is followed by hydrolysis in and movement of the next subunit like a hand-over-hand rope pulling mechanism but with six hands (Figure 47). An active AAA⁺ protein is thus fully asymmetric with its AAA⁺ ring assuming a so-called spiral staircase: Each subunit interacts with the second next amino
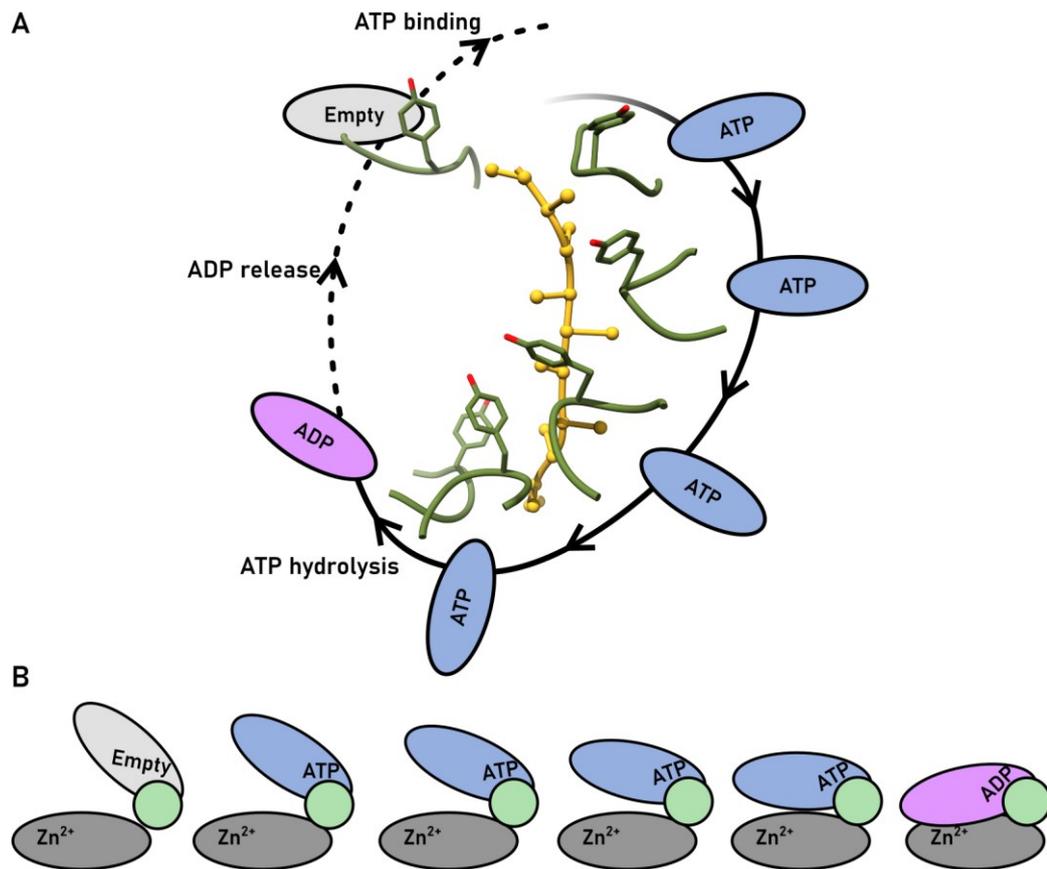
Figure 47: The hand-over-hand translocation mode of AAA+ proteins.

A) The spiral staircase formed by pore-1 loop in Yme1 (EMD-7023). B) The nucleotide induced movement is predominantly a change of the angle between the AAA and the protease module in FtsH-like proteases. The subunits are ordered by most open to most closed. Based on Puchades et al., 2017.

acid or nucleotide in the substrate and is therefore angled differently relative to the plane of the AAA+ ring. This simple mechanism has for example been observed in Yme1 and is so far unchallenged for FtsH-like AAA proteases (Puchades et al., 2017). ATP hydrolysis is only in the bottom, most closed subunit possible, as the *trans*-stimulating arginine finger of the previous subunit in a more open conformation is required. This mechanism also ensures the direction of the staircase. After hydrolysis, this subunit changes into the open conformation and thereby stimulates ATP hydrolysis in the next subunit. Subsequently, new ATP occupies the subunit forcing a closing movement which cannot be achieved in one step but requires all preceding subunits to undergo hydrolysis first.

The substrate is engaged by loops protruding into the axial pore. Pore-1 loop is between β2 and α2 and contains a highly

conserved motif ([FY][IV]G) in all AAA$^+$ proteases (Figure 46, Sauer & Baker, 2011). Especially the aromatic residue of that motif has been linked to substrate translocation for a while (Martin et al., 2008; Park et al., 2005; Schlieker et al., 2004; Yamada-Inagawa et al., 2003; F. Zhang et al., 2009). Interestingly, a defining feature of the AAA clade is an additional α-helix directly preceding this motif (Miller & Enemark, 2016). Also, the pore-2 loop between β3 and α3 can interact with the substrates but its sequence is not as good conserved and harbours several glycines (Figure 46). The recent cryo-EM models demonstrate how interaction occurs (Puchades et al., 2017, 2019). The pore loops assume a similar spiral like the subunits they are belonging to. The aromatic amino acids, with side- and mainchain from four ATP loaded subunits intercalate between or are at least positioned above every second sidechain of the substrate (Figure 47A). The ATP binding induced downwards movement of each subunit is therefore directly transferred onto the substrate. The substrate has to be in a linear form, slightly helical, thus each downwards pushing translocation step unfolds the substrate and translocates by approximately 6 Å.

Under suitable reaction conditions, *E. coli* FtsH can degrade the rather stable membrane bound GlpG while translocating 0.5–1.7 amino acids per ATP (Yang et al., 2018). The authors compiled an average of 0.2–5 amino acids per ATP for the other AAA$^+$ proteases as comparison. According to the cryo-EM models of Yme1 and AFG3L2, we can assume a theoretical maximum efficiency of two residues per ATP due to mechanical reasons (Puchades et al., 2017, 2019). Another study further characterises FtsH (Yang et al., 2019). The maximum hydrolysis rate is achieved at low 60–80 μM ATP, owing to a cooperative mechanism not found in other AAA$^+$ proteases. The authors compare this to the ClpXP which reaches its highest ATP hydrolysis rate in a linear concentration dependent fashion at 500 μM ATP. FtsH has to make around 20 hydrolysis events per minute and hexamer for a successful degradation, i.e. unfolding while 80 or 100 are necessary for peak efficiency unfolding GplG or casein.

4.4.4   Critique of the simple hand-over-hand mode

Studies of the chaperone ClpX do not agree directly with that simple model, yet spiral and pore loops exist nonetheless (Fei et al., 2020). The authors propose a single power stroke from the most open to the most closed state, thus explaining numerous reports about 10–20 Å translocation steps. They also hypothesise a less strict regulation of ATP hydrolysis, as such, other subunits in the middle of the spiral could build up stress leading to a burst of several translocation steps in quick succession. This moves the substrate 20–40 Å at once. For unfolding, this could mean to completely remove a β-strand from a β-sheet instead of partial removal and a subsequent refolding (Martin et al., 2008; Sen et al., 2013). It could also potentiate the applied force in order to unfold tightly bound secondary or tertiary structures and might explain why different AAA+ proteases are better unfoldases than others (Koodathingal et al., 2009).

The basic model also fails to explain how hexamers with a few hydrolysis deficient subunits are still able to translocate as they should stall further processing (Beckwith et al., 2013; Cordova et al., 2014). However, they could be rescued by mechanisms similar to ClpX. Eventually, more complex AAA+ proteins like *hetero*-hexameric AFG3L2 are composed of two distinct AAA+ proteins and the proteasomal AAA+ ring even of six different ones. Experiments demonstrate accordingly a difference between these proteins of the yeast 26 S proteasome (Beckwith et al., 2013). AAA+ hexamers with hydrolysis deficient mutants of Rpt3, Rpt4 and Rpt6 cannot translocate substrates anymore while mutating the other three subunits only reduces degradative activity of the proteasome by 40–60 %.

The recent cryo-EM analysis of the 26 S proteasome reveals additional complexity on top (Dong et al., 2019). The authors link differently nucleotide occupied AAA+ rings and different conformational compositions with three key stages of substrate processing. The first is ubiquitin binding and deubiquitination followed by translocation initiation and ubiquitin release.

Eventually, there is the third phase in which they associate the typical 'spiral staircase'-like arrangement with the degradative translocation.

AAA$^+$ proteins/proteases evolved to ever more complex machines. However, the prototypic translocator made of six identical subunits constitutes a spiral rather than a ring and translocation depends on subsequent ATP hydrolysis stimulated by the previous subunit in the ring. However, deviation from that plan most likely tunes the AAA$^+$ machinery to its actual task: Steady translocation (Gao et al., 2019; Puchades et al., 2017) or removal of tight secondary and tertiary structures (Fei et al., 2020); a second AAA$^+$ hexamer (Deville et al., 2017) for disaggregation; pumping of viral DNA in capsids under high pressure (Fuller et al., 2007); loading of protein complexes on the DNA (Kelch et al., 2012) or the movement along microtubules (Bhabha et al., 2016).

4.5 FtsH is a zinc metalloprotease

As similar as the AAA$^+$ modules might be across all AAA$^+$ proteases, as divergent are the actual proteases. HslV, the proteasome and its archaea counterpart utilise N-terminal threonines as peptide bond nucleophilic attacker; ClpP and Lon are characterised as serine proteases with a His-Asp-Ser triade and Lys-Ser dyad, respectively, and FtsH itself is the M41 family of zinc metalloproteases (Sauer & Baker, 2011).

FtsH has the metalloprotease typical conserved HEXXH motif in which XX are alanine and glycine. The histidines coordinate the Zn^{2+} cation and the glutamate deprotonates the catalytic water bound to the ion. The motif is found in the very first α-helix of the protease, α9 if extending the canonical AAA$^+$ annotation. The third zinc ligand, an aspartate, is positioned right after an angled π-helix at the N-terminus of α11 (Figure 46, Bieniossek et al., 2006).

There are in total six α-helices and three very short β-strands in one sheet. This β-sheet helps aligning the substrate above the active site but the so-called edge strand, interacting with the substrate, is only transiently part of the sheet. Together with the

conserved preceding three positions it can also form an α-helix or be completely disordered in the crystal structures (Bieniossek et al., 2006, 2009; Vostrukhina et al., 2015). This is different in the cryo-EM models of Yme1 and AFG3L2 in which the edge stands are part of the β-sheet in all conformations (Puchades et al., 2017, 2019). In addition, substrates are found at the edge strands in the four AMP-PNP loaded subunits of AFG3L2. Hence, substrate release might be coupled to the AAA$^+$ cycle but on a closer inspection the question arises how the substrates exit the hexamer at all. The crystal structures have secondary pores between the subunits close to the protease active sites, however, they have often also a disordered i.e. missing edge strand region. These secondary pores are sealed in the cryo-EM models by the edge strand, the C-terminus and the AAA$^+$ modules due to a rotation of the AAA$^+$ against the protease ring (as further outlined in our paper, section 5.1.2).

The simple translocation mechanism is based on steps of two residues per power stroke and subunit. All subunits are proteolytic active, hence, the produced peptide should probably have lengths of two amino acids as well. However, based on the cleavage pattern of pepsin and α-casein, 85 % of all products are smaller than 3 kDa which are roughly 27 amino acids (assuming 110 kDa/amino acid, Asahara et al., 2000). The authors admit that their statistic based on mass spectrometry data might not be representative but it indicates that proteolysis is not compulsorily following a translocation step. The found cleavage sites are predominantly in hydrophobic regions i.e. the substrate can perhaps not be aligned correctly and thus is not cleaved.

4.6 The FtsH project – The goals

Our group is characterising FtsH for several years. (Bieniossek et al., 2006, 2009; Uthoff & Baumann, 2018; Vostrukhina et al., 2015). When I started my master's thesis, the structure of the truncated, i.e. non-membrane bound version of *Aquifex aeolicus*, a hyperthermophile bacterium, by Marina Vostrukhina was about to be published. However, they had some reservation about that

structure: Its C2 symmetry was potentially a crystallographic artefact. My goal was to improve that structure and I eventually succeeded with a slightly different construct (FtsH$_{Aa,tr}$). Based on two similar crystal forms (R32 at 3.41 Å and P312 strongly twinned at 3.10 Å) I made early predictions of the translocation mechanism. However, resolutions were low and the models themselves of poor quality.

4.6.1   Refine and publish the results and predictions of the master's thesis

The initial tasks for my PhD thesis were therefore clear: Improve these structures and refine the predictions. I was indeed able to get a better diffracting crystal (2.95 Å) and refined the structure to publication quality. Together with a more in depth analysis, this work was published as my first first-authorship paper and is included as the following chapter (Uthoff & Baumann, 2018). A key result is the determination of a second more compact pore loop conformation. It probably does not interact with the substrate anymore and is required during the upwards movement of the ADP loaded subunit without clashing into the substrate or other subunits.

4.6.2   Establish new activity assays and validate predictions

Besides further structural work, I wanted to validate my predictions of the master's thesis and the paper. Some of the key amino acids (Figure 46) found are well-known elements like the arginine fingers (Arg313 and Arg316), or elements of the Walker motifs (E255) (Karata et al., 1999; Miller & Enemark, 2016). Others including D287, R303 and D305 have been analysed before and no loss in proteolysis has been observed upon mutation of these residues (Karata et al., 2001). However, their FtsH from *E. coli* was truncated, thus lacking the membrane and periplasmic part. More-over, their sequence numbering does not agree with deposited sequence in UniProt (P0AAI3). We assumed D287 to be critical for remodelling of the pore loop and coincidentally, it matches the description of the position of sensor 3 in MCM or the papilloma-virus E1 (Enemark & Joshua-Tor, 2006; Li et al., 2015). Of great importance are of course the pore-1 loop phenylalanines F224 and

F228, of which the latter is again well known. Moreover, Q281 is like D287 probably important for pore loop remodelling.

Simple activity assays had been established before but were usually of lower throughput because reactions had to be stopped at the chosen time points and be further processed. In addition, the ATP hydrolysis assay was always perceived as unreliable. As an AAA$^+$ protease, FtsH has two main activities, ATP hydrolysis and proteolysis, but also less obvious ones: translocation and unfolding. A goal for all assays was to establish easy to handle 384-well plate based medium to high-throughput assays so several mutants can be analysed in rapid succession.

Analysis of the proteolytic activity is usually carried out by degradation of a protein substrate, however, this requires at least the successful translocation and unfolding. A pure proteolytic assay could use short peptides with fluorophores at each terminus. A signal is only emitted after cleavage when both groups are separated. If the assay works as intended, the peptides are small enough to diffuse into the proteolytic chamber without ATP consumption. This assay was theorised but not developed since the focus of the project was not on proteolysis itself.

Translocation can be measured by the usage of an unfolded protein. Tested were chromophore (master's thesis) and fluorophore labelled caseins. The former requires the removal of uncleaved but still coloured proteins and effectively means the addition of trichloraceate for precipitation and stopping of the assay. The latter method allows a readout continuously because the labelled casein is internally quenched and only cleavage releases fluorescently active peptides. In any case, the proteolytic activity is used for report. A strict proteolysis assay (as above) can ensure correct interpretation but no FtsH mutant should interfere with proteolysis.

Testing the unfolding capability requires a folded protein. Casein however has no tight tertiary structure and is often used as 'universal' protease substrate because of that. This third type of assay was originally sought after only as an alternative to the

translocation assay but is in fact different because of unfolding as demonstrated in this thesis. A continuous assay can utilise GFP, a well-folded protein which loses its fluorescence due to unfolding (and degradation). GFP has been used for ClpXP (Nager et al., 2011), but FtsH is a weak unfoldase for soluble proteins, hence the very stable GFP cannot be processed (Christophe Herman et al., 2003). However, circular permutation of GFP yielded a good substrate for Lon a similarly weak unfoldase and thus may be processable by FtsH as well (Wohlever et al., 2013). Circular permutation reorders the termini of a protein. GFP is an eleven stranded β-barrel and the termini are of course in front of the first and behind the eleventh β-strand. However, in $_{cp6}$GFP they are behind β6 and in front of β7. As Wohlever et al. demonstrated, this destabilises the β-barrel enough so that Lon can process it. Further destabilising modifications are known (Perez-Jimenez et al., 2006; Wohlever et al., 2013).

AAA$^+$ protease activities are usually measured in presence of an ATP regeneration system using creatine phosphate and creatine phosphokinase (Nager et al., 2011; Wohlever et al., 2013). During my master's thesis, I could show that the regeneration system enhances activity and that FtsH is probably ADP inhibited. Moreover, with 32 mM creatine phosphate, FtsH has up to 40 mM ATP available, increasing the assay time.

Last but not least, ATP hydrolysis was determined by a derivative of the GgpS activity assay. This means ADP production was coupled to NADH oxidation via pyruvate kinase and lactate dehydrogenase. Interestingly, pyruvate kinase essentially regenerates ATP. Hence, the regeneration system could be changed to pyruvate kinase and enables, so far in theory, the simultaneous readout of translocation/unfolding and hydrolysis.

4.6.3 Determine the full-length structure of FtsH via cryo-electron microscopy

The spiral staircase was back in 2015/2016 not more than a faint hypothesis and a full-length high resolution structure of FtsH i.e. with membrane and periplasmic part was not available—as it is

not by now. I tried to crystallise it briefly during my master's thesis but we are not equipped properly for membrane protein crystallisation. Hence, we started a collaboration with Elmar Behrmann, at the time independent group leader at Caesar, Bonn. The goal was and is still to resolve the full-length structure of *Aquifex aeolicus* FtsH (hereafter FtsH$_{Aa,FL}$) at 'atomic resolution' via cryo-electron microscopy (cryo-EM). *A. aeolicus* is a hyperthermophilic bacterium living at up to 95 °C (Deckert et al., 1998). Proteins of such bacteria are more stable and believed to be better crystallisable (Deller et al., 2016). This hopefully holds true for cryo-EM.

4.6.4 Rethink FtsH solubilisation or reconstitution

Cryo-EM requires the protein frozen in thin ice films on a holey carbon grid. The formation of these films is heavily influenced by the presence of surface-active agents. They can either help or prevent successful preparations. FtsH has always been solubilised with n-dodecyl-β-ᴅ-maltopyranoside (DDM), a micelle forming detergent. This restricts freezing options because micelles are only stable if the detergent concentration is high enough, thus heavily influencing surface tension. They have an additional drawback: A micelle has usually a relatively small radius, the formed sphere has therefore a very high curvature. It can distort the structure of protein. On the other hand, the small size helps to solubilise only the protein or tightly bound complexes. Another problem with detergents is the usual lack of natural lipids in the micelle and hence the incapability to analyse lipid:protein interactions. Different alternatives were therefore tested.

DDM is an uncharged maltose head group with an unbranched aliphatic tail often used for membrane protein preparation (Bloch et al., 2020). CHAPS is a detergent as DDM, but zwitterionic in nature and successfully used by our collaboration partners for other proteins. Styrene-maleic acid copolymers (SMAs) solubilise proteins and preserve the natural lipid environment. They are stable after formation and do not required constant supply of

surface active molecules (Dörr et al., 2016; S. C. Lee & Pollock, 2016).

The last class of solubilisation agents are composed of protein-lipid discs. These so-called nanodiscs have been used for some time (Bayburt & Sligar, 2010; Denisov & Sligar, 2017; Hagn et al., 2013). They provide a flat natural-like lipid environment and the process is therefore usually referred to as reconstitution. The hydrophobic side of the discs is covered by two long helical amphipathic helices derived from apolipoprotein A1. These membrane scaffolding proteins (MSPs) shield the hydrophobic lipid surfaces from the solution. The membrane patch sizes depends on the length of the used MSPs and the correct size is critical. If the nanodisc is too small, reconstitution will fail, and if it is too big, the target protein might move in the discs leading to heterogeneous particles abolishing single particle cryo-EM. However, a new invention can help (Frauenfeld et al., 2016). Saposin A lipoprotein discs are also stable flat membrane complexes, but the disc surrounding MSPs are replaced by small rather globular proteins. The disc diameter only depends on the actual concentrations of saposin A, lipids and the target protein. Successful reconstitution is therefore only a matter of titration, in theory.

5 FTSH – PUBLISHED ARTICLE

Secondary structure numbering in the paper was based on the secondary structures found in the models. The numbering was change to better fit the general AAA$^+$ annotation. The article under the title '*Conformational flexibility of pore loop-1 gives insights into substrate translocation by the AAA+ protease FtsH*' was published in the Journal of Structural Biology (Uthoff & Baumann, 2018).

5.1 Results

5.1.1 Two new trigonal crystal forms of soluble FtsH$_{Aa,tr}$

The previously reported crystals of *A. aeolicus* FtsH were grown using an N-terminally truncated (142–634) quadruple mutant (I250M, F360L, K552R, E627G). The mutations were unintentionally introduced by PCR (Vostrukhina et al., 2015). In order to avoid a possible bias by these unintentional mutations and circumvent the assumed crystal artefacts, the crystals described here were grown using an even further N- and C-terminally truncated construct (only positions 151–608) without those mutations. The C-terminus of *A. aeolicus* FtsH has an unusual length compared with other species, and only *E. coli* has a similarly long but not conserved C-terminus. It is reported to be cleaved autocatalytically (Yoshinori Akiyama, 1999) but its purpose is unknown. Similar to the quadruple mutant, the new construct is in solution monomeric and thus virtually inactive in ATPase and protease assays.

The truncation led to easily reproducible crystals in space group R32 in presence of 60 % Tacsimate at neutral pH without removal of the N-terminal His-tag. Twinned crystals in space group P312 appeared only upon partial removal of the His-tag. Both crystal forms contain fully ADP-bound FtsH hexamers. Hereafter, the crystals and models are named after their space group.

Data reduction was carried out with XDS (Kabsch, 2010) and molecular replacement with Phaser (Adams et al., 2010), using each individual chain or each domain from the quadruple mutant as search model for the P312 crystal form, and later on the P312 model for the dataset in space group R32. The here presented model
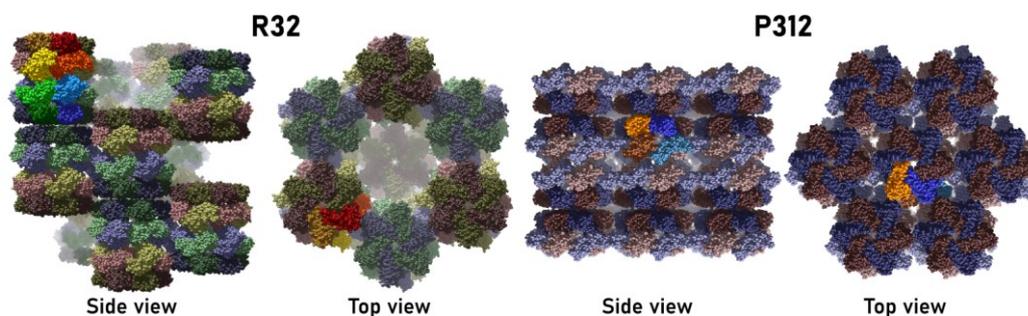
Figure 48: Packing analysis of the two crystals with space groups R32 and P312.

Highlighted (more saturated colours) are the asymmetric units with four (R32) or two (P312) monomers, respectively. The darker colours represent the protease disks, the brighter ones the AAA⁺ rings.

in space group R32 is based on two merged datasets. The subsequent refinement and building cycles were done in Phenix (Adams et al., 2010). Refine (Adams et al., 2010) and Coot (Emsley et al., 2010). All data collection and refinement statistics are listed in Table 13.

The R32 model comprises four monomers in the asymmetric unit and was refined to a R_{work}/R_{free} of 0.216/0.257 at a resolution of 2.9 Å. The four monomers belong to two independent hexamers that are generated by the crystallographic threefold symmetry (Figure 48). The two hexamers stack on top of each other with their AAA⁺ rings (head-to-head interaction). Two of these dodecamers form an assembly of a 24mer generated by the crystallographic twofold axis and therefore stack via their protease rings (back-to-back interaction). The 24mer assembly is repeated in the c-direction after a gap of some 70 Å. Lateral stacks are shifted up- or downwards by 1/3 or 2/3 of the height of a stack along the c axes, respectively. This leads to different crystal contacts between each monomer and its surrounding symmetry mates (Figure 49).

The P312 model is in many aspects very similar to the R32 model. It was refined to a R_{work}/R_{free} of 0.235/0.264 at a resolution of 3.3 Å. However, the crystals were twinned with a twin operator of -h,-k,l and a twin fraction of approximately 0.43. With only two molecules in the asymmetric unit, just one hexamer is generated by the threefold crystallographic symmetry (Figure 48). These hexamers stack infinitely by head-to-head and back-to-back interactions along the c axis and are arranged in layers in the plane
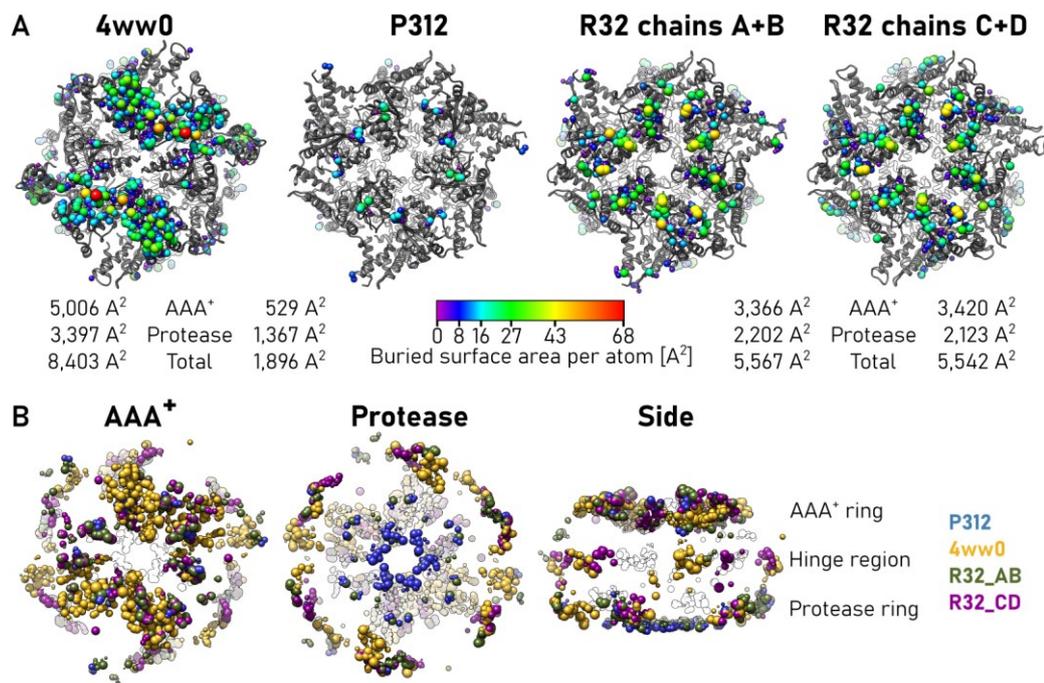
Figure 49: Analysis of buried surface of the two new structures in the crystals with space groups R32 and P312.

The hexamers and their environment truncated to 14 Å around the hexamer were first calculated in Pymol and then parsed to areaIMol from the CCP4 package (B. Lee & Richards, 1971; Schrödinger, 2010; Winn et al., 2011). A) Grey: Ribbon representation of the corresponding hexamer; Rainbow coloured spheres: Atoms, which are buried by crystal contacts, colour and radius are proportional to third root of the buried surface area. Owing to low electron density, residues 230–234 were not built in the open conformation of the P312 model. B) All buried atoms from all four hexamers superimposed in different perspectives. Crystal contacts differ in general in the number of atoms on top of the AAA^+ ring, their distribution and along the helical domains of the AAA^+ modules near the hinge regions. In 4ww0 the contacts are stronger and shifted to the 'left' (side view).

of the *a* and *b* axes. Therefore, crystal contacts differ between both trigonal crystals.

We have already earlier observed this formation of dodecamers (head-to-head as well as back-to-back) in the crystal structures of the quadruple mutant and of FtsH from *T. maritima* but concluded that they have no biological significance. In order to compare our structures with the previously published *A. aeolicus* quadruple mutant, we calculated the surface of the hexamers that is buried by crystal lattice contacts using areaIMol from the CCP4 package (Lee and Richards, 1971, Figure 49). Our two new trigonal crystal structures show different crystal contacts and have at least 40 % less buried surface area compared to the previous orthorhombic

crystal form of the quadruple mutant, indicating lesser packing constraints and thus representing more likely the biologically important picture.

5.1.2   The angle between AAA$^+$ and protease module classifies the open and the closed conformation

The protease ring of each hexamer in both space groups adopts an almost perfect sixfold symmetry, as observed in all other FtsH structures. The AAA$^+$ ring, however, adopts threefold symmetry and the major difference between the individual monomers in the asymmetric unit is the intermodule angle in the hinge region between AAA$^+$ and protease modules (Figure 50B). Two classes of conformations can be defined, an open conformation with a larger angle between both modules and a closed conformation with a smaller one. As listed in Table 5 in the supplemental material, the intermodule angle between open and closed conformation ranges between 19° and 23°. All rigid body transformations exhibit a high closure, meaning with almost no twisting component (calculated using DynDom Server, Hayward and Berendsen, 1998).

We also analysed the conformations in the twofold symmetric, but also fully ADP bound *A. aeolicus* and *T. maritima* structures (Bieniossek et al., 2006; Vostrukhina et al., 2015). The results are also listed in Table 5 in the supplemental material. Both structures have two subunits with similar intermodule angles as in our new crystal forms; however, there are also subunits with an even smaller angle. In total, the angle varies by 35° across all structures in all space groups, and all the models have in common that adjacent subunits always adopt different conformations. An additional translational component further differentiates the previous *A. aeolicus* quadruple mutant structure from both new ones. We concluded that stronger and differently positioned crystal contacts around the AAA$^+$ rings led to the distorted symmetry.

A short comparison with the sixfold symmetric apo-structure of *T. maritima* revealed each of its monomers adopting an open-like conformation, however, a rotation by around 30° of the AAA$^+$ and protease rings against each other has occurred additionally.
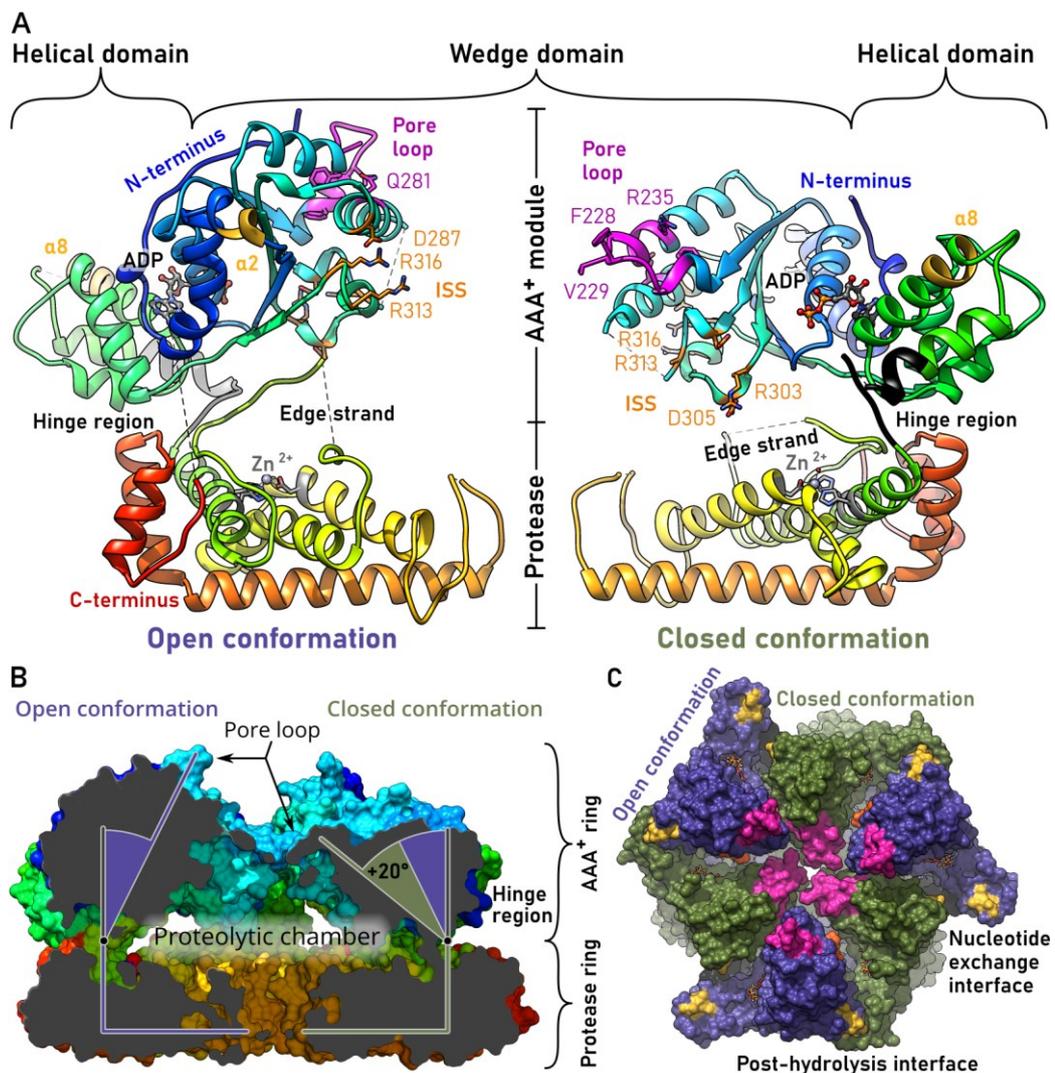
Figure 50: Annotated monomer and hexamer of the crystallised FtsH$_{Aa,tr}$ construct (residues 151–608).

A) Monomer with highlighted features and rainbow colouring (blue to red, chain A from R32). The edge β strand that binds the substrate is disordered. B) R32-Hexamer in clipped side view (coloured as the monomer). C) Top view on the R32-hexamer coloured according to conformational state (closed: green, open: blue) and features coloured as in (A). The pore loops of the subunits in closed conformations are very close to each other with their pore loops sealing the pore.

Interestingly, this is also visible in the cryo-electron microscopy structure of YME1. In general, if compared with the conformers of YME1, our closed and open subunits resemble the very closed ADP-bound and the apo or the most open ATP conformation, respectively. Whether a ring rotation is completely artifical or an additional feature remains open. Similar ring rotations are already known with lower extent (7°) from HslUV, but also of similar size (20°) from p97 (J. Wang et al., 2001; Yeung et al., 2014). p97 is a

Table 5: DynDom Server Analysis (Hayward & Berendsen, 1998).

The letter after the underscore represents the analysed chains. Closed conformations are present in chains A and C, open conformations in chains B and D in the new structures in space groups R32 and P312. Residue numbering is according to FtsH from A. aeolicus. The hinge regions is the intermodule linker (all values are based on this hinge), further hinge regions are at/around the active site switch and a small part of the protease module at the C-terminus where A. aeolicus FtsH has a stabilizing disulphide bond (C588–C606).

Conformers	Rotation Angle	Trans-lation	Closure	Hinge Regions			
				Inter-module	Active Site Switch		C-terminus
R32_A : R32_B	21.3°	-0.8 Å	97.1 %	395–407	438–439	441–442	
R32_C : R32_D	19.3°	-0.6 Å	99.3 %	393–394	439–442	460–461	
R32_A : R32_C	No dynamics were found						
R32_B : R32_D	No dynamics were found						
R32_A : P312_A	6.7°	0.0 Å	93.0 %	391–407			
R32_B : P312_A	15.4°	-0.6 Å	87.9 %	394–407	438–439	445–462	
R32_A : P312_B	28.3°	0.2 Å	98.9 %	395–407			
P312_A : P312_B	22.5°	-0.1 Å	94.3 %	394–413	438–440	442–461	
R32_A : 4WW0_A	19.5°	-0.8 Å	94.0 %	393–407	438–439	442–459	
R32_A : 4WW0_B	27.0°	-6.5 Å	78.8 %	396–407			
R32_A : 4WW0_C	No dynamics were found						
R32_B : 4WW0_A	No dynamics were found						
R32_B : 4WW0_B	34.3°	-0.5 Å	76.7 %	405–406			
R32_B : 4WW0_C	30.0°	-1.6 Å	97.9 %	405–408		438–461	
R32_A : 2CEA_A	No dynamics were found						
R32_A : 2CEA_B	13.2°	-0.3 Å	78.2 %	394–407			584–585
R32_A : 2CEA_C	27.9°	-2.1 Å	64.2 %	394–406			
R32_A : 2CEA_D	No dynamics were found						
R32_A : 2CEA_E	No dynamics were found						
R32_A : 2CEA_F	25.2°	-1.7 Å	42.2 %	406–407			
R32_B : 2CEA_A	No dynamics were found						
R32_B : 2CEA_B	17.4°	-1.6 Å	99.5 %	391–392		440–459	584–585, 589–595
R32_B : 2CEA_C	46.6°	0.3 Å	85.1 %	409–409		442–458	
R32_B : 2CEA_D	20.0°	-1.3 Å	99.1 %	391–392			584–585, 589–590

cytoplasmic AAA protein from yeast and shares about 40 % identity with the AAA$^+$ module from FtsH.

5.1.3 The intersubunit signalling network differs between post-hydrolysis and nucleotide exchange interfaces

The threefold symmetric FtsH hexamer with alternating open and closed conformations of the monomers has two types of interfaces between adjacent AAA$^+$ modules. We name them 'post-

hydrolysis interface' and 'nucleotide exchange interface' according to our model of ATP hydrolysis; with the post-hydrolysis interface presenting the ADP-bound state after ATP-hydrolysis and ejection of the inorganic phosphate ion, and the exchange interface presenting the state before ADP is leaving and being replaced by ATP (Figure 50C, Figure 51). Both interfaces are mainly characterised by the two α-helices α0.1 and α7 (Phe180–Leu183 and Ala373–Ala377, respectively), as well as by pairs of interacting arginine and aspartate residues (Arg261, 303, 316, and Asp257, 287, 305; Figure 51). Some of the latter have been previously implied to take part in intersubunit communication (Augustin et al., 2009). An annotated alignment of the FtsH proteins from the species discussed here can be found in the supplementary material (Figure 46).

In the following, the subscripts *Op* (open) and *Cl* (closed) will indicate the conformation to which the residue belongs. Figure 51 depicts the interfaces and Figure 52 the same scene with corresponding electron density.

In the post-hydrolysis interface, the intrasubunit ion pairs of $Arg316_{Op}:Asp287_{Op}$, $Arg261_{Cl}:Asp257_{Cl}$, $Arg303_{Cl}:Asp305_{Cl}$ are separated by around 20 Å and the main interaction between both AAA^+ modules is mediated by the two α-helices: $α0.1_{Op}$ and $α7_{Cl}$ at the outer rim of the ring (Figure 51). This interaction is mainly hydrophobic in nature: $Phe180_{Op}$, $Leu183_{Op}$ as well as $Ala373_{Cl}$, $Leu374_{Cl}$, $Ala377_{Cl}$, in addition the hydrogen bonds between $Lys179_{Op}$ and the peptide bond oxygen of $Arg378_{Cl}$ (at the top of $α7_{Cl}$), as well as the salt bridge between the sidechains of $Arg378_{Cl}$ and $Glu172_{Op}$. The only other interaction between both AAA^+ modules that is not at the outer rim of the ring is an ionic bond between $Arg313_{Op}$ and $Glu255_{Cl}$, the latter being part of the Walker B motif. The position of the sidechain of $Arg313_{Op}$ is not well defined, but at the currently modelled position the guanidinium group is around 13 Å apart from the centre of the β-phosphate of ADP_{Cl}. At the same time, $Arg316_{Op}$ is better resolved and with 15 Å in similar distance to the β-phosphate of ADP_{Cl}. We assume this arrangement of Walker B $Glu255_{Cl}$, arginine finger $Arg313_{Op}$ and
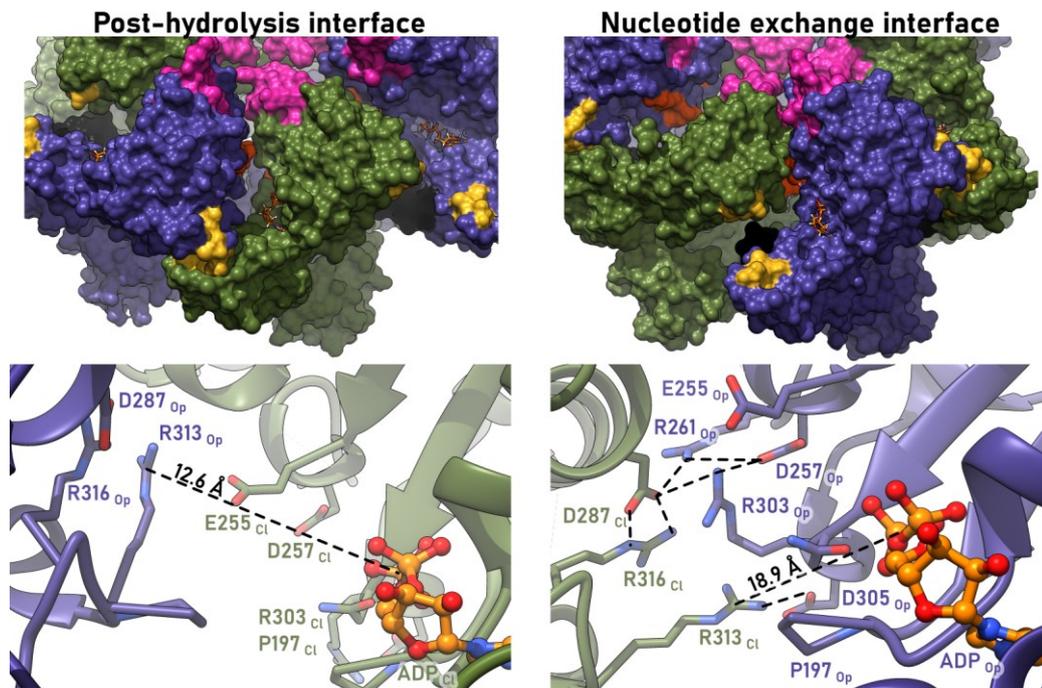
Figure 51: The post-hydrolysis and nucleotide exchange interfaces in the threefold symmetric FtsH hexamer (R32).

Top row: Overview of the two interfaces. Blue: open conformation, green: closed conformation, orange: intersubunit signalling network magenta: Phe180—Leu183 (α0.1) and Ala373—Ala377 (α7), black: hinge region, orange sticks: ADP. Bottom row: Detailed view of the intersubunit signalling network. Colour as above. Nucleotide exchange interface: Pairs of aspartates and arginines (Asp257$_{Cl}$/Arg261$_{Cl}$, Asp287$_{Op}$/R316$_{Op}$, Asp305$_{Cl}$/Arg303$_{Cl}$) interact with each other. There is only an interaction between arginine finger Arg313$_{Op}$ and Walker B Glu255$_{Cl}$ in the vicinity of the ADP phosphate groups. Post hydrolysis interface: The aforementioned ionic pairs take part in an extensive intersubunit interaction network, including Arg313.

Arg316$_{Op}$ and nucleotide after hydrolysis of the ATP, hence the name post-hydrolysis interface.

In the nucleotide exchange interface, on the other hand, the guanidinium group of Arg313$_{Cl}$ is about 19 Å away from ADP$_{Op}$ and in addition, Pro197$_{Op}$, and Asn302$_{Op}$ block the passage between both. The electron density is again low around this Arg313$_{Cl}$, but the final map shows weak electron density indicating interactions with Asp305$_{Op}$ (sensor 1, Figure 52). Arg316$_{Cl}$ is 17 Å away from the β-phosphate of ADP$_{Op}$ and is now forming a salt bridge with Asp287$_{Cl}$. This aspartate is further interacting with Arg261$_{Op}$ and Arg303$_{Op}$, which are both interacting with Asp257$_{Op}$. This hydrogen bond network is located within the inner part of the AAA$^+$ ring and replaces the α0.1$_{Op}$–α7$_{Cl}$ interaction framework established in the
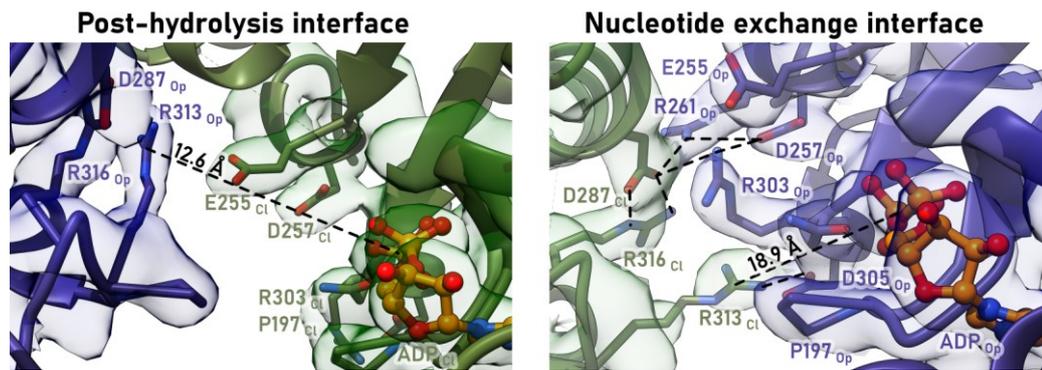
Figure 52: The post–hydrolysis and nucleotide exchange interfaces in the threefold symmetric FtsH hexamer (R32) with feature enhanced map at 1 σ RMSD.

outer part of the post hydrolysis interface. The helices are now around 14 Å apart from each other and thereby opening the outer part of the cleft between both monomers increasing the solvent accessibility of the bound nucleotide, hence the name: nucleotide exchange interface.

5.1.4 Conformational changes critically alter the position of Phe228 in the pore loop

The so-called pore-1 loop with the highly conserved FVG motif is located between the helices αP (AAA insertion into AAA$^+$ fold) and α2 (Figure 53A) and is responsible for substrate binding and translocation. Our new structures resolve for the first time two different conformations of this loop and show critical differences linked to substrate translocation and release. However, the loop of the open conformer is more flexible in the crystal of space group P3₁2, thus the loop was only partially built and lacks the residues 230–234.

In the closed conformation, Phe228 points into the pore and seals it, together with Val229 and both of their symmetry equivalents (Figure 50). The C-terminal helix α2 starts at Gly230 and Arg235 interacts with the peptide bond carbonyl oxygen of Asp223. Helix αP comprises the amino acids between Ser220 and Phe224. The described conformation of the pore loop is expected if the pore motif interacts with a substrate and is already known from previous structures (Bieniossek et al., 2006) and visible in both of our structures.
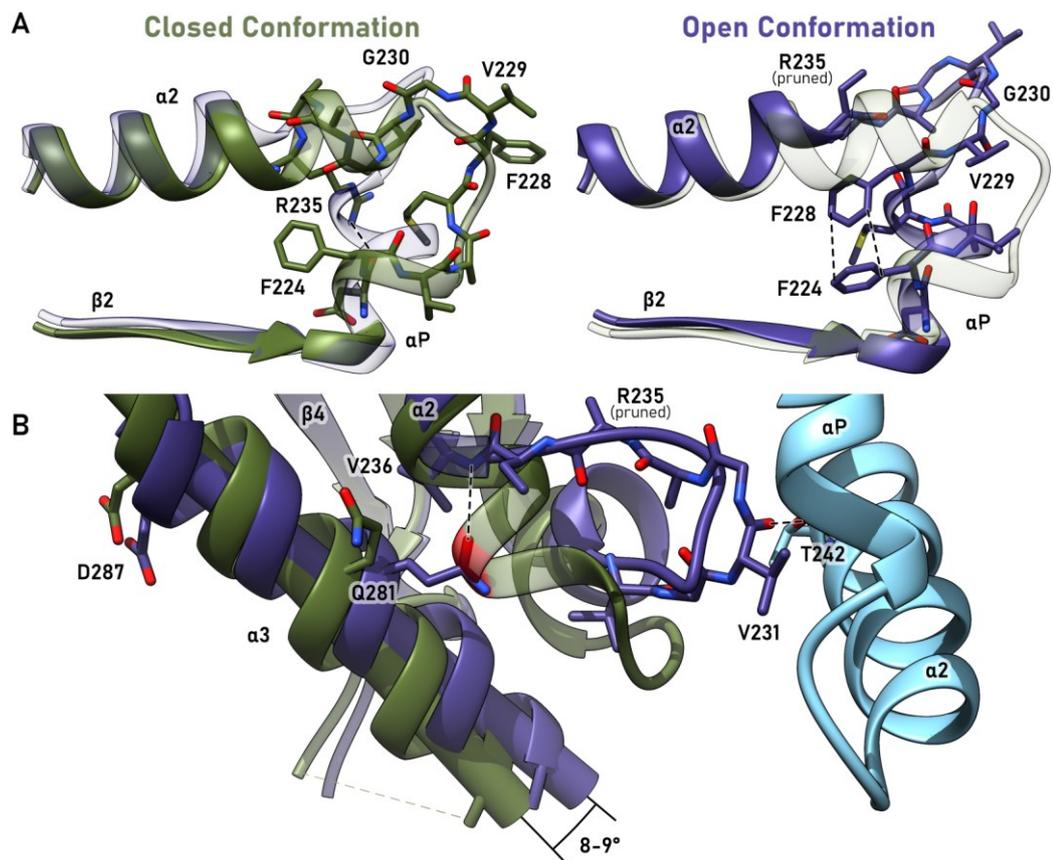
Figure 53: The different conformations of pore–1 loop in the crystals of space group R32.

The substrate interacting pore motif consist of the three amino acids [228]FVG[230] and lies between helices αP and α2. A) The pore loop adopts a compact fold in the open conformation (blue), Phe228 undergoes a π–π interaction with Phe224 and αP is extended by one turn. At the same time, α2 is shortened similarly and Arg235 saturates the carbonyl oxygen of Phe224 which is in the current conformation no longer part of a helix. B) Conformational changes go in hand with a shift of helix α3 by 8°–9°. Helix α3 with Asp287 is part of the ISS and its Gln281 has displaced the N–terminal end of helix α2 in the open conformation. The compact pore loop form is stabilised via a polar crystal contact (Val231– Thr242). Sky blue: Pore loop of an NCS-related closed conformer.

On the other hand, in the open conformation, major changes can be observed (Figure 53A): αP is extended by residues Phe224 to Val229 in α-helical conformation, while helix α2 is simultaneously shortened and starting now with Arg235 instead of Gly230. More-over, the sidechain of Arg235 is no longer interacting with the carbonyl oxygen of Asp223, since it now takes part in the newly formed α-helical part. In the current model of space group R32, the sidechain is not properly resolved but weak unexplained density suggests a position centrally over the pore loop, where it would be able to interact with the peptide bond oxygens of Gly230 and Ala233 of the loop. Thus, Arg235 acts maybe as kind of a switch by

saturating the carbonyl oxygens of Asp223 or Ala233 if they are not part of an α-helix. This part (residues 230–234) of the polypeptide chain is disordered in space group P312.

Aside of Arg235, the most important change in this conformation is the newly formed π–π interaction between Phe228 and Phe224, a sidechain conserved across all FtsH orthologues. Phe228 can engage in this interaction because it is part of the newly formed helix and the phenyl ring is now pointing away from the pore. This way the pore motif adopts a more compact form compared to the more extended one in the closed conformation.

These changes are most likely induced by disrupting the intersubunit signalling network (ISS) with the adjacent closed subunit during the upwards movement while changing to the ADP bound opened state (Figure 51, post hydrolysis interface). Asp287 is moved by 2.1 Å, leading to a shift of helix α3 by around 8.5° (Figure 53B). Subsequently, Gln281 is positioned further towards helix α2 thus displacing its N-terminal end and inducing most likely the compact pore loop conformation.

A polar crystal contact (Val231 with sidechain of Thr242 of another hexamer) stabilises the compact pore loop conformation. Nevertheless, our structure links the ISS with the conformational change in the pore loop required for the upwards movement of the current subunit. Otherwise, the pore loop would push the substrate out of the pore again.

5.2 Conclusion

The general finding of this study is supported by experiments from Augustin et al. (2009) on the FtsH homologue m-AAA in yeast. This protease is a heterohexamer of two FtsH-like paralogs most likely arranged in an alternate sequence. Analysing a model based on published FtsH structures (Bieniossek et al., 2006; Suno et al., 2006), they reported several residues as important in intersubunit signalling and they could prove that hydrolysis in one subunit depends on the adjacent one.

Our structures give a more detailed view on these residues in an overall structural perspective. Arg313/316 (Arg447/450 in AFG3, Arg435/438 in YME1) are either stimulating the ATP hydrolysis or take part together with Walker B Asp257 in the ISS. This enables ATP hydrolysis only in closed subunits with a neighbouring (more) opened one. This is in general in agreement with the recently published Yme1 model (Puchades et al., 2017). There, owing to just 5°–10° difference of the intermodule angle, the adjacent ATP loaded subunits show primarily an interaction of the phenylalanines 342, 344, 378 and 411. They correspond to residues 216, a not conserved one, 252 and 289 in *A. aeolicus* FtsH. In Yme1, these phenylalanine sidechains do not interact in the ADP-bound subunits, as they do not in our FtsH$_{Aa,tr}$ structures. As described, Asp287 (Asp409 in YME1) on α3 bound to Arg261/303 of the adjacent subunit is important in our model for pore loop conformation and thus substrate interaction. Puchades et al. (2017) identify the interaction of the phenylalanine quartet (with F290 C-terminal of α3 in *A. aeolicus*) as crucial but do not report or discuss any conformational change of the pore loop. They only highlight that depending on the intermodule angle the pore loop is further away or close enough for substrate interaction.

We believe that the inactive or compact pore loop should also be induced in Yme1 if the phenylalanine quartet does not interact with each other. An analysis of the deposited map and coordinates reveals, that one of the final five deposited structures, all based on the same map, show an inward folded pore-tyrosine (Phe228 in FtsH) in the corresponding subunit. However, the map is poorly resolved in this area and the significance of the coordinates is questionable (Figure 54).

A clear difference between Yme1 and all other here discussed proteins are the highly conserved arginines 261 and 303. They seem to be crucial in our *A. aeolicus* structures and are present or conservatively substituted in virtually all of the other homologues but have no counterpart in Yme1. Arg303 is universally present in e.g. the FtsH-like yeast m-AAA protease subunits AFG3 and YTA12 as well as the human homologue paraplegin (SPG7) and the
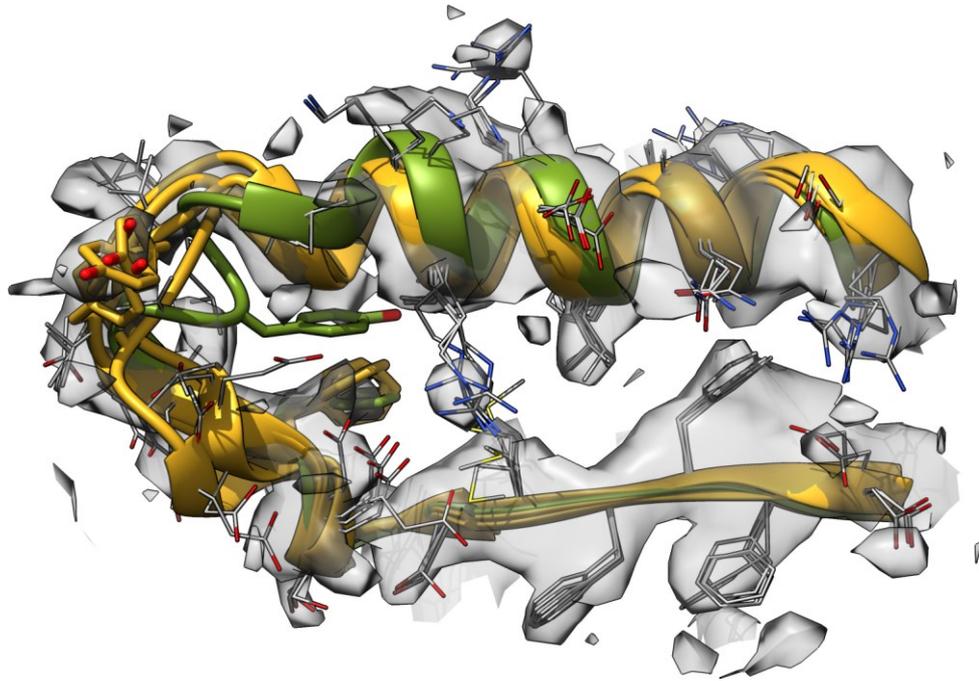
Figure 54: YME1 pore loop of the open conformation (EMDB: 7023).

All five models from the ensemble refinement are shown. The green one has a similar fold to our compact/inactive pore loop, i.e. Tyr354 folds inwards and interacts with F350 (thick sticks). However, the EM map is very poor, the volume is contoured at 0.04 in contrast to 0.08 as recommended by the authors.

functionally unrelated yeast proteins CDC48 and Pex1. This residue is replaced by a phenylalanine in Yme1. Similarly, Arg261 is conserved or conservatively replaced by lysine in most other homologues but it is changed to a glycine in Yme1. Yme1 appears therefore as an exception and its interfaces are obviously different in this regard. The unique character of Yme1 is also corroborated by the observation that it is the only FtsH homologue with a topology where AAA$^+$ and protease domain face the exterior.

Figure 55 depicts a simplified ATPase cycle based on a threefold symmetric hexamer. It gives an overview over the here discussed structural features with the focus on the two observed intersubunit interfaces, whose general structure also occurs in the asymmetric YME1 cryo-electron microscopy model. Our structures show the closed conformation directly after ATP hydrolyses and the subunits in open conformation prior to nucleotide exchange. Upon exchange, the pore motif of the open conformer intercalates between the sidechains of the substrate, as reported by Puchades et al. (2017). Thereafter, the subunit switches to the closed ATP bound
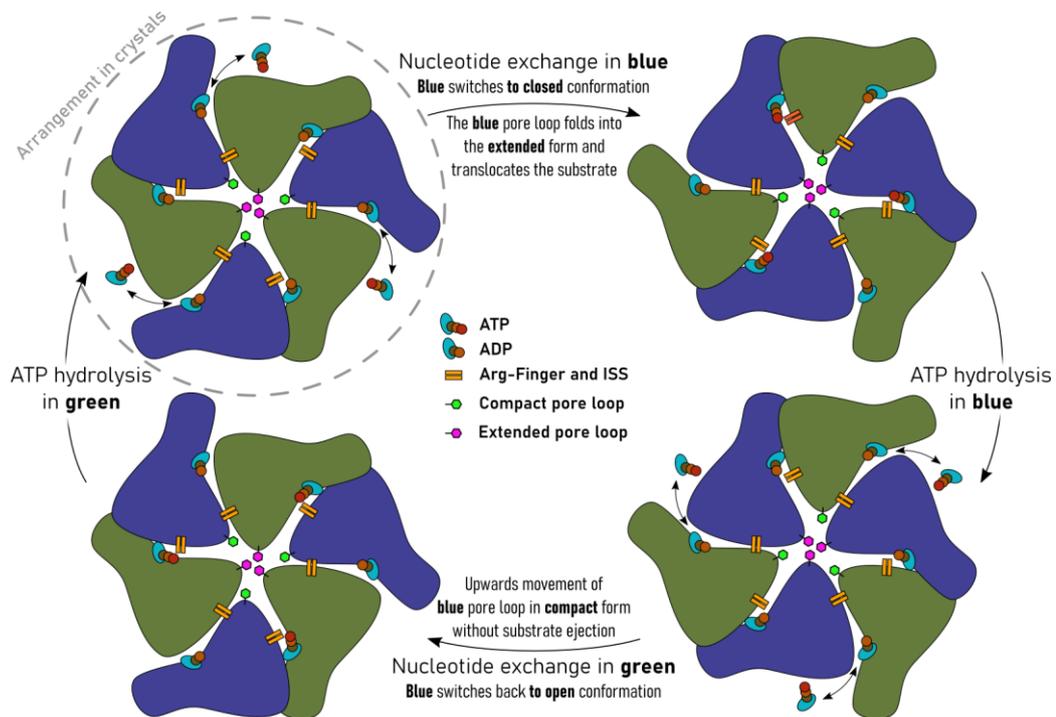
Figure 55: Very simplified overview of the ATPase cycle based on a threefold symmetric FtsH.

In order to advance thorough the ATPase cycle while starting with our crystal structure, the ADP in the open conformations (blue) are exchanged against ATP. Note: Regardless of the conformational state, the colour remains the same during the cycle, thus in the right column the blue subunits are in closed state. Post hydrolysis interface: between green and blue subunits, anti–clockwise manner, nucleotide exchange interface: blue/green. All colours are the same as in Figure 51.

state pulling simultaneously the substrate into the proteolytic chamber. Subsequently, $Arg313_{Op}/316_{Op}$ of the adjacent conformation stimulates the ATP hydrolysis yielding the here presented closed ADP bound state. While changing back to the open ADP bound state, the ISS is disrupted and thus induces the compact pore loop conformation retracting Phe228 from the substrate. This way the substrate is not pushed out of the pore again.

6 FTSH – RESULTS AND DISCUSSION

6.1 Purification strategies for FtsH and GFP

6.1.1 Expression and purification of *A. aeolicus* full-length FtsH was improved

In the beginning, a 12 l expression yielded enough bacterial pellets for up to two usual purifications. However, cryo-EM and the characterisation of mutants require a lot of fresh protein. Improving expression and purification was therefore advantageous.

At first, the LB medium was replaced with TB medium and additional phosphate buffer, magnesium chloride and antifoam B. TB medium has 0.4 % glycerol and more than twice as much tryptone and yeast extract compared to LB medium. Antifoam B reduces foam formation on top of the cultures thus increasing oxygen exchange. Mg^{2+} is believed to limit reproduction if oxygen supply is assured. Moreover, medium volume was reduced from 1 l to 0.5 l per flask to increase aeration and the old medium of the pre-cultures was exchange against fresh one before inoculation. β-lactamase can accumulated in that old medium thus reducing the selective pressure for expression vector containing bacteria. Eventually, expression overnight was induced with 1 mM IPTG and an additional dosage of ampicillin and antifoam B at around an OD_{600} of 2.0. After induction, flasks were cultivated at 30 °C as fast as possible and only covered with sponge-like plugs. The given reasons for the changes can be considered anecdotally and have not been further researched. The most critical changes are probably the late induction and the reduced temperature for expression as changing those parameters reduced the growth. Interruption of the shaking of culture is probably catastrophically, as perceived during several expressions (also with GgpS). As such, all interruptions were reduced to the minimum with minimal length. With half of the volume per incubator, almost three times more bacterial pellets were harvested (136 g per 6 l). During this expression, a cell density of 12.8 (OD_{600}) was achieved and protein quality was not affected as evaluated by SEC and negative stain EM.
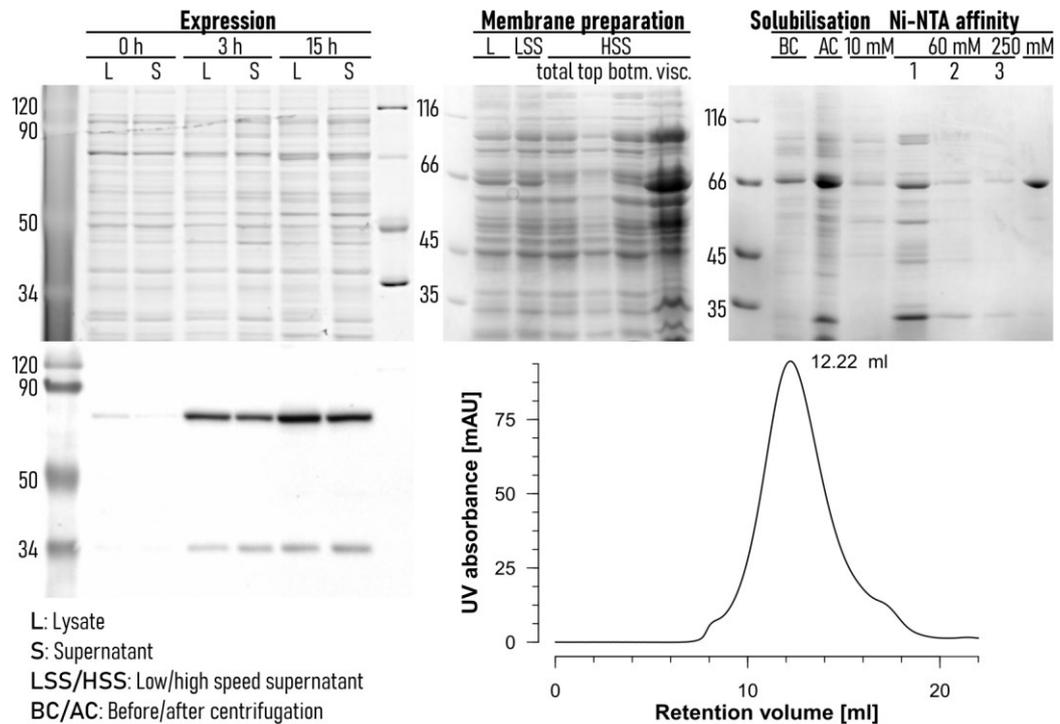
Figure 56: Expression and purification of full–length FtsH from *A. aeolicus*.

SDS-PAGE gels are either TCE (expression) or coomassie stained. Proteins in the western blot for the expression were labelled with α–penta–His–HRP. Solubilisation sample 'BC' is the product of membrane preparation with DDM before removal of non–solubilised proteins via centrifugation. Size: 71.75 kDa, runs at 66 kDa.

The first purification step is the isolation of the membrane fraction via low and high-speed centrifugation with 20,000 *g* and 256,000 *g* respectively. The precipitated membranes are taken up in TBS and are solubilised in TBS with 100 g/l n-Dodecyl β-D-malto-pyranoside (DDM). Roughly 20 % of the bacterial pellet are later isolated membranes. The supernatant after high-speed centrifu-gation often showed a visible gradient with a small viscous phase above the membrane pellet. The visible gradient matches a protein/FtsH gradient as determined by SDS-PAGE (Figure 56) and could be reduced by addition of 600 mM sodium chloride. However, this much salt often decreased solubility of the bacterial pellet hampering lysis and should be modified for better handling in the future.

A further improvement was the adjustment of the ratio of affinity tag chromatography bed volume and prepared membranes. It was evaluated that the binding capacity of Ni-NTA is less than 1 mg FtsH$_{Aa}$ per 1 ml resin. However, this is a rough and safe

estimation since the protein was first quantified after size exclusion. Finally, 20 ml of membranes per 6 ml Ni-NTA can yield around 2.0 mg of FtsH$_{Aa}$ after size exclusion (0.4 mg/ml protein/resin). The old protocol recommended five times more protein with the same amount of resin.

Combined, both improvements increased the yield from around 0.4 mg/l to more than 5 mg/l (final wt$_{Aa,FL}$ per culture) or 4.8 mg per incubator to 30 mg per incubator. The mutants behaved similar.

6.1.2 Expression and purification of *E. coli* full-length FtsH has to be improved

The purification of full-length FtsH from *E. coli* for activity assays at room temperature has to be improved. Bacteria of C41(DE3) and C43(DE3) accumulated wt$_{Ec,FL}$ while BL21(DE3) bacteria grew but did not express. The expression is usually characterised by a double band at/below 66 kDa in SDS-PAGE gels (Figure 57A). Only the lower one appears after induction. Blots indicated the upper band to be a random protein overlapping with the inducible His-tagged FtsH. The protein was mostly insoluble and levels apparently decrease after 2–3 h of expression. Solubilisation in DDM led to aggregates as depicted in the size exclusion profile (Figure 57C).

6.1.3 Expression and purification of cpGFP variants was satisfactory

$_{cp6}$GFP-ssrA was heterologously expressed for 4 h at 37 °C in either BL21(DE3) Star and C+RIPL or C43(DE3). Both BL21 strains produced greenish pellets indicating successful expression, however, C+RIPL also expressed another undesired protein (not shown) and was not used for later expressions. After lysis using a CF Cell Disruptor (Constant Systems), the His-tagged proteins were isolated via Ni-NTA affinity chromatography. A gradient was initially applied to estimated best imidazole washing concentrations of about 25 mM (Figure 58). The final purification step was a size exclusion chromatography with a Superdex 75. A representative SDS-PAGE analyses of the purification is depicted in Figure 59B.
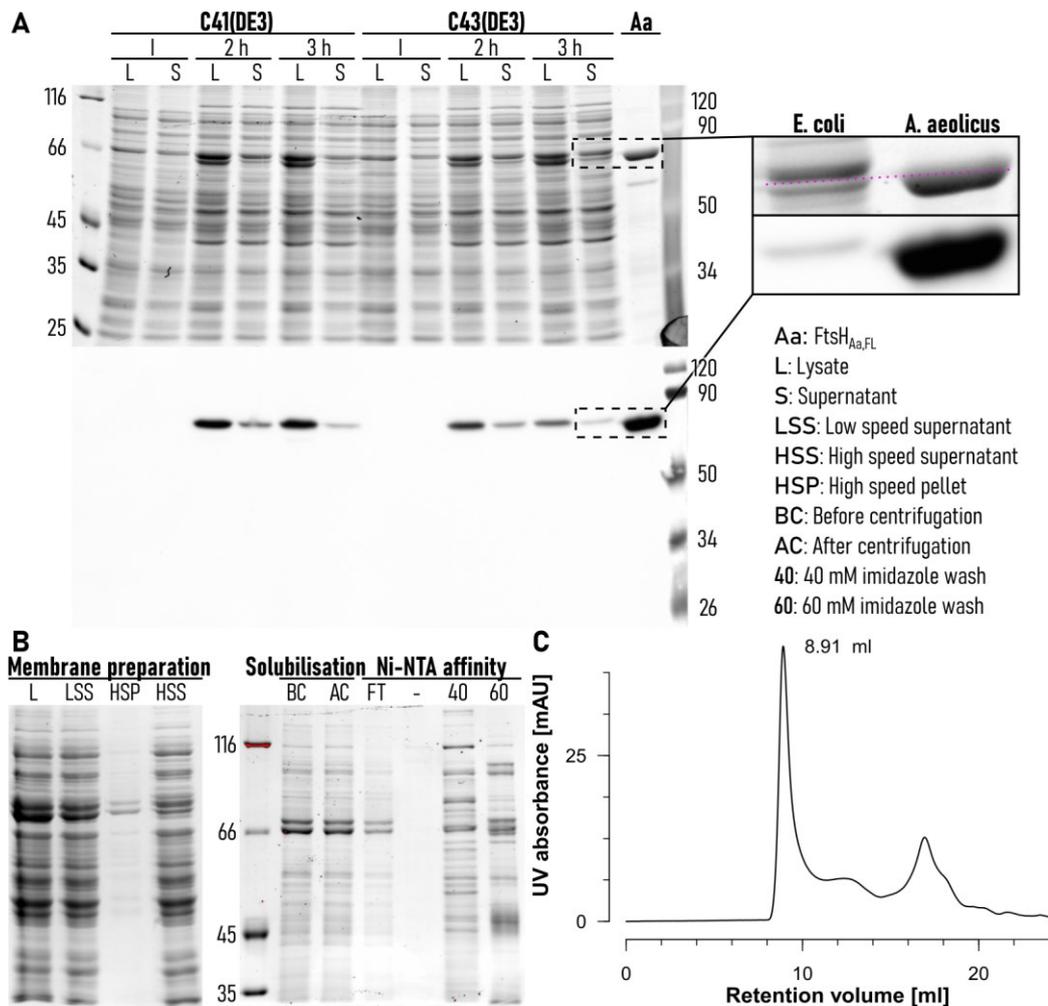
Figure 57: Expression and first purification of full–length FtsH from *E. coli.*

A) SDS–PAGE gel TCE stained and western blot with α–penta–His–HRP of expression in C41 and C43 bacteria. Inset: Zoom to show which band produces luminance i.e. has a His–tag. B) SDS–PAGE gel Coomassie (membrane preparation) or TCE stained (subsequent steps). C) Size exclusion profile of affinity chromatography eluate of purification in B). No eluate sample was loaded into the gel. Size: 71.77 kDa, runs above 66 kDa.

Besides $_{cp6}$GFP-ssrA, also the non-permutated template of the superfolder GFP ($_{sf}$GFP-ssrA) and further destabilised mutants were purified. However, purification of these mutants was not as simple. The size exclusion profiles indicated contaminations, and low yields hint for successful destabilisation but reduced expression yield (Figure 58). Nonetheless, the proteins should be useable for initial FtsH activity testing.
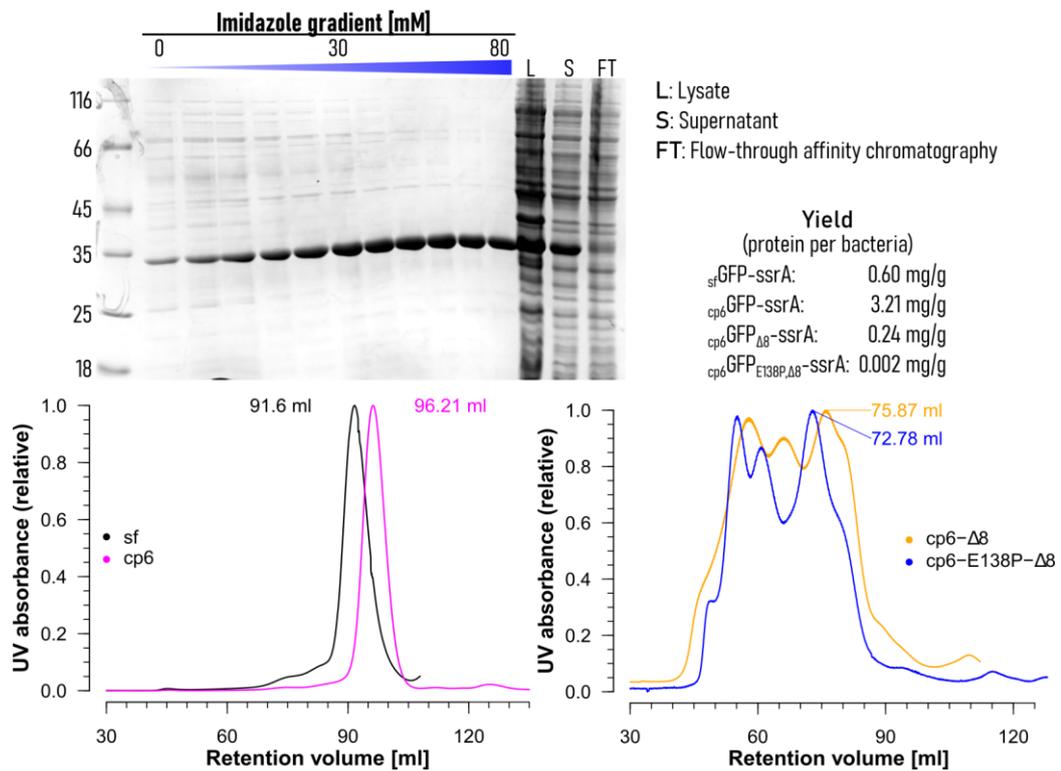
Figure 58: Purification of GFP variants.

The Coomassie stained SDS–PAGE gel depicts samples from $_{cp6}$GFP–ssrA but a different purification as the size exclusion profile. These are grouped according to the used column and were normalised to their maximum value each. As such, the peak of $_{sf}$GFP and mutant E138P–Δ8 were in fact only half and one sixth as high as their counter parts. Size: 30.85 kDa, runs below 35 kDa.

6.2 Mutants F224A and F228A cannot translocate casein

All assays show a lag time of around 30 min, most likely because the employed FtsH is from *A. aeolicus*, a hyperthermophile bacterium, but the assays were carried out at 37 °C without pre-heating. Two tests demonstrated temperature dependent activity between 20–60 °C. Thus, the 384-well plate warms up and activity increases slowly during the beginning of the assay; visible as initial lag phase. The assays were started by adding one volume of protein to one volume of reaction mixture. Pre-incubation of the plate with protein and addition of warmed reaction mixture reduced that lag phase only partially. In addition, the new lag phase was not consistent probably because of different cool down times while adding the reaction mixture; Different number of samples, replicates and proteins varied the time for pipetting. Hence, all ingredients were simply pre-incubated in the climate controlled laboratory, mixed and directly measured.

The full-length FtsH is expected to be predominantly hexameric. All molar concentrations represent the amount of hexamers and all derived variables described accordingly the hexamer.

6.2.1   Mutants F224A and F228A have transient ATP hydrolysis activity

The pyruvate kinase-lactate dehydrogenase ATP hydrolysis assay as developed for GgpS was successfully modified for FtsH (Figure 59A). Wild type k_{cat} was estimated based on two biological replicates with either 1.19 s^{-1} or 1.68 s^{-1} at 37 °C. One time measurements indicate lower k_{cat}s of 0.47 s^{-1} and 0.72 s^{-1} for the mutants F224A$_{Aa,FL}$ and F228A$_{Aa,FL}$ respectively. The true hydrolysis rate of the mutants could be higher because only 0.3 mM ATP were consumed, whereas the assay is designed for five times as much. As such, the lag phase overlaps with the early stopping. This could indicate that the active site is actually fully functional (as it has not been altered) but blocked after some cycles of hydrolysis. This further implies a strict connection between ATP site and pore loop, as proposed in our publication (section 5.1.4).

6.2.2   Mutants F224A and F228A cannot not translocate

The translocation capability of FtsH is analysed by a fluorophore labelled casein. Casein has no stable fold and requires no unfolding. The fluorophore is quenched in the uncleaved substrate but can fluoresce after processing. The development of the assay was complicated due to a few pitfalls. First, the aforementioned lag phase which could not be shortened as expected by pre-incubation at assay temperature. Second, the internal quenching is loosened due to the DDM for full-length FtsH solubilisation. This increased the background by a factor of two, compared to DDM-free samples (not shown) and of course reduced the range of possible fluorescence increase. Third, saturation was achieved quickly, thus interfering with the lag phase, although controls with trypsin indicated much more fluorescence gain is possible. Increasing the substrate concentration while decreasing the concentration of wt$_{Aa,FL}$ initially helped but was not sufficient. Using eventually six times as much substrate as recommended by the manufacturer and decreasing the concentration of wt$_{Aa,FL}$ by around 20 times
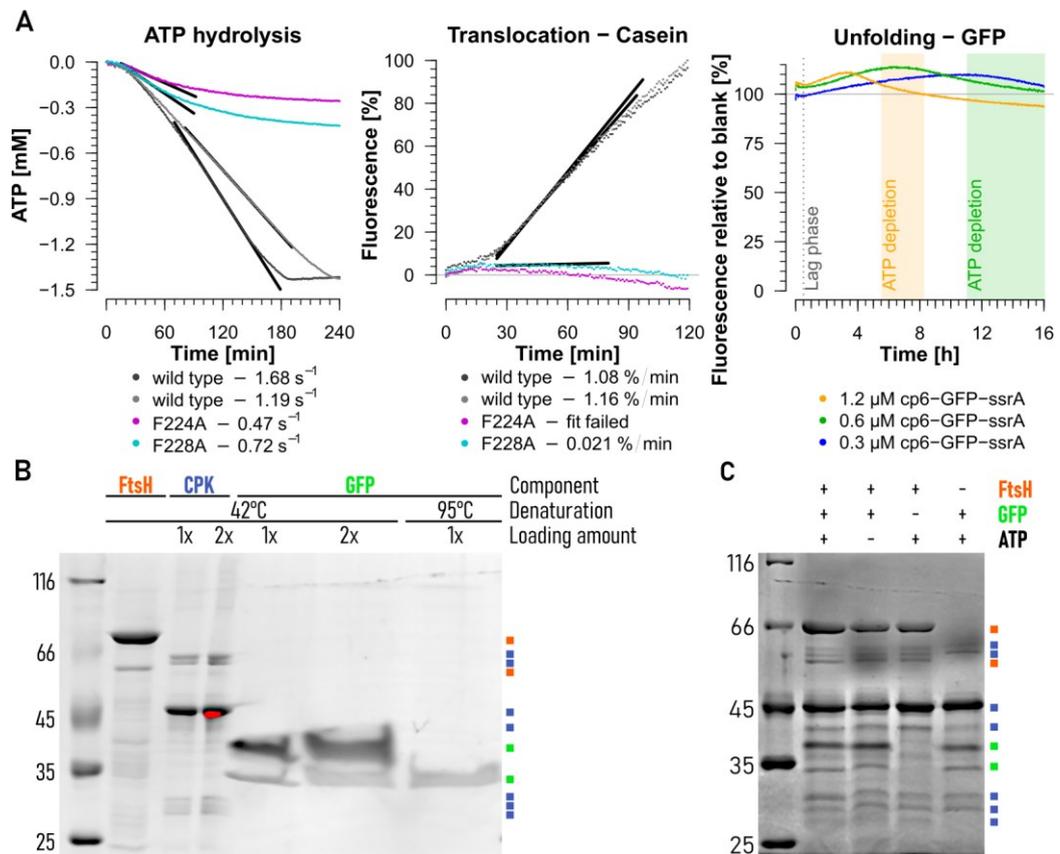
Figure 59: Activity assays for *A. aeolicus* full–length FtsH variants.

A) ATP hydrolysis tested via pyruvate kinase and lactated dehydrogenase coupling, translocation via fluorophore labelled casein and unfolding via GFP. The latter two utilise the creatine phosphokinase ATP regeneration system. k_{cat}s (hydrolysis) and velocities (translocation) are estimated based on linear regressions as indicated for the steepest slopes of at least 30 min (hydrolysis) or 15 min (translocation) length. Great reduction in the GFP unfolding assay is expected but did not occur. ATP depletion range is estimated based on lowest/highest ATP hydrolysis k_{cat} and with or without lag phase. B) SDS–PAGE gel stained with TCE of the proteins used in the unfolding assay. FtsH samples smear at higher denaturation temperatures. Hence, all samples have to be denatured at 42 °C. This leads to two GFP bands. C) SDS–PAGE gel stained with TCE of phenantroline inhibited unfolding-assay samples after incubation over-night. Colour key indicates proteins as derived from B). A complete active sample (first lane) shows a slightly truncated GFP (lower band).

yielded stable signals for k_{cat} estimations. However, the best wt$_{Aa,FL}$ concentration has not been titrated yet.

The mutants F224A$_{Aa,FL}$ and F228A$_{Aa,FL}$ have been preliminarily assayed (Figure 59A) and are found to be completely inactive. This result is not totally unexpected, considering their problems with ATP hydrolysis (just 1/5 is hydrolysed, then they stop completely). However, no translocation at all indicates further problems with in both mutants. Pruning one of them prevents their interaction with

each other in any case. According to my model, this interaction is required for proper pore loop retractions and subunit upwards movement without a substrate push back or clash (section 5.1.4).

6.3 The GFP unfolding assay is not useable with FtsH$_{Aa,FL}$

Probing the unfolding capabilities of FtsH requires a folded protein substrate. This is also closer to *in vivo* conditions since real FtsH substrates can be both, folded and undamaged. Moreover, it was planned as an alternative for the translocation assay, due to its initial problems. GFP has been used as proteolysis substrate before, but FtsH is a weak unfoldase and cannot unfold GFP (Christophe Herman et al., 2003). Hence the $_{cp6}$GFP-ssrA developed in the lab of Robert Sauer was tested. Degradation of GFP should decrease fluorescence since the fluorophore is degraded or at least solvent exposed. The latter has a quenching effect.

My results are similar to the reported results about FtsH's weak unfoldase activity by Herman et al. (Figure 59A). The fluorescence increases first and is followed by a steady decrease below initial values. SDS-PAGE gels indicate the truncation of GFP (Figure 59C). The fluorescence maximum is wt$_{Aa,FL}$ concentration dependent and happens before ATP is depleted. Hence, wt$_{Aa,FL}$ most likely recognises the ssrA tag and degrades it. Most of the GFP however is too stable and cannot be unfolded. ATP consumption and thus translocation/unfolding attempts stabilise the wt$_{Aa,FL}$:GFP complex transiently. This probably further shields the fluorogenic group from the solvent, enhancing fluorescence. The complexes dissociates at latest after all ATP is consumed or earlier due to random dissociation. The missing degraded part leads probably to stronger quenching by the solvent, thus decreasing the fluorescence slightly compared to initial values.

$_{cp6}$GFP-ssrA was further destabilised with the mutant E138P to proline and deletion of the last eight amino acids before the ssrA tag according to Perez-Jimenez et al., 2006; Wohlever et al., 2013. However, this did not change anything. The GFP-unfolding assay is not usable with FtsH from *A. aeolicus* at 37 °C.

6.4 Crystallisation of F224A, F228A and Q281A

Similar to the assays, mutants F224A$_{Aa,tr}$ and F228A$_{Aa,tr}$ but also Q281A$_{Aa,tr}$ were crystallised in the known conditions. Crystallising the mutants led to smaller and worse crystals of which Q281A$_{Aa,tr}$ diffracted best in comparison (Table 13). No (processable) crystals/datasets were obtained from F228A$_{Aa,tr}$. The pore loop expected to be in the extended conformation of F224A$_{Aa,tr}$ is always disordered. This was also observed in a second but worse dataset of this protein. Besides this, no structural changes became apparent. However, none of the mutations should induce or prevent a certain fold. Pruning the sidechains to alanine only removes constraints and allows for more flexibility as demonstrated by the bad quality of crystals or their absence.

New initial screens led to the discovery of slightly different crystallisation conditions (Morpheus G6, Molecular Dimensions). A fine screen yielded promising crystals in carboxylic acids (25 mM of each: sodium format, ammonium acetate, disodium citrate, K/Na-tartrate) and 18 % ethylene glycol and 9 % polyethylene glycol 8,000 in presence of 0.1 M Na-HEPES/MOPS pH 7.25. Being more cubic in shape, the crystals were different from the usual ellipsoid from. Indexing yielded a slightly thicker but lower cell of $a = b = 208$ Å and $c = 280$ Å compared to the normal one of $a = b = 198$ Å and $c = 328$ Å. However, resolution was always worse than 7 Å, hence no structure could be derived.

The results do not disagree with the proposed model. The amino acids in question only stabilise or induce the compact (F224/F228A) and extended (Q281A) conformations. Both forms are found nonetheless, probably because of the crystallographic environment (Figure 53B) and because the mutation do not actively prevent these conformations.

6.5 New solubilisation/reconstitution strategies? – No DDM alternatives!

DDM is a detergenent commonly used for membrane protein purification (Bloch et al., 2020) and has been used successfully for
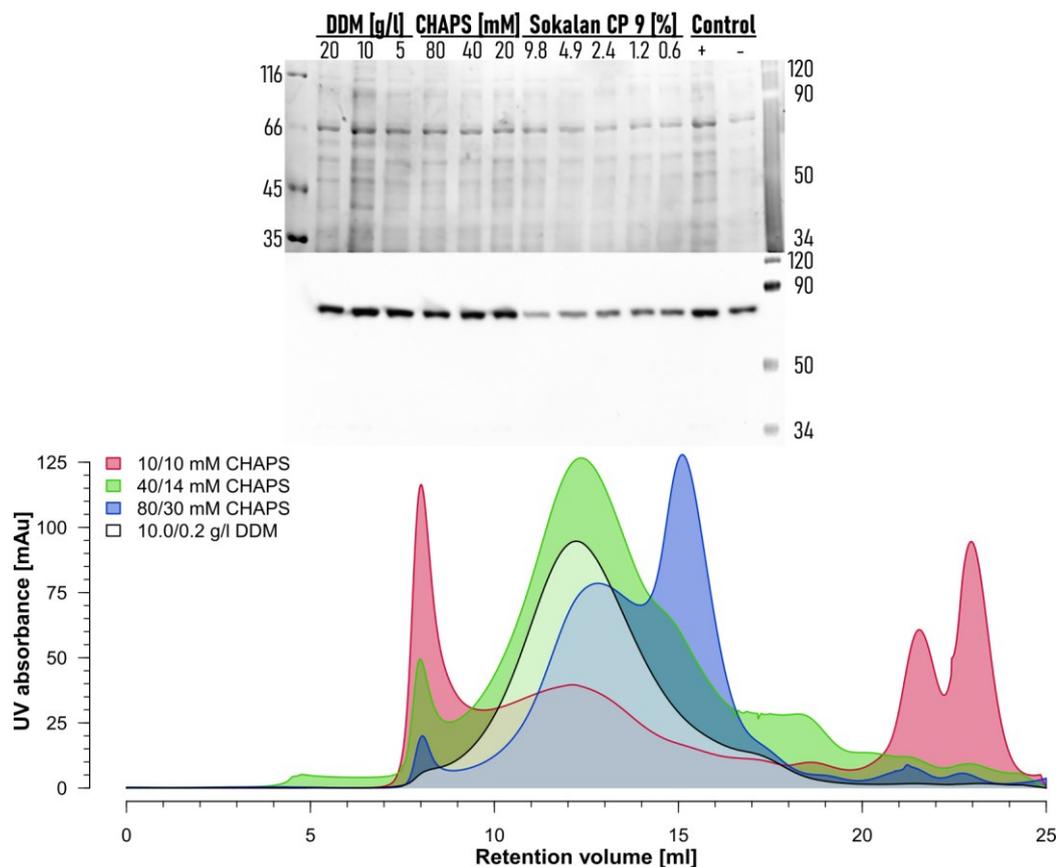
Figure 60: Solubilisation test with full–length FtsH from *A. aeolicus*.

100 µM samples were solubilised (according to established DDM protocol) and subsequently centrifuged to pellet all non–solubilised proteins. The controls are treated as the samples but the positive control was not centrifuged. The negative control contains a lot of FtsH. Samples were only centrifuged at 20,000 *g* for 20 min leading to incomplete removal of not solubilised proteins. The SDS–PAGE gel is TCE stained. The plots compare size exclusion chromatograms after different CHAPS solubilisation with a DDM standard. '80/30 mM CHAPS' indicates solubilisation in 80 mM CHAPS and later buffer supplementation with 30 mM CHAPS. Critical micelles concentration of CHAPS is 6–10 mM according to the datasheet from the supplier Sigma Aldrich.

cryo-EM (Deng et al., 2018; Iadanza et al., 2016; Parey et al., 2018). Initial tests showed a strong preference to the carbon of the grid itself, depleting the holes. Higher concentrations of DDM also hamper ice film formation in the holes and the high curvature of detergent micelles could distort the protein. Hence, alternative solubilisation methods were analysed.

6.5.1   SMA and CHAPS cannot replace DDM as solubilisation agent

Maleic acid-olefin copolymer (Sokalan CP 9) and CHAPS were tested as DDM solubilisation alternative. All these three agents can solubilise $wt_{Aa,FL}$ but Sokalan CP 9 is less effective (Figure 60) and CHAPS concentrations have to be carefully chosen as purifications
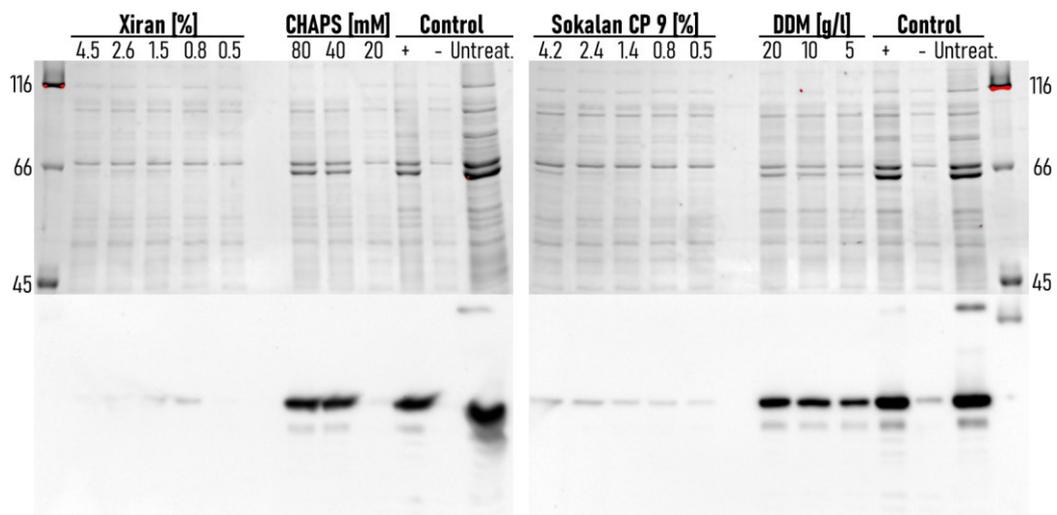
Figure 61: Solubilisation tests with full-length FtsH from *E. coli*.

Samples were solubilised (according to established DDM protocol) and subsequently centrifuged to pellet all non-solubilised proteins. The controls are treated as the samples but the positive control was not centrifuged and the untreated control was diluted from stock membranes directly before loading buffer addition. Samples were centrifuged equivalent to the usual purifications by centrifugation at approximately 190,000 g for 9 min. All SDS–PAGE gels are TCE stained.

demonstrates. Too much CHAPS tears the $wt_{Aa,FL}$ hexamer apart and too less, close to the critical micelle concentration, facilitates aggregation. Initial solubilisation with 40 mM and later supplementation with 14 mM CHAPS is seemingly a local optimum but still leads to aggregation and an assumed monomer shoulder (green profile). Less aggregates and no monomers are visible if solubilised according to the default protocol with 10 g/l DDM initially and later supplementation of about twice the critical micelle concentration. The translocation assay shows accordingly a smaller activity (Figure 62A). Consequently, zwitterionc detergents are often considered as harsh in contrast to non-charged ones as DDM (Bloch et al., 2020). Neither Sokalan CP 9 nor CHAPS were further used.

Since full-length FtsH from *E. coli* has been purified for the first time in the lab, testing alternatives to DDM was obvious (Figure 61). In addition to CHAPS, DDM and Sokalan CP 9 also the styrene maleic anhydride copolymer Xiran was tested. As previously described, $wt_{Ec,FL}$ runs at the position of a different similar sized protein but the western blot clearly indicates almost no solubilisation with Xiran and Sokalan CP 9 and 20 mM CHAPS.
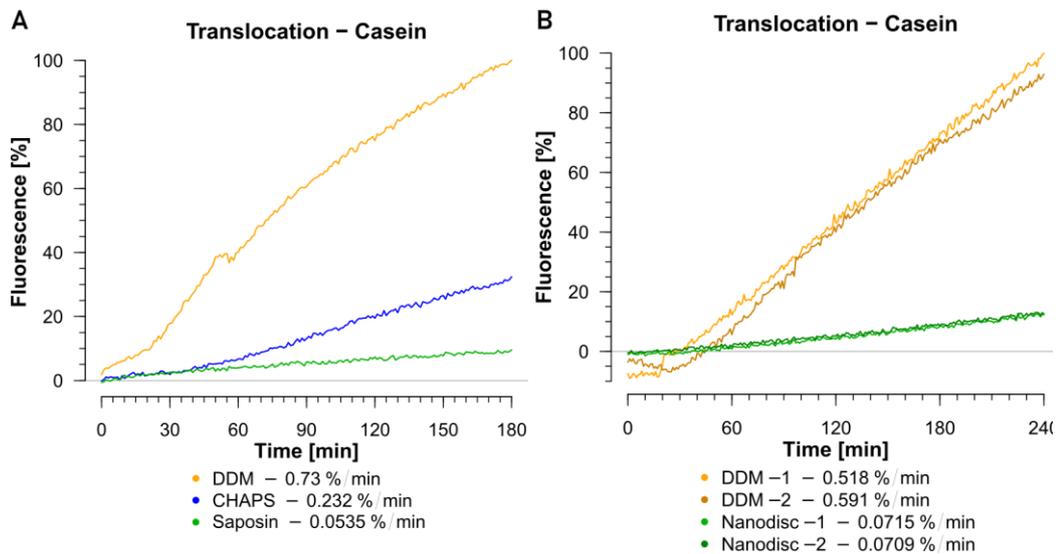
Figure 62: Translocation assays with differently solubilised/reconstituted wt$_{Aa,FL}$.

A) CHAPS (40/14 mM, main peak) and Saposin as technical triplicates and B) Nanodiscs with biological duplicates for DDM and two different HPLC fractions of Nanodiscs. Assays have been performed before a standardised assay was established. Hence, due to different detector gains, substrate and protein concentration amongst others, velocities are not comparable between A), B) and Figure 59A. Assay properties per plot are of course identical, however, stoichiometry of the saposin disc is unknown. Protein concentration was estimated by BCA assay and a ratio of one hexamer to five saposins was assumed and. In addition, the ratio of FtsH containing and empty particles is unknown but the former should dominate strongly according to EM images.

Subsequent purifications with higher CHAPS concentrations have not been performed yet.

6.5.2   Nanodiscs and saposin

Protocols for nanodiscs and saposin discs were developed in Elmar Behrmanns group by Gayathri Jeyasankar and subsequently tested in negative stain or cryo-EM preparations. In addition, both were tested once in the translocation assay (Figure 62). Folch I fraction of bovine brain extract was used with saposin A and phosphatidylcholine POPC with MSP1D1 nanodisc protein.

According to the assays, saposin or nanodiscs reconstituted wt$_{Aa,FL}$ are approximately ten times less active. The quantification for saposin particles is less accurate, as the assay protocol was an early version and the stoichiometry of FtsH hexamers to saposin could only be assumed. However, the estimated stoichiometry of 1:5 (from cryo-EM) corresponds to a protein mass ratio of 430.5 : 59.5 kDa, hence a saposin more or less has no tenfold
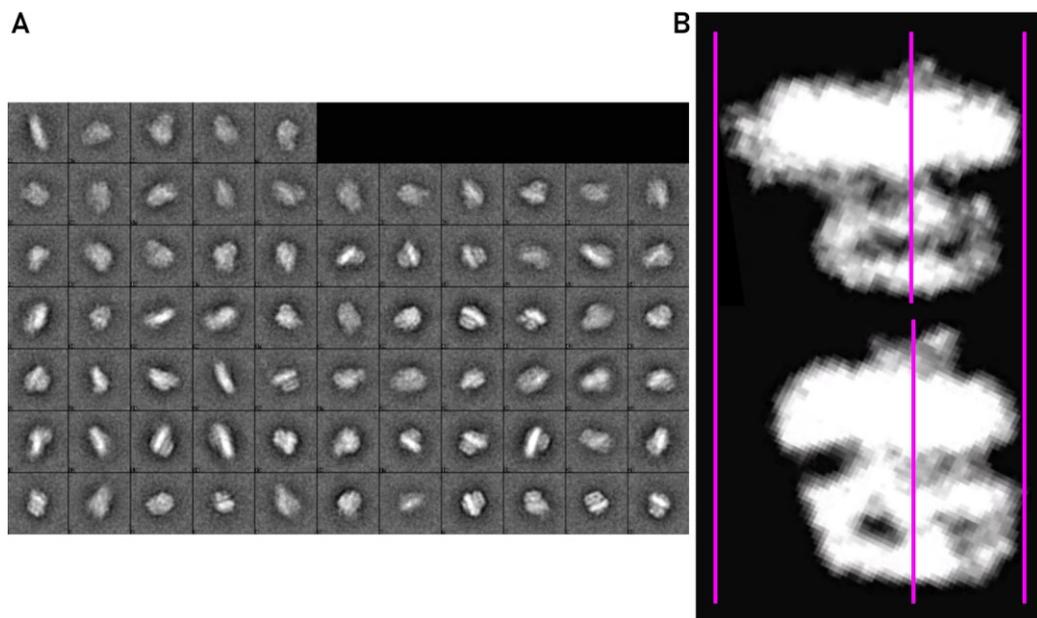
Figure 63: FtsH$_{Aa,FL}$ in saposin A discs with Folch I fraction of bovine brain extract.

A) 2D classes in no particular order. B) Selected 2D classes. FtsH is not in the centre of the lipid disc and the disc itself has varying sizes.

impact. In addition, both activities can be underestimated since empty particles were not completely separated (empty nanodisc: 49.7 kDa), however, at most 10 % of the lipid discs should be empty, thus not explaining the strongly reduced activity.

The reconstitution with saposin A was initially perceived as a good solution due to early negative stain images, but a dataset with 9,000 particles revealed the lipid belt to be of no constant diameter (Figure 63). Moreover, the FtsH hexamer was not in central position. Both phenomena increase heterogeneity and thus raise the number of required particles. In addition, it was unclear if this heterogeneity could be filtered correctly by the classification programs. The development of the correct protocol for nanodisc formation and freezing took longer than anticipated. Eventually, FtsH in DDM was preferred, especially after the results described in the next section.
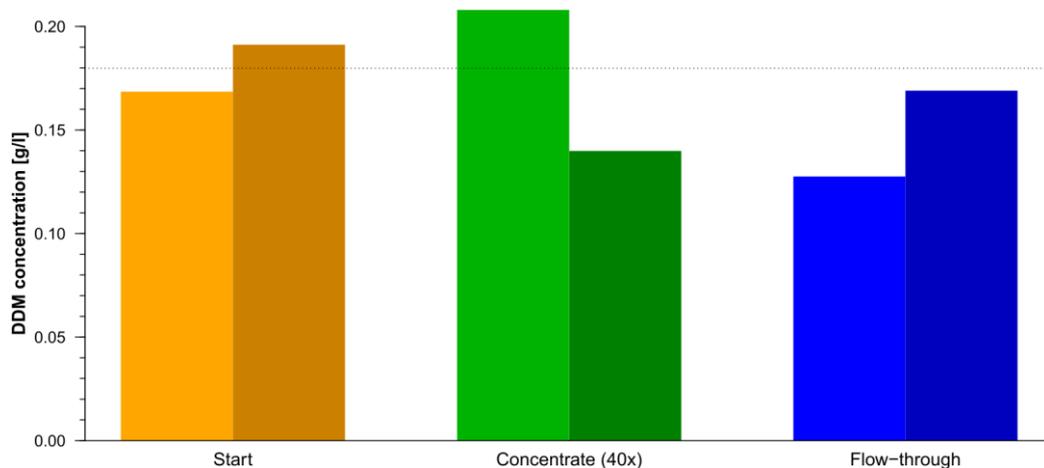
Figure 64: 100 kDa molecular weight cut–off ultrafiltration membranes do not concentrate DDM.

DDM concentration was with 0.18 g/l close to the desired 0.20 g/l before centrifugation. The 40x concentrate (0.17 g/l) and the flow–through (0.15 g/l) contain both a similar amount afterwards. Given values are averages of the individually plotted duplicates.

6.5.3   DDM is not concentrated while concentrating the protein

A usual step during the purification is the concentration of the protein using an ultrafiltration membrane with a molecular weight cut-off of 100 kDa. All co-workers with membrane protein experience were afraid that the 74 kDa DDM micelles (Strop & Brunger, 2005) concentrate during this step as well. This would lead to 20–40 times as much DDM as actually aimed for, and hence severely aggravating grid preparation. Excess DDM could be removed by incubation with Biobeads SM-2 also used for hydrophobic interaction chromatographies and DDM removal during reconstitution in nanodiscs. Since not all DDM had to be removed, its concentrations had to be evaluated using an assay with phenol and sulphuric acid. The addition of those compounds leads to the formation of phenol derivatives with absorption maxima between 400 nm and 500 nm (Urbani & Warne, 2005).

Long story short, while establishing the assay I tested the flow-through of a typical concentration run. DDM is not concentrated in a 100 kDa molecular weight cut-off ultrafiltration device as also reported by Strop and Brunger and thus DDM does not have to be removed at all (Figure 64).
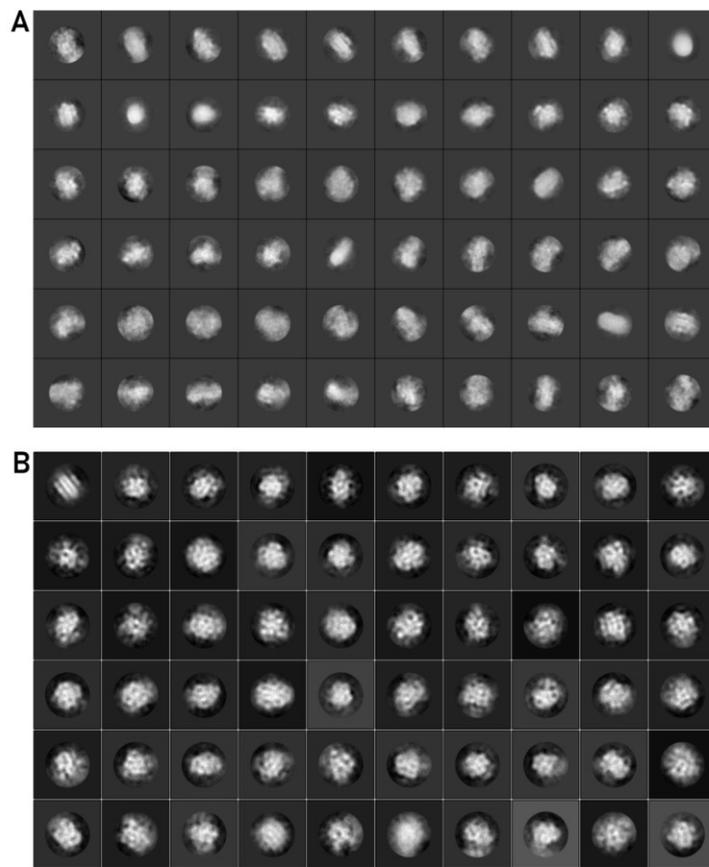
Figure 65: First cryo–EM datasets of wt$_{Aa,FL}$ with DDM.

Initial 2D classes of A) the very first dataset collected at the ESRF and B) of the second collection session. The latter are crispier but still do not show any secondary structural features except from, the most abundant first class on the upper left. C) Two selected 2D classes of wt$_{Aa,FL}$ in saposin A discs demonstrate the size difference of the lipid belt and the eccentric position of FtsH.

6.6 The first cryo-EM models of FtsH$_{AaFL}$ at 6–7 Å resolution

While several negative stain and cryo-EM datasets were collected, most suffered from certain pathologies. Initially FtsH in DDM preferred to stick to the carbon of the grid thus depleting the ice film. Graphene oxide helped getting better particle distribution but data collection at the ESRF, Grenoble, France yielded due to still low density only 6,832 particles from 1,824 micrographs. 2D classes including 5,651 particles (Figure 65A) were used to calculate a 23 Å model. After applying a second time at the ESRF, conformational heterogeneity became apparent as 2D classes did not improve (Figure 65B) although more than 16 times more particles were picked automatically via neuronal networks in crYOLO
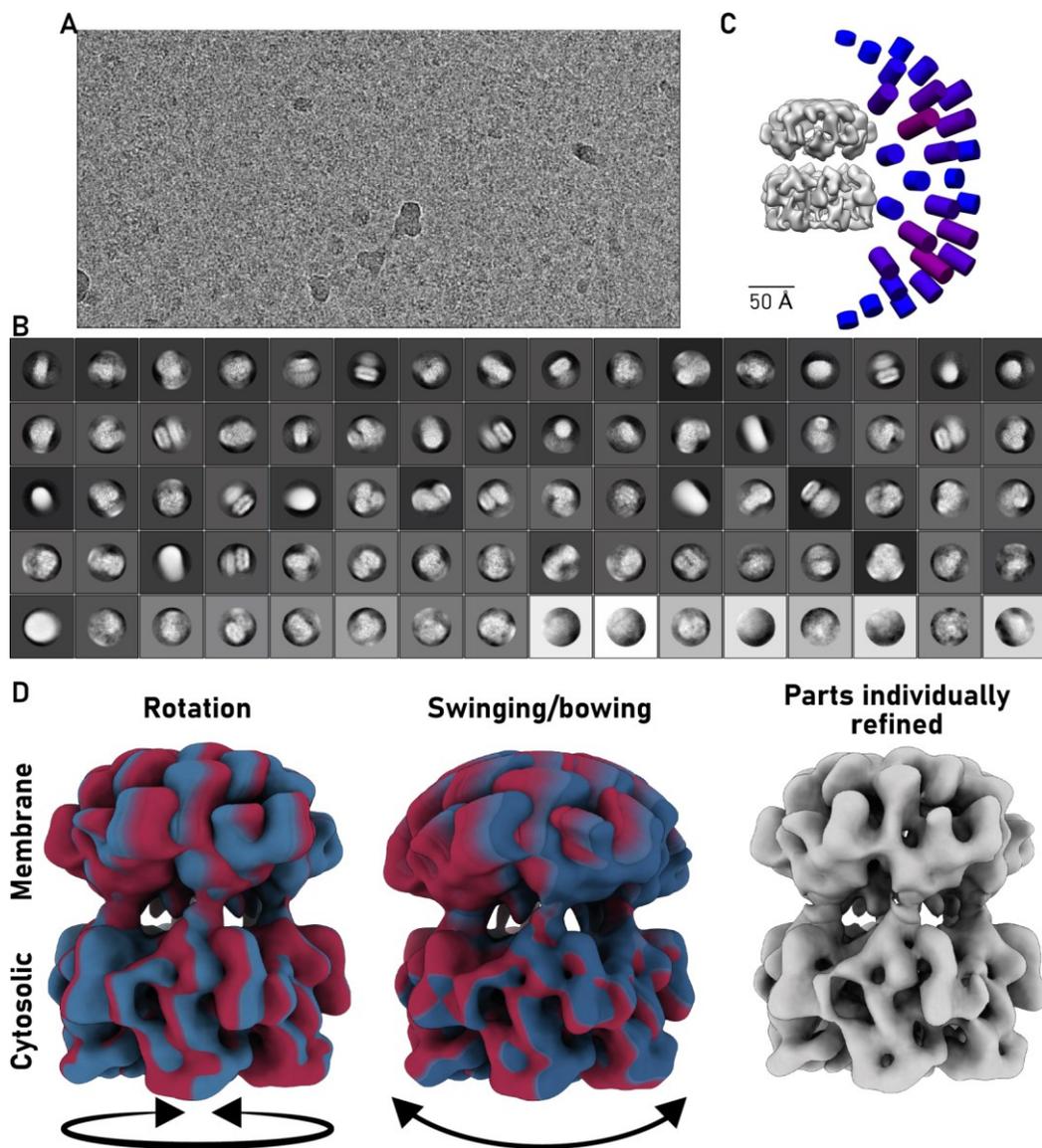
Figure 66: First dataset with ADP of wt$_{Aa,FL}$ in DDM on continues amorphous carbon supply.

A) Lower half of a representative micrograph with continuous amorphous carbon support. B) Initial 2D classes of 1,206,508 particles, note the lack of top/bottom views. Population size decreases from left to right and top to bottom. C) Ab initio 3D model with enforced C6 symmetry of a subset of 458,940 particles. Cylinder length and redness represent frequency of views of that direction. D) 3D multi–body refinement revels two major types of movements of cytosolic and membrane parts against each other (gradient red to blue). An individual refinement of both multi–body refinement models enhances resolution to 6–7 Å (grey).

(Wagner et al., 2019). Subsequent 3D classifications with *ab initio* and *in silico* models did not helped either.

Conformational heterogeneity could be reduced by incubation with ADP beforehand and vitrifying on continuous amorphous carbon (Figure 66A). Using a newly trained neural network based
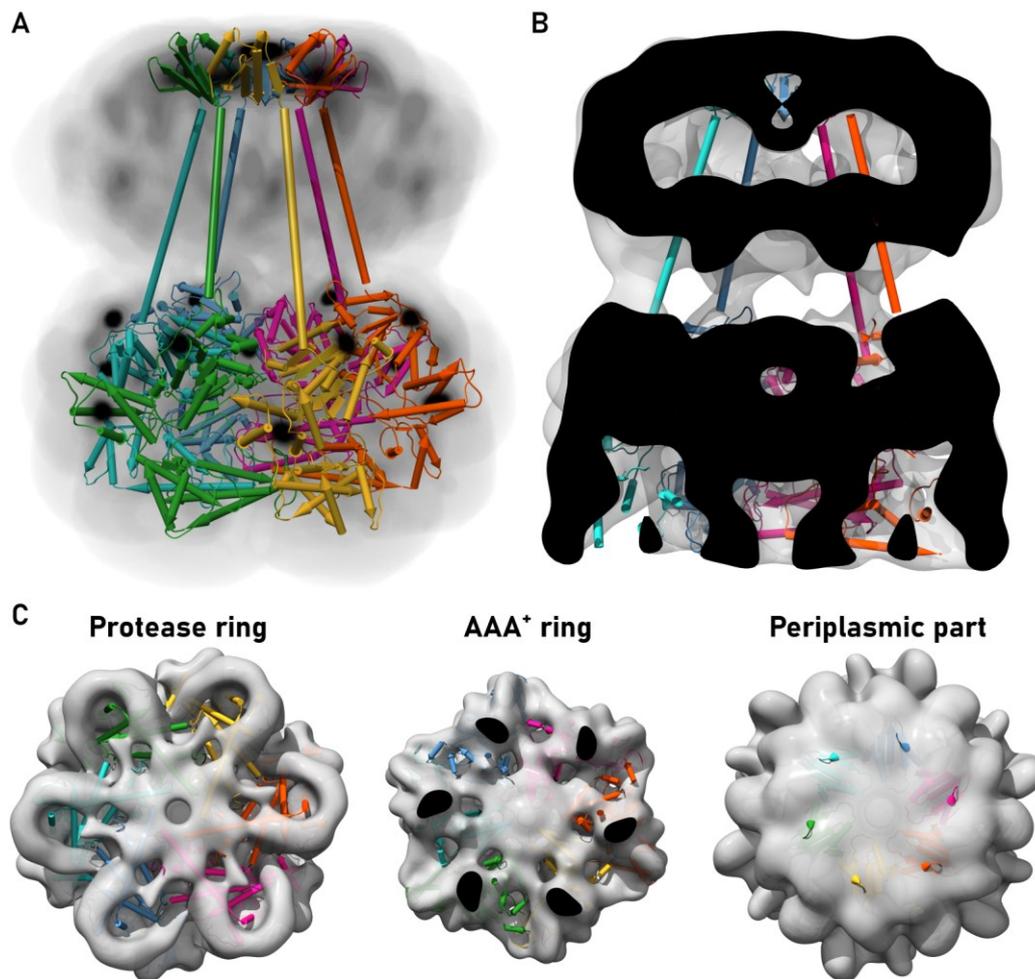
Figure 67: Cryo–EM model of FtsH. Sixfold symmetric periplasmic domain of *E. coli* (4v0b, Scharfenberg et al., 2015) and the threefold symmetric cytoplasmic part from *A. aeolicus* (6gcn, Uthoff & Baumann, 2018).

The second transmembrane helix and the linker between membrane and cytoplasmic part of each subunit is indicated by a long tube. A) Black dust: C6 symmetric Cryo–EM map, opacity scales linearly with map density. B) Cut through the middle of the hexamer. Black faces are the clipped map, map contour is depicted in transparent grey. A cavity with lower coulomb potential becomes visible in the membrane part. Spaces of low potential are also found in the protease ring. C) Views on the different parts of the hexamer. Black faces are again the clipped map.

on 16 micrographs in crYOLO yielded 1,206,508 particles from the 8,312 micrographs of the dataset collected at the Titan Krios (Thermo Scientific) at Caesar in Bonn (Figure 66B). An *ab inito* 3D model with C6 symmetry was composed of 458,940 particles (Figure 66C) and 3D classification further focused on 188,553 particles. An auto-refinement with enforced C6 symmetry yielded eventually an 8 Å model similar to refinements with C3 and C1 symmetry. However, clear distinction is not possible as the resolution is yet too low. Moreover, the dataset has still less
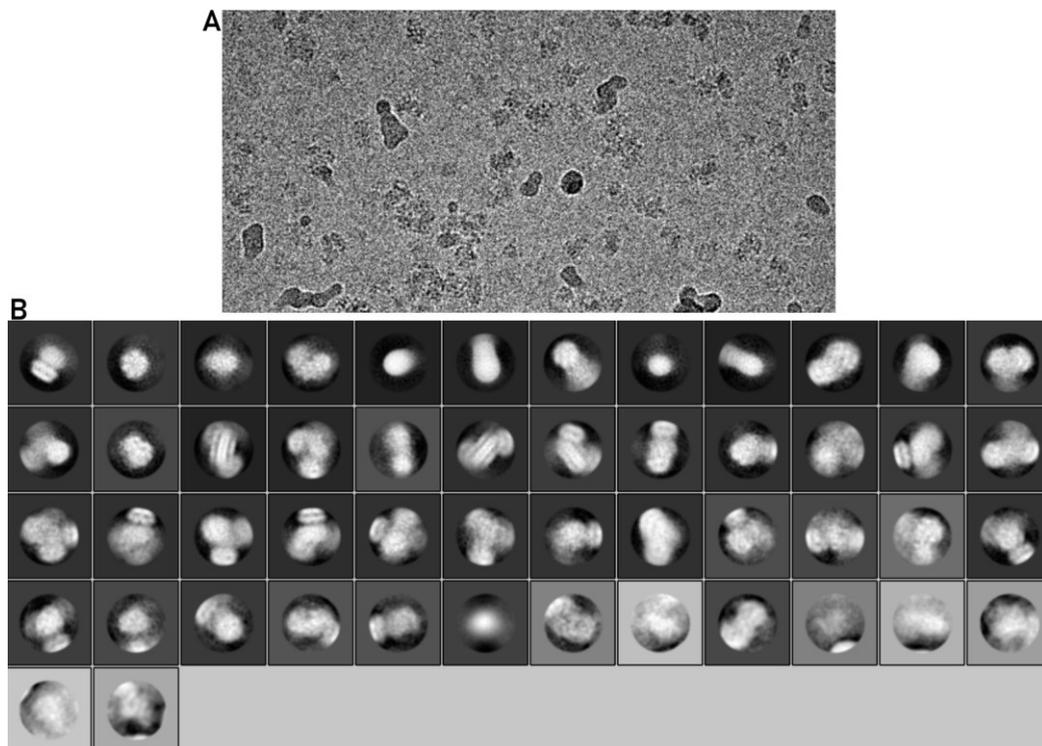
Figure 68: New preliminary dataset with ADP of wt$_{Aa,FL}$ in DDM without support.
A) Upper half of representative micrograph. B) Initial 2D classes with side and top/bottom views. Population size decreases from left to right and top to bottom.

top/bottom than side views (Figure 66C). Nevertheless, a 3D multibody refinement of the C6 symmetric model improved the resolution to 6–7 Å and revealed two different modes of motion (Figure 66D). Most prominently is a rotation around the symmetry axis, the whole cytosolic region rotates against the membrane/periplasmic region. Of lesser prominence is a swinging/bowing motion in which the distance between both regions decreases on one side and increases on the other (Figure 66D).

The C6 symmetric map with C6 symmetric periplasmic domain from *E. coli* (Scharfenberg et al., 2015) and the C3 symmetric cytoplasmic part from *A. aeolicus* (Uthoff & Baumann, 2018) is depicted in Figure 67. The orientation was chosen based on the small and round shape of the assumed membrane/micellar part. The detergent molecules further away from the protein are usually less ordered. This can potentially also explain the cavity inside the micelle. The transmembrane helices of each subunit are apparently not in contact with each other as indicated by the connecting potential between AAA$^+$ ring and membrane part (Figure 67C,

middle). A lower coulomb potential can also be observed in the protease ring (Figure 67B and left in C). The reasoning behind this is so far unclear but the phenomenon is present in all other preliminary low resolution cryo-EM models. The fitted periplasmic domains are in the flat and not the crown-like conformation. Figure 43 in the introduction is based on this fitting here and gives the height of each part.

The here presented model is a huge step towards a full-length model of FtsH using cryo-EM. However, the resolution has to be increased. New grids without any support layer were vitrified using a microdispenser (SPT labtech, Figure 68A). This reduced preferential orientation as demonstrated by 2D classes of a small preliminary dataset collected using a Talos Arctica (Thermo Scientific) at the Max Plank Institute of Molecular Physiology in Dortmund (Figure 68B). A bigger collection (Titan Krios (Thermo Scientific) at MPI) of this grid is currently combined with the one described before.

7 FTSH – SUMMARY AND CONCLUSION

7.1 The assays are partially established

The ATP hydrolysis assay was easily adapted from GgpS yielding very promising results. The translocation/protease assay with fluorophore labelled casein is in a useable state as well. However, the allowed FtsH concentration range should be determined by a simple titration with the wild type before further assays are done. The GFP unfolding assay on the other hand does not work, so far. The GFP is after destabilisation via circular permutation, proline insertion in β7 and removal of amino acids which seal one end of the barrel and thereby stabilising it, still no functional substrate. A similar assay was used for AFG3L2 but with a differently destabilised GFP (Puchades et al., 2019). The linker between the two last β-strands was cleaved by thrombin before the assay. This does not harm the GFP but the last β-strand is less likely to insert back into the β-barrel once it is extracted. This enables the complete degradation of that β-strand and loss of fluorescence of the remaining GFP.

All assays show a lag phase of about 30 min. This certainly comes from not pre-incubating the assay components. A reliable pre-incubation method (except from heating the whole room) had not been found as the small volumes in the 384-well plates cooled down rather quickly. *A. aeolicus* is a hyperthermophilic bacterium and thus adapted to temperatures as high as 95 °C (Deckert et al., 1998). Its FtsH has thus a reduced activity at lower temperatures and more than room temperature is needed. Another hypothesis partially explaining the lag phase could be a loss of the 'spiral staircase' order. The crystal structures all show some sort of symmetry which is apparently not the active yet a possible assembly. Addition of ATP might slowly start the ordering but requires some hydrolysis steps before full activity and translocation is possible.

7.2 Implication on the proposed model of the new results

We proposed a model for substrate translocation in our publication (section 5.2, Uthoff & Baumann, 2018). It is based on a threefold symmetric hexamer which is probably not the correct state. However, as already discussed there, the key points in our model about pore loop conformations and inter-subunit signalling remain valid.

The here reported results do not contradict the model. $F224A_{Aa,FL}$ and $F228A_{Aa,FL}$ exert no translocation of a loosely folded casein and further analysis revealed a problem with the ATP hydrolysis. The latter is seemingly possible for a short period of time with at least 30 % and 50 % activity for $F224A_{Aa,FL}$ and $F228A_{Aa,FL}$ respectively. However, hydrolysis stops around consumption of 1/5 of the available NADH reporter. This links both amino acids directly to ATP hydrolysis as proposed. Initial hydrolysis could be explained by the requirement of correctly ordering the subunits after purification in absence of fresh ATP. Hydrolysis happens at a reduced rate under this conditions thereby potentially also explaining to some degree the lag phase. Translocation however, is not possible by an unordered stage. Once the order is established the mutated pore loops cannot reliable fold into the compact form and hinder translocation and hydrolysis as proposed.

The crystal structures reveal unfortunately now new insights other than that crystal growth is more difficult. The substituted amino acids only induce or stabilise the compact pore conformation but the new introduced residues do no actively prevent it. Hence, crystals probably grow worse than with the wild type protein but the same conformations can still appear. New crystals with a differing cell but same space group apparently have a different fold, however in the worst case the indexing is already wrong due to very low resolution. Interesting experiments for better characterisation would the substitution of the F224 and F228 by charged amino acids. Opposite charged residues might retain activity while similar charges could actively abolish remodelling

and thus the activity. A less severe change would be the substitution of just one of the phenylalanines. Such mutations with structural analysis have never been done before.

7.3 Structure determination of FtsH$_{Aa,FL}$ via cryo-EM is going on

Membrane proteins are always more complicated than water-soluble polypeptides. Reconstitution in saposin or nanodiscs was technically successful, however, the activity was severely reduced at 37 °C. Moreover, saposin particles were of no equal diameter and nanodisc were apparently harder to vitrify than FtsH$_{Aa,FL}$ in DDM. ADP helped eventually to reduce conformational heterogeneity while ATP has not been used yet. The benefit of ADP can be explained by the hypothesis of an unordered hexamer due to the absence of ATP during purification and reordering upon nucleotide interaction. If this is true, ATP could potentially be more potent, but likewise increase heterogeneity if too much ATP is hydrolysed and not replaced.

The current dataset with initially more than 1.2 million particles and eventually 188,553 for an 8 Å model indicate that heterogeneity is still present. This could be a feature, as the movements observed after 3D multi-body refinement are expected and at least one was also observed in a different detergent (Carvalho et al., 2020). The first mode is a rotation of the cytosolic against the membrane part, most likely enabled by the unordered linker between them. The second movement is probably helping to recognise and unfold substrates. First of all it enlarges the space between membrane and cytosolic part on one side of the hexamer. The substrate has to travel through this space and the larger it is, the easier it might be to thread it through initially. This deformation could also absorb forces during unfolding as the substrate must be located besides the hexamer; the space above the hexamer is limited by the membrane and the transmembrane part of FtsH. As such, the substrate will pull with equal force on FtsH but sideways. This could distort FtsH, but is of less concern if the cytoplasmic part inclines. Secondly, this inclination also agrees

with the assumption that FtsH is a potent membrane extractor (Yang et al., 2018). The authors report about FtsH exceptional membrane protein unfolding capabilities and explain it with the same hypothesis that the substrate must be beside the transmembrane part of FtsH. As such, the polypeptide is not extracted perpendicular from the membrane but in a sideways inclined fashion. Coincidentally are most transmembrane helices not perpendicular in the membrane but inclined by around 20° (Bowie, 1997). Hence the flexibility of the linker would allow FtsH to pull with maximum force on the trans-membrane helices of a substrate while minimising the friction of the substrate with the AAA^+ ring and the membrane.

Currently another ADP dataset on support-free grids is evaluated. It can be merged with the old one, thus increasing particle count and hopefully yielding a higher resolution model.

7.4 Recent publications in the field

7.4.1 Carvalho et al., 2020

The most recent publication is by Carvalho et. al from Delft University, a preprint was available three years earlier (Carvalho et al., 2017, 2020). They ask for our $FtsH_{Aa,FL}$ for biochemical analysis and decided to solubilise in LMNG instead of DDM. They claim it to be more active but do not present any evidence. Moreover, their active particles include a big fraction of dodecamers instead of proper hexamers, as visualised by a cryo-EM model at 25 Å. These dodecamers are distorted similar to the movements observed in our models and they conclude that inclination is an important feature. Thus, deleting all amino acids belonging to the linker would prove its importance and the mutant is indeed inactive. However, as they removed the linker, they also moved the pore entry directly beneath the membrane. It is doubtful that any substrate could enter at all because there is now space left for the substrate. Hence they validate the importance of the linker and the space it opens but not necessarily the flexibility in question.

They also present a 6.6 Å C6 symmetric and a 15.9 Å model of a subset of the particles without implying any symmetry. Their results are in general agreement with ours and the 15.9 Å is potentially as asymmetric as Yme1 i.e. has the spiral staircase.

7.4.2 Puchades et al., 2019

The C-terminus of FtsH is often predicted to be unstructured, lacks conservation and is sometimes shorter or longer (Figure 46). It was found to be autocatalytically cleaved in E. coli but with unknown purpose (Yoshinori Akiyama, 1999). As the C-terminal module is the protease, the terminus is expected to be far away from the membrane. However, already my published structures indicated that the C-terminus folds upwards but it was truncated strongly and hence not further evaluated. The last visible residues (up to position 779) in the new cryo-EM model are interacting with α0 on top of the AAA module and the linker between the AAA and protease module. The final 18 residues are neither visible nor do they effect activity in contrast to the removal of positions 750-780.

Another interesting fact is the correlation of several disease associated mutations with the 'central protrusion'. This is a subdomain on the bottom of the 'AAA-barrel' protruding into the proteolytic chamber. The loops belonging to this protrusion are often disordered in the crystal structures. They occupy the proteolytic chamber to some degree and potentially help directing the substrates to the proteolytic site.

The authors also describe a disease associated mutation of Asn432. This asparagine is interacting with Arg416. These two residues are equivalent with Asn280 and Arg261 or Arg263 in *A. aeolicus* (Figure 46). According to my model of pore-1 loop remodelling, Gln281 is actually inducing the compact conformation by disturbing the C-terminus of α2. Arg261 is found in the intersubunit signalling network while Arg263 is disordered. Arg263 is effectively part of the pore-2 loop, a feature rich in glycines interacting with the substrate in the author's models.

Appendix

8 METHODS

8.1 Site directed mutagenesis

Site directed mutagenesis primers were bought from Eurofins and PCR was performed as generalised in Table 6. 9 µl products were treated with 0.2 µl DpnI in 1 µl CutSmart buffer (NEB) for 2 h at 37 °C and subsequently transformed in 50 µl chemical competent *E. coli* DH5α cells, which were thawed and incubated on ice with 10 µl ligated vectors for 20 min. The bacteria were heat shocked at 42 °C for 2 min and incubated for 2 min on ice as well as 1 h at 37 °C with additional 300 µl SOC medium (20 g/l tryptone, 5 g/l yeast extract, 0.58 g/l NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 2.5 mM KCl, 20 mM glucose). They were streaked out on LB agar (10 g/l tryptone, 2.5 g/l yeast extract, 10 g/l NaCl, 17.5 g/l agar) containing the appropriate antibiotic (1,000x stock) and incubated o/n at 37 °C. On the next day, 5 ml cultures in LB medium (10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl) with antibiotics were inoculated and incubated o/n at 37 °C. Plasmids were isolated using the Gene Elute Plasmid Miniprep kit from Sigma Aldrich. They were eluted in 55 µl 10 mM Tris pH 8.5 and sequenced by GATC with primers T7 and pET-RP.

Table 6: PCR protocol for site directed mutagenesis and thermal cycler program.

Reagent	Final concentration	Volume
10 µM primers	1.0 µM	2.5 µl each
dNTP (each 10 mM)	0.2 mM	0.5 µl
5x Q5 Buffer	1x	2.5 µl
2 U/µl Q5 polymerase	0.04 U/µl	0.5 µl
Template	0.5 µg/µl	? µl
H2O		Ad 25 µl

Time	Temperature	
2.00 min	95 °C	
0.45 min	95 °C	
1.00 min	68 °C	17x in total
3.5 min/kb	72 °C	
10.00 min	72 °C	

8.2 Plasmids and their origins

pET22b-EcFL-FtsH-wt was synthesised by General Biosystems, Inc. pCOLADuet-1_cp6-sfGFP-ssrA and pCOLADuet-1_sfGFP-ssrA were kindly provided by Karl Schmitz from the Sauer lab. All other used plasmids (GgpS: pET28a with N-terminal Strep-tag; $FtsH_{Aa,tr}$: pET28a; $FtsH_{Aa,FL}$: pET22b) were cloned by myself or property of the lab for over a decade.

8.3 Protein production

Bacteria with pET22b were incubated in presence of 100 µg/ml ampicillin and bacteria with pET28a in presence of 50 µg/ml kanamycin.

8.3.1 Transformation and pre-culture for Expression

50 µl chemical competent E. coli cells (GgpS: BL21(DE3), $FtsH_{Aa,FL}$: C43(DE3), $FtsH_{Aa,tr}$: BL21(DE3) CodonPlus RIPL, GFP: BL21(DE3) Star), were thawed and incubated on ice for 20 min with 50–100 ng vectors containing the desired gene. Afterwards, the bacteria were heat shocked at 42 °C for 2 min and incubated for another minute on ice. About 300 µl SOC medium was added and the bacteria were further incubate between 15–60 min. Half the volume was streaked out on LB agar (10 g/l tryptone, 2.5 g/l yeast extract, 10 g/l NaCl, 17.5 g/l agar) containing the appropriate antibiotic and incubated at 37 °C.

Agar plates were stored in the fridge until evening and three colonies were picked for each pre-culture. Roughly, 30–40 ml LB medium (10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl) per 1 l expression culture were incubated in presence of the appropriate antibiotic (1,000x stock). Cultures were incubated at 180 rpm at 37 °C.

8.3.2 Expression

The standard protocol was as the following: 1 l LB medium with appropriate antibiotic (1,000x stock) and 150 µl Antifoam B in 2 l baffled flasks were inoculated to an OD_{600} of 0.1 and incubated at 37 °C at 220 rpm. 1 mM (GgpS) or 0.5 mM ($FtsH_{Aa,tr}$) IPTG was used for induction at around an OD_{600} of 0.7. GgpS was expressed at

Table 7: Purification buffers for all proteins.

Buffer were filtrated and additionally degassed before size exclusion. AC: affinity chromatography, AE anion exchange chromatography, SEC: size exclusion chromatography, Strep.: Strep–Tactin column regeneration.

Protein	Usage	NaCl [mM]	Buffer [mM]		pH	Additional agents	Elution agent [mM]		Remarks
$FtsH_{Aa,FL}$, $FtsH_{Ec,FL}$	Lysis	600	20	Tris/HCl	7.5				
	Membrane resuspension	600	20	Tris/HCl	7.5	10 % glycerol			Glycerol is added after resuspension
	AC wash	300	20	Tris/HCl	7.5	5 % glycerol, 0.2 g/l DDM	60	imidazole	DDM is added right before use
	AC elution	300	20	Tris/HCl	7.5	0.2 g/l DDM	250	imidazole	
	SEC	150	20	HEPES/NaOH	7.9	0.2 g/l DDM			
$FtsH_{Aa,tr}$	Lysis	600	20	Tris/HCl	7.5				
	Ac wash	150	20	Tris/HCl	7.5		30	imidazole	
	Ac elution	150	20	Tris/HCl	7.5		250	imidazole	
	Dialysis	40	20	Tris/HCl	7.5				
	AX-low		20	Tris/HCl	7.5				
	AX-high	1000	20	Tris/HCl	7.5				
	SEC	150	20	HEPES/NaOH	7.9				
GFP	Lysis	300	20	Tris/HCl	7.5				
	AC wash	150	20	Tris/HCl	7.5		25	imidazole	
	AC elution	150	20	Tris/HCl	7.5		250	imidazole	
	SEC	150	20	HEPES/NaOH	7.9				
GgpS	Lysis	600	20	Tris/HCl	7.5				
	AC elution	150	20	HEPES/NaOH	7.6		2	D–desthio-biotin	
	SEC_{assay}	150	30	HEPES/KOH	7.6				
	$SEC_{crystallisation}$	150	10	HEPES/NaOH	7.6				Or desired sodium salt
Strep.	Purge	150	100	Tris/HCl	8.0				
	Wash		100	Tris/HCl					Untitrated
	Storage	150	100	Tris/HCl	8.0	1 mM EDTA			

37 °C for 3 h and $FtsH_{Aa,tr}$ at 20 °C for around 20 h, GFP for 3–4. All were incubated at 220 rpm.

Pre-cultures of the $FtsH_{Aa,FL}$ were centrifuged for 10 min at 2,000 g and pellets resuspended in TB medium (20 g/l tryptone, 24 g/l yeast extract, 4.6 ml/l 87 % glycerol). TB medium with phosphate buffer and 8 mM $MgCl_2$, 100 µg/ml ampicillin and 133 µl Antifoam B were inoculated with resuspended pre-culture to an OD_{600} of 0.1 and incubated for a bit more than 2 h at 37 °C and 260 rpm in 2 l baffled flasks. Expression was induced with 1 mM IPTG, additional 100 µg/ml ampicillin and 133 µl Antifoam B.

All bacteria were harvested by centrifugation at 3,200 *g*, 4 °C for 20 min. Lysis-1 buffer was used for resuspension and transfer into 50 ml tubes, usually one falcon per culture. 1 l FtsH$_{Aa,FL}$ yielded around 20 g pellet (GgpS 1.8 g, GFP 4.5 g) and was divided into 4—5 g pellets for easier resuspension. FtsH$_{Aa,FL}$ pellets were washed again with lysis buffer and all pellets were stored at -80 °C.

8.3.3   Lysis

All steps were carried out at 4 °C or on ice.

All bacteria were lysed in 6 ml/g (FtsH) or 4 ml/g (GgpS, GFP) lysis buffer but at least 20 ml with additional 10 µg/ml DNase I (1,000x stock) using a cell disruptor at 2.5 kbar twice. Cell debris was removed at 20,000 *g* for 20 min.

8.3.4   Purification (excluding full-length FtsH)

All steps were carried out at 4 °C or on ice for GFP and FtsH$_{Aa,tr}$ but at room temperature for GgpS.

0.5 ml (GFP) or 1 ml Ni-NTA (FtsH$_{Aa,tr}$) Ni-NTA resin per gram bacterial pellet were equilibrated with lysis buffer and subsequently washed and eluted with imidazole containing buffer. Flow rate was set to 1–1.5 ml/min.

For GgpS, a small scale purification utilised 0.5 ml streptactin resin in total while a large scale 2 ml. It was equilibrated with lysis buffer and subsequently washed again before and eluting with D-desthiobiotin containing buffer. Flow rate was set to 2 ml/min for all steps except from elution with 1 ml/min.

1–2 µl of all 1 ml elution fractions were blotted in filter paper and proteins visualised with coomassie brilliant blue as used for SDS-PAGE gel staining. The paper was briefly washed in water and destaining solution. Fractions were pooled accordingly.

FtsH$_{Aa,tr}$ pools were dialysed overnight against 100x excess of anion exchange loading buffer in a 3,500 kDa dialysis tubing (Spectra/Por 3, Spectrumlabs). The dialysed protein was applied to a HiTrap Q FF column (5 ml, GE) equilibrated with 4 % AX-high against AX-low. The elution was carried out with a linear gradient

over 10 CV to 25 % AX-high. 5 CV 100 % AX-high were used to purge the column free of potentially stronger binding proteins. Fractions were pooled according to chromatogram.

The proteins were concentrated (GgpS 100 kDa, FtsH$_{Aa,tr}$ 30 kDa and GFP 10 kDa molecular weight cut-off, Amicon) to a volume depending on the size exclusion chromatography column, usually 100–150 µl for a 24 ml column and 1–2 ml for a 120 ml column. The proteins were always centrifuged after concentrating for 10 min at 21,000 g. Size exclusion chromatographies were performed with a Superdex 200 10/300 Increase (GgpS), Superdex 200 16/60 (FtsH$_{Aa,tr}$) or Superdex 75 16/60 (GFP) and an ÄKTA Prime, Prime plus or Purifier. Columns and ÄKTA were from GE. The flow rate was set to 0.5 ml/min (GgpS) or 1 ml/min (FtsH$_{Aa,tr}$, GFP). Fractions were pooled according to chromatogram and concentrated. The proteins were always centrifuged after wards for 10 min at 21,000 g. Concentration of all FtsH proteins was evaluated via BCA assay and for GgpS and GFP spectrophotometrically via NanoDrop (Thermo Scientific) and theoretical absorption coefficient.

8.3.5   Membrane preparation of full-length FtsH

All steps were carried out at 4 °C or on ice.

Supernatant after lysis were centrifuged at 257,000 g for 1h. Membranes were resuspended in 3.54 ml/mg resuspension buffer without glycerol using a 5 ml syringe with a blunt cannula. The 0.46 ml 87 % glycerol was added after resuspension and the concentration was meausred via BCA assay. Membranes were adjusted to 10 mg/ml protein with Lysis-2 and glycerol buffer and stored at -80 °C in 10 ml or 20 ml aliquots.

8.3.6   Purification of full-length FtsH

All steps were carried out at 4 °C or on ice.

20 ml membranes were thawed in cold water and solubilised with 10 g/l DDM in SEC buffer (1,000x stock) while rotating 1 h. Membranes were diluted with 20 ml wash buffer containing 20 mM imidazole and were subsequently centrifuged at 20,000 g

for 30–60 min to remove non solubilised proteins. The supernatant was applied to 6 ml wash buffer (10 mM imidazole) equilibrated Ni-NTA resin, washed (60 mM imidazole) and subsequently eluted. All affinity chromatography steps were carried out at 1 ml/min. The protein was subsequently concentrated (100 kDa molecular weight cut-off, Amicon) at 4,000 g in steps of less than 10 min to around 15–20 mg/ml and subsequently centrifuged for 10 min at 21,000 g. The protein was transferred to Gayathri Jeyasankar for crypo-EM applications and usually purified via HPLC-SEC. For own usage, the protein was applied to a Superose 6 10/300 size exclusion chromatography at 0.5 ml/min, followed by concentration as desired. Concentration were estimated by BCA assay and a rough guidance was that the raw absorption at 280 nm equals BCA estimate.

8.3.7 BCA-assay

Bicinchoninic acid (BCA) assay was performed as described by the manufacturer (Pierce BCA Protein Assay Kit, Thermo Scientific). Protein and working reagent (10 µl and 200 µl) were incubated in 1.5 ml reaction tubes at 60 °C for 30 min, pipetted into a 96-well microtiter plate (flat bottom and transparent, Brand (781602)) and read with a BioTek Synergy HT at 562 nm. A 20 mg/ml concentrated BSA (Sigma Aldrich) standard was diluted to 0.4 mg/ml and used for start point for a calibration series in 0.04 mg/ml increments. Calibration standards were always normalised to the effective concentration of the 0.4 mg/ml standard as estimated with a NanoDrop (Thermo Scientific) in BSA mode.

8.3.8 SDS-PAGE

0.6 ml *OD600 culture were centrifuged at 21,000 g for 5 min and the pellet was frozen. After thawing lysis was done by sonication in 200 µl lysis buffer for 16 s in total with 2 s 30 % pulses and 4 s pauses. The 1.5 ml reaction tube was always kept in ice with surrounding water for better heat dissipation. 40 µl samples were taken before and after 10 min centrifugation at 21,000 g at 4 °C.

Table 8: SDS–PAGE gel composition.

Reagent	Stacking gel	Separation gel 10 %	Separation gel 12 %
Water		4.17 ml	3.50 ml
LGB†		2.50 ml	2.50 ml
Stacking gel solution‡	4.0 ml		
30 % acrylamide with 0.8 % bisacrylamide		3.33 ml	4.00 ml
2,2,2-trichlorethanol		50 µl	50 µl
TEMED	8 µl	10 µl	10 µl
10 % ammonium persulphate	20 µl	50 µl	50 µl

†: 182 g/l Tris pH 8.8, 4 g/l SDS; ‡: 7 ml 30 % acrylamid/0.8 % bisacrylamid, 30 µl 2 % bromphenolblue, 2.5 ml water, 12.5 ml 60.5 g/l Tris pH 6.8, 4 g/l SDS

All purification samples were usually diluted 1/10 with appropriate buffer before further diluting 1/5 with loading buffer (0,71 mg/ml sodium dodecyl sulphate, 10 mM Tris pH 6.8, 13,3 % glycerol, 126 mM β-mercaptoethanol and some bromphenolblue). The dilution was adjusted so usually 10 µl (3 µg purified protein) samples were loaded into gel slots. The samples were incubated for 5 min at 95 °C or 10 min at 42 °C (full-length FtsH) and briefly centrifuged. The Pierce Unstained Protein MW Marker was used as standard. SDS polyacrylamide gel electrophoresis (SDS-PAGE) gels were prepared as described in Table 8. Each gel part was allowed to polymerise for at least 15 min. The gels were stored at 4 °C and run at 30–40 mA per gel in 3.02 g/l Tris, 14.4 g/l glycine (pH 8.3) and 0.5 g/l SDS for 40–65 min. Afterwards, the gels were stained in 2.5 g/l Coomassie Brilliant Blue R 250, 30 % ethanol, 10 % ethanoic acid and heated for 1 min in a microwave oven. After 20–30 min, the gels were destained in 30 % ethanol and 10 % ethanoic acid and images were eventually taken by using a ChemiDoc XRS+ from Bio-Rad. Due to the addition of 2,2,2-trichlorethanol images after activation with UV light could be taken directly before staining in coomassie.

8.3.9   Western blot

SDS-PAGE gels were blotted via wet-blot method. The gel-methanol activated PVDF membrane combo was sandwiched between around 4 mm filter papers and transferred at 400 mA, 100 V for 1 h at 4 °C in 3.09 g/l boric acid and 10 % methanol. The

membrane was blocked afterwards with 5 % skim-milk powder in TBST (150 mM NaCl, 20 mM Tris/HCl pH 7.5, 0.05 % Tween-20) for 1 h at room temperature. For His-tag detection, 1/2,000 α-penta His HRP (Miltenyi Biotec) in 5 % skim-milk powder TBST was used. Strep-tag was labelled with Streptavidin conjugated horseradish peroxidase (IBA) *after* washing with TBST thrice for 5 min. The labelling compound was washed away with TBST thrice for 5 min and detection was carried out by addition of 10 ml 100 mM Tris/HCl pH 8.8 with 25 µl 90 mm p-coumaric acid (in DMSO), 50 µl 250 mM luminol (in DMSO) and 60 µl 30 % hydrogen peroxide in a ChemiDoc XRS+ from Bio-Rad.

8.4 Assays

8.4.1 GgpS – NHS-Cy5 labelling

GgpS was labelled with the Sulfo-Cyanin 5 NHS-Ester kit from Lumiprobe. 4.2 nmol lyophilised coupling reagent were added with 1.05 nmol tetramers (4:1 ratio). The protein was concentrated/diluted to around 30 µM in SEC$_{assay}$ buffer. The pH was adjusted to pH 8.30 as recommended by the manufacturer with 0.1 mM NaOH. A usual final reaction volume was 70–100 µl and had to be incubated overnight at 23 °C. 3 h were not enough. Unbound fluorophore was removed as advised with SEC$_{assay}$ buffer.

8.4.2 GgpS – Two-K_d model for interaction and activity fit

A biphasic mode for interaction and activity can be explained by different responses of the three possible states of GgpS: (i) empty, (ii) one chloride site occupied, (iii) both sites occupied. Therefore, a model explaining the observed responses must accommodate how much of each state is present simultaneously in this three-state equilibrium. I developed a simple model based on the law of mass action assuming each degree of occupation coincides with only one state/conformation and transition only occurs obligatory upon binding or release. This model comes naturally with one dissociation constant (K_d) per transition or one per binding site. Since both binding sites are equal in the fully chloride bound crystal structure, I would expect $K_{d_1} = K_{d_2}$.

However, the asymmetry of the W110P mutant may not imply this directly but gives a basis for an interaction of both sites and hence, the first binding event might influence the second site and thereby increases or decreases its affinity. This is the definition of a cooperative binding mode and would yield $K_{d_1} \neq K_{d_2}$. Therefore, this model is not restricted to a cooperative or non-cooperative binding.

The K_ds are defined in (3) and (4), with $[A^-]$ and $[GgpS_x]$ as the actual free anion and GgpS tetramer concentrations. The index x further defines whether total GgpS or empty, half and fully occupied by anions is meant.

$$K_{d_1} = \frac{[A^-] * [GgpS_{empty}]}{[GgpS_{half}]} \tag{3}$$

$$K_{d_2} = \frac{[A^-] * [GgpS_{half}]}{[GgpS_{full}]} \tag{4}$$

The total concentration of GgpS tetramers equals the sum of empty, half and fully occupied tetramers:

$$[GgpS_{total}] = [GgpS_{empty}] + [GgpS_{half}] + [GgpS_{full}] \tag{5}$$

Thus, the fraction of empty tetramers is defined as:

$$\frac{[GgpS_{empty}]}{[GgpS_{total}]} = \frac{[GgpS_{empty}]}{[GgpS_{empty}] + [GgpS_{half}] + [GgpS_{full}]} \tag{6}$$

Eventually, the total concentration of the anion $[A^-_0]$ is the sum of free and bound anions:

$$[A^-_0] = [A^-] + [GgpS_{half}] + 2 * [GgpS_{full}] = [A^-] + [A^-_{bound}] \tag{7}$$

However, since 8 nM or 12 nM concentrations of tetramers have been used, $[A^-]$ can only be as small as twice that amount below $[A^-_0]$. The lowest amount of added anions is 400 µM – four orders

of magnitudes more. Therefore, we assume the concentrations of free anions to be identical to the total concentration (8) and conclude (9).

$$[A^-{}_{bound}] \ll [A^-] \text{ and } [A^-{}_{bound}] \cong 0 \qquad (8)$$
$$[A^-] \cong [A^-{}_0] \qquad (9)$$

Transforming and inserting equation (3) for the concentration of half-occupied and equations (3) and (4) for the concentration of fully occupied GgpS yields equation (10).

$$\frac{[GgpS_{empty}]}{[GgpS_{total}]} = \frac{[GgpS_{empty}]}{[GgpS_{empty}] + \frac{[A^-] * [GgpS_{empty}]}{K_{d_1}} + \frac{[A^-]^2 * [GgpS_{empty}]}{K_{d_1} * K_{d_2}}} \qquad (10)$$

Taking into account (9) and further simplifying, gives the fraction of empty over total GgpS tetramers in a biphasic two K_d dependent manner.

$$\frac{[GgpS_{empty}]}{[GgpS_{total}]} = \frac{1}{1 + \frac{[A^-{}_0]}{K_{d_1}} + \frac{[A^-{}_0]^2}{K_{d_1} * K_{d_2}}} \qquad (11)$$

Likewise, equations (12) and (13) represent the relative amount of half and fully occupied GgpS tetramers.

$$\frac{[GgpS_{half}]}{[GgpS_{total}]} = \frac{1}{1 + \frac{K_{d_1}}{[A^-{}_0]} + \frac{[A^-{}_0]}{K_{d_2}}} \qquad (12)$$

$$\frac{[GgpS_{full}]}{[GgpS_{total}]} = \frac{1}{1 + \frac{K_{d_1} * K_{d_2}}{[A^-{}_0]^2} + \frac{K_{d_2}}{[A^-{}_0]}} \qquad (13)$$

The final dose-response model is the sum of each state multiplied by the response factors a, b and c, respectively.

$$Response = a * \frac{[GgpS_{empty}]}{[GgpS_{total}]} + b * \frac{[GgpS_{half}]}{[GgpS_{total}]} + c * \frac{[GgpS_{full}]}{[GgpS_{total}]} \quad (14)$$

In contrast to chloride and bromide, binding of acetate and sulphate exhibits only a monophasic response. A second binding event may still be possible, but does not impede the K_d estimation with a simple one-K_d model. Equivalent to the two-K_d model, we derive (17) as the dose-response approximation for acetate and sulphate.

$$\frac{[GgpS_{empty}]}{[GgpS_{total}]} = \frac{1}{1 + \frac{[A^-_0]}{K_d}} \quad (15)$$

$$\frac{[GgpS_{half}]}{[GgpS_{total}]} = \frac{1}{1 + \frac{K_d}{[A^-_0]}} \quad (16)$$

$$Response = a * \frac{[GgpS_{empty}]}{[GgpS_{total}]} + b * \frac{[GgpS_{half}]}{[GgpS_{total}]} \quad (17)$$

8.4.3   GgpS – Activity

The concentration of all four salts were chosen based on equal ionic strengths. Each replicate was the average of two measurement of same protein batch and master mix. 40 µl total volume were measured in 384-well pureGrade transparent PS plates with F bottom (Fisher Scientific) in a Synergy H4 (BioTek) plate reader at 23 °C. 20 µl master mix was added to each well and the plate was shaken 'fast' for 10 s to allow spreading of the entire bottom. The reaction was started by adding 20 µl 60 nM protein (1 mM NaCl, 30 mM HEPES/KOH pH 7.6) using the dispenser of the plate reader followed by 30 s 'fast' shaking for mixing. The decreasing NADH concentration was measured for about 2 h at 320 nm with an extinction coefficient of 5.037 mM^{-1}cm^{-1}. Most assay components were mixed in a pre-master mix which was then split and further supplemented with a sodium salt, G3P and DNA. Water was added instead of G3P for blank samples as well as DNA-free samples. A complete measurement series included DNA-free and DNA containing samples. Final assay concentration are listed

Table 9: GgpS activity assay composition and final concentrations in the well.

Ingredient	Concentration	
Na$_2$–NADH	1.8	mM
Na$_3$–PEP	2.0	mM
Na$_2$-ADP-glucose	4.0	mM
Li$_1$-G3P	4.0	mM
Mg-acetate	4.0	mM
Lactate dehydrogenase	32.0	U/ml
Pyruvate kinase	12.0	U/ml
Sheared herring sperm DNA	1.25	µg/µl
HEPES/KOH pH 7.6	30.0	mM
GgpS	30.0	nM
NaCl	0.5	mM

in Table 9 and estimated total ion concentrations in Figure 26. HEPES was titrated with KOH to supply K$^+$ for pyruvate kinase (Kachmar & Boyer, 1953). This enzyme requires Mg^{2+} in addition for higher activity (Baek & Nowak, 1982; Larsen et al., 1997).

Fits were carried out with *nls.multstar* package in R (Padfield & Matheson, 2020; R Core Team, 2020).. Parameter were constrained if applicable as follows:

- $a = \begin{cases} [0, Inf] & \text{if aceate or not normalised before averaging} \\ 1 & \text{if other anion and normalised before averaging} \end{cases}$
- $b = c = K_{d_1} = K_{d_2} = [0, Inf]$

8.4.4   GgpS – Fluorescence quenching

Samples were measured in a Monolith NT.115 (Nanotemper) with standard capillaries similar to the manufacturer's protocol. A serial dilution (usually 2/3) of the ligand solution (sodium salt as desired, 30 mM HEPES/KOH pH 7.6, 0.05 % Tween-20) was pipetted to gain concentrations as desired. Eventually 3 parts of ligand were mixed with 2 parts of target (20 nm Cy5 coupled protein, 30 mM HEPES/KOH pH 7.6, 0.5 mM NaCl). This generalised protocol was used for the W110P mutant. Exact ratios varied for the wild type slightly as the assay is not refined yet.

Fits were carried out with *nls.multstar* package in R (Padfield & Matheson, 2020; R Core Team, 2020). All parameters were constrained to $[0, Inf]$.

8.4.5 FtsH – ATP hydrolysis

Table 10: FtsH ATP hydrolysis activity assay composition and final concentrations in the well.

Ingredient	Concentration	
Na$_2$–NADH	1.75	mM
Na$_3$–PEP	4.00	mM
Na$_3$–ATP	4.00	mM
MgCl$_2$	4.00	mM
HEPES/NaOH pH 7.9	20.00	mM
NaCl	150.00	mM
Lactate dehydrogenase	10.00	U/ml
Pyruvate kinase	10.00	U/ml
FtsH hexamers	200.00	nM
DDM	0.20	g/l

Each protein was measured in technical triplicates with same protein batch and master mix. 40 µl total volume were measured in 384-well pureGrade transparent PS plates with F bottom (Fisher Scientific) in a Synergy H4 (BioTek) plate reader at 37 °C for 3 h. 20 µl 0.2 µM protein was added to each well and the reaction was started by addition of 20 µl master mix. The decreasing NADH concentration was measured for about 2 h at 320 nm with an extinction coefficient of 5.037 mM^{-1}cm^{-1} and evaporation was reduced by sealing with 'Clear Heat Seal' (4titude). Most assay components were mixed in a pre-master mix which was then split and further supplemented with ATP or water for blanks. Final assay concentration are listed in Table 10.

8.4.6 FtsH – Bodipy TR-X casein translocation

Lyophilised Bodipy TR-X casein (EnzCheck, Thermo Fisher Scientific) was solved in 200 µl 100 mM NaHCO$_3$ pH 8.3 for a couple of hours and diluted with 800 µl volume so eventually the substrate was in SEC buffer. Each protein was measured in technical triplicates with same protein batch and master mix. 40 µl total volume were measured in 384-well pureGrade black PS plates with F bottom (Fisher Scientific) in a Synergy H4 (BioTek) plate reader at 37 °C for 2 h. 20 µl 0.6 µM protein was added to each well and the reaction was started by addition of 20 µl master mix. Emission was measured at 617 nm with a 13.5 nm band pass filter

Table 11: FtsH casein translocation assay composition and final concentrations in the well.

Ingredient	Concentration	
Bodipy TR-X casein	15.0	µg/ml
Creatine phosphate	16.0	mM
Na$_3$-ATP	4.0	mM
MgCl$_2$	4.0	mM
HEPES/NaOH pH 7.9	20.0	mM
NaCl	150.0	mM
Creatine phosphokinase	20.0	U/ml
FtsH hexamers	300.0	nM
DDM	0.2	g/l

and a detector gain of 102 while the fluorophore was excited at 589 nm with a 9 nm band pass filter. Evaporation was reduced by sealing with 'Clear Heat Seal' (4titude). Most assay components were mixed in a pre-master mix which was then split and further supplemented with ATP or water for blanks. Final assay concentration are listed in Table 11.

8.4.7 FtsH – GFP unfolding

The GFP assay was derived from the Bodiy TR-X casein assay. Instead of casein 1 µM destabilised GFP mutant utilised. The blank sample was without FtsH. The GFP was excited at both possible wavelengths of 430 nm or 488 nm and emission were detected at 512 nm. A bandpass filter of 9 nm was used everywhere. The detector gain was set to around 90. No default protocol was established because GFP was not degraded. FtsH concentrations between 1 µM and 0.1 µM were analysed. Other assay components were as in Table 11.

8.4.8 Estimation of DDM concentration

The DDM concentration was measured according to Urbani & Warne, 2005. 50 µl sample and 250 µl 5 % phenol (w/v) in water were transferred into a 2 ml polypropylene tube with tight lock. 600 µl of 100 % sulphuric acid were added and the tube was closed immediately as the exothermic reaction releases a lot of heat. The absorbance was measured at 490 nm after the sampled cooled down. The assay is valid for concentrations up to 0.4 g/l and potentially more.

Dynamic light scattering was performed with a DynaPro NanoStar from Wyatt Technology using 60 µl single-use 'UVettes' from Eppendorf at 25 °C. The protein was diluted to 0.5 mg/ml and centrifuge for 5 min at 21,000 g. A refractive index of 1.33, a viscosity of 0.906 cp and a globular protein was assumed. A measurement consisted of ten acquisitions of 10 s length. Data plotted are from cumulant analysis, but Inverse Laplace Transform regularisation was checked for bi- or oligomodality.

8.5 Crystallography

8.5.1   Crystallisation

Final crystallisation conditions are listed in Table 12 and Table 13.

Sitting drop initial crystallisation screens with drop ratios of protein:reservoir of 100:200 nl, 150:150 nl and 200:100 nl (50 µl reservoir in SwissCi 96 well 3-drop plate) were set up using a Mosquito (SPT Labtech). The protein was usually concentrated to 10 mg/ml. Used screen: JCSG-plus, Midas HT96, Morpheus, PACT premier HT96, Proplex HT96, Stura Footprint Combination HT-96 (Molecular Dynamics), Index HT, Natrix, PEGIon HT, Salt RX (Hampton Research), Wizard 1 and 2 (Rigaku Reagents) and Sigma for Complexes (Sigma Aldrich). Finer grid screen were employed around the initial hit by varying main precipitant and buffer pH in drop ratios of 1:2 µl, 1.5:1.5 µl and 2:1 µl. 48-well MRC Maxi plates (SwissCi) with 100 µl reservoir volume were used.

Sodium chloride in the $SEC_{crystallisation}$ buffer was changed to the appropriate sodium salt for crystallisation of GgpS in sulphate, acetate or bromide. For bromide each buffer during the whole purification was changed. Some screens contained hydrochloric acid titrated buffers and malonic acid was used instead.

GgpS crystals appeared within one day but usually as clusters. They were harvested and used for microseeding using the Beads-for-Seeds kit (Jena Bioscience) and eventually 0.25 µl of 10^{-3}–10^{-5} dilutions into so far empty drops. FtsH crystals grew sometimes

within 20 min. The crystals used for the published dataset with space group R32 were harvest one year after set up and incubation at 4 °C. All subsequent experiment were started in the cold room. The mutants were microseeded similarily to GgpS with mediocre results.

All GgpS crystallisation conditions turned out to require further cryo-protection. Increasing the precipitant concentration and supplementation with polyethyleneglycol 400 failed due to rapid phase separation. Most crystals were eventually cryo-protected in a 50 % saturated sucrose solution made from reservoir solution or reservoir solution of an adjacent condition. If sucrose was undesired, Paratone N (wt_{Cl} and $wt_{Cl,G3P}$), Paraffin ($wt_{Cl,ADP}$), and crystallisation buffer supplemented with 20 % glycerol ($wt_{SO4,GOL}$) were used.

Crystals were mounted using 18 mm Mounted CryoLoops (Hampton Research) or Micro Loops (MiTeGen) in various sizes models.

8.5.2 Data collection and processing

Datasets were collected at beamlines ID23-1 (ESRF, Grenoble, France), P13 (PETRA, Hamburg, Germany), and PXI as well as PXIII (SLS, Villigen, Switzerland). We thank all the staff involved of beamline supervision, maintenance and organisation. Data were indexed, integrated and scaled with XDS (Kabsch, 2010). Data reduction strategy was guided by statistics from aimless (P. R. Evans & Murshudov, 2013) and pointless (P. Evans, 2006) as implemented in CCP4 (Winn et al., 2011), as well as xdsstat (Diederichs, n.d.) and xdscc12 (Assmann et al., 2016). Datasets were truncated so that the $CC_{1/2}$ in their highest resolution shell (CORRECT step in XDS) was close to 0.6 (GgpS) or I/σ_I close to 2 (FtsH). XSCALE was used to merge datasets and/or Friedel pairs. Phaser (McCoy et al., 2007), phenix.refine (Afonine et al., 2012), xtriage and feature enhanced maps (Afonine et al., 2015) from Phenix (Adams et al., 2010) were used to further process and evaluate the datasets. Models were build and modified manually in Coot (Emsley et al., 2010). Visualisation was done using UCSF

Chimera (Pettersen et al., 2004) with Polder maps from Phenix (Liebschner et al., 2017) for ligands.

The first GgpS dataset was phased by molecular replacement with coordinates from PDB entry 2wtx, a homologous GT20 glucosyltransferase with 31 % sequence identity (Errey et al., 2010). The first FtsH was phased by molecular replacement with coordinates from PDB entry 4ww0 (Vostrukhina et al., 2015). Subsequent datasets were phased by molecular replacement with the best model to date, or by rigid body phasing with subsequent randomisation of the coordinates by 0.2 Å and setting all B factors to the dataset's Wilson B.

9 SUPPLEMENTAL INFORMATION
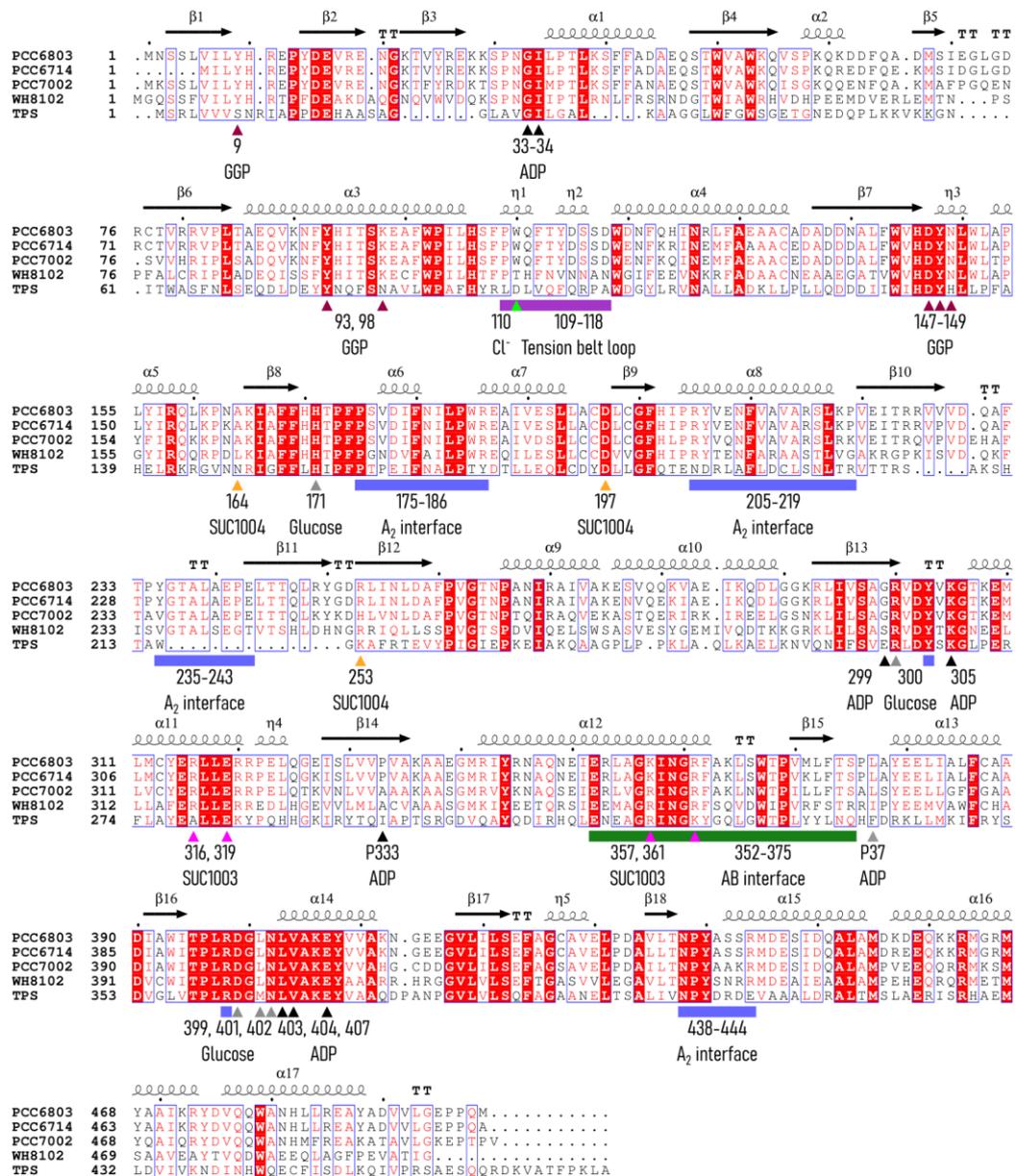
Figure 69: Annotated alignment of GgpS from Synechocytis sp. PCC6803, PCC6714; Synecococcus sp. PCC7002, WH8102 and TPS from E. coli.

Secondary structures are taken from chain A of $wt_{Br,SUC}$. 'A$_2$ interface' refers to the interface between both monomers forming the chloride bridge. The region around the chloride interactor Trp110 is part of that interface as well. 'AB interface' refers to the interface between the dimer usually found in the asymmetric unit.
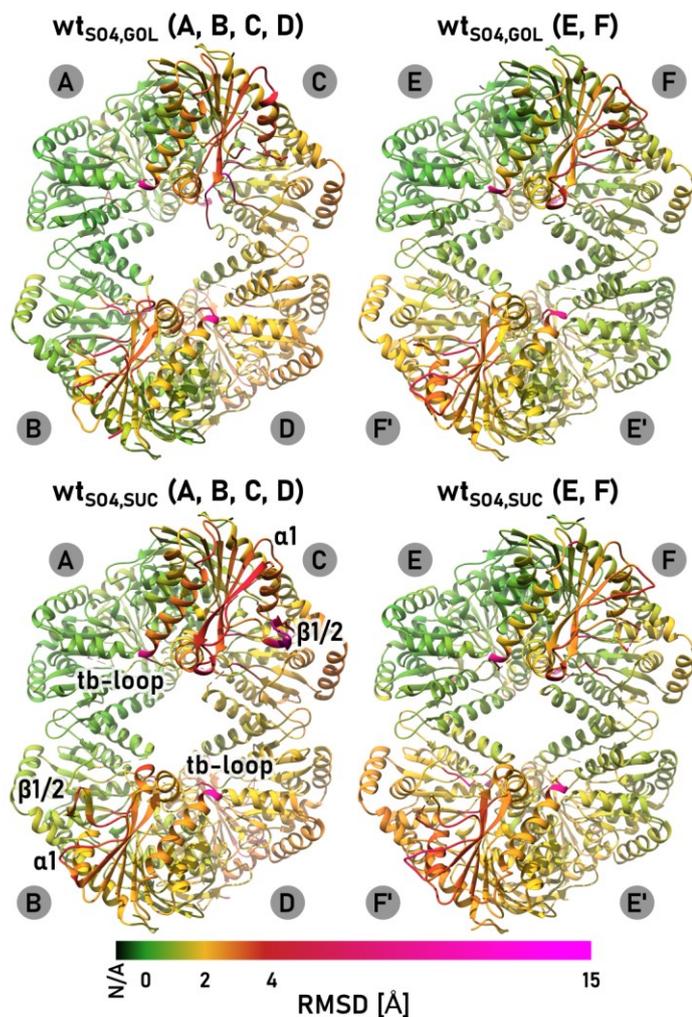
Figure 70: Deviations of the tetramers found in the crystals grown in presence of sulphate from the two-bromide bound state.

All tetramers were aligned to chain A (always top left) from the $wt_{Br,SUC}$ model. The RMSD of the mainchain atoms of each residue of each model from the equivalent chain in the $wt_{Br,SUC}$ model is visualised, e.g. deviation between $wt_{SO4,SUC}$-C and $wt_{Br,SUC}$-A'. Grey circles identify the subunit. Only the $wt_{SO4,SUC}$ (A, B, C, D) tetramer is annotated. β1/2 is usually missing as most of the tb-loop except from α4 are. α1 is present everywhere. The dimer around one chloride site consists of subunits AC, BD, and EF.
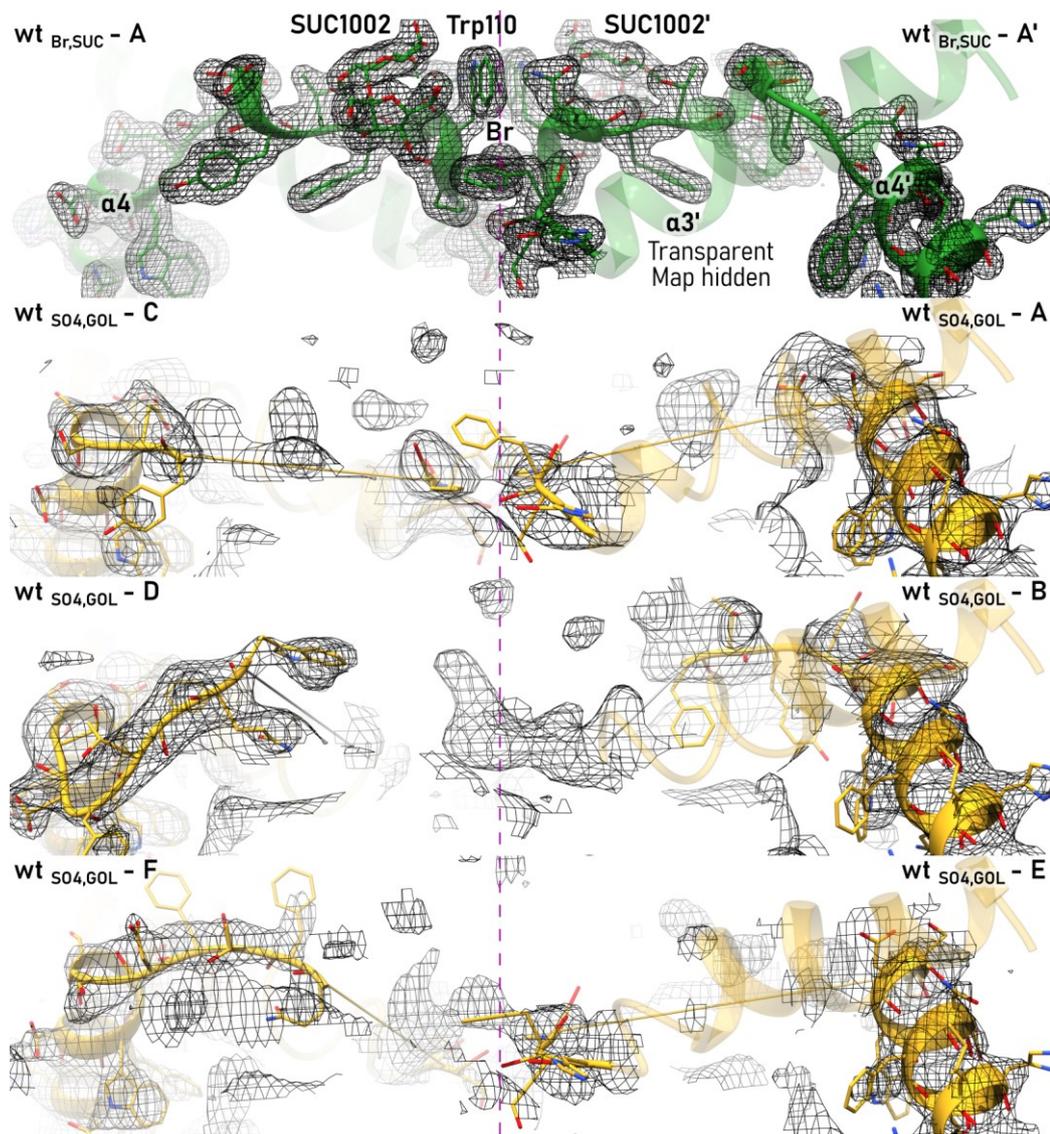
Figure 71: Tension belt loop with $2mF_o$–DF_c map of $wt_{SO4,GOL}$.

The map is drawn at 1σ RMSD up to 1.9 Å around the $wt_{Br,SUC}$ model (top). The map for the three different chloride sites of the two tetramers found in the $wt_{SO4,GOL}$ model are visualised. Map boundaries are 3.5 Å away from the coordinates of the aligned $wt_{Br,SUC}$, or $W110P_{Cl,SUC}$ models. The expected position of a bound sulphate molecule is indicated by a dashed magenta line. Some density can belong to residues of the C–terminus of α3. This helix is drawn transparent as reference. The map around the transparent part was hidden, visible electron density is located behind the helix.
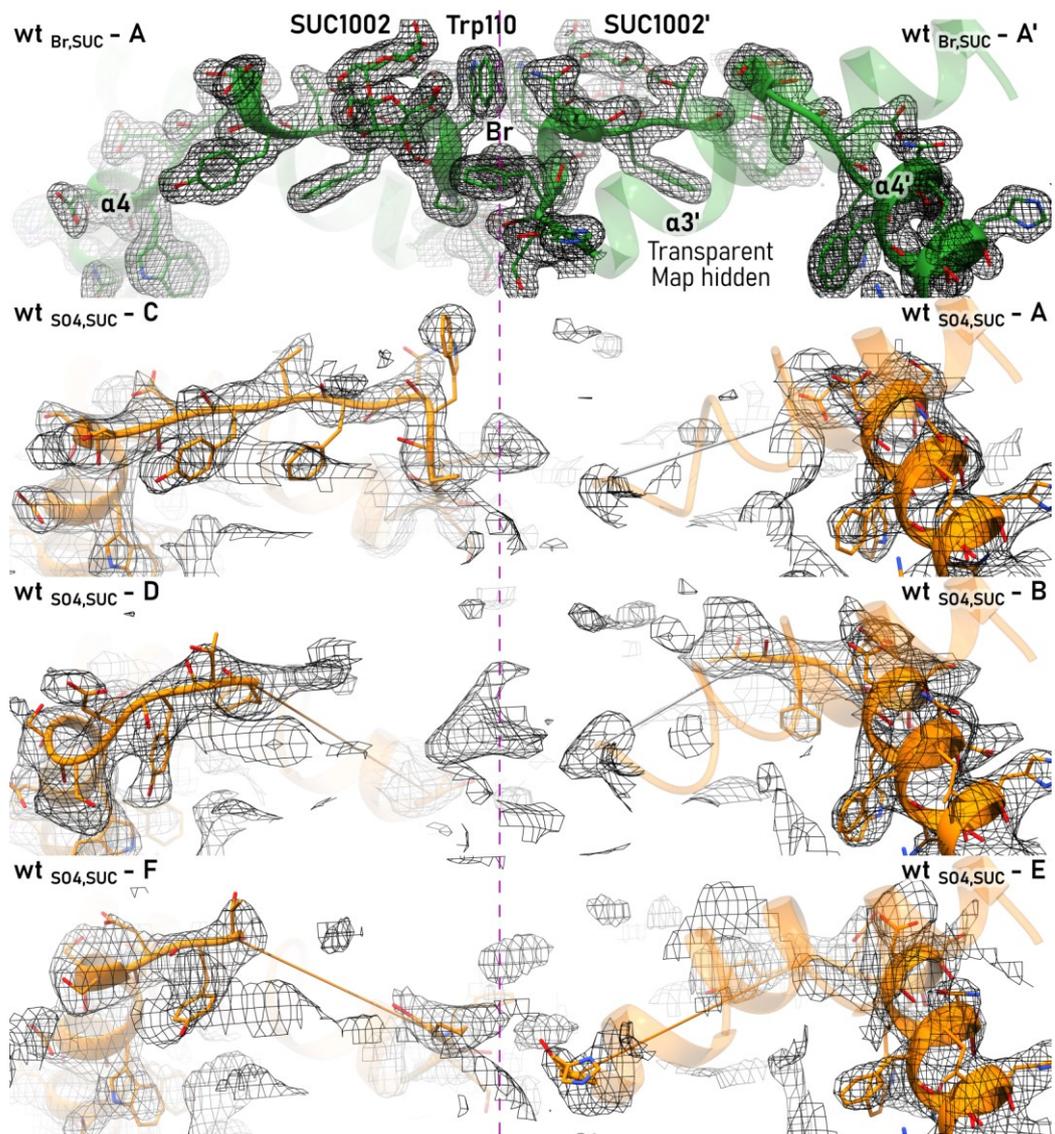
Figure 72: Tension belt loop with $2mF_o–DF_c$ map of $wt_{SO4,SUC}$.

The map is drawn at 1σ RMSD up to 1.9 Å around the $wt_{Br,SUC}$ model (top). The map for the three different chloride sites of the two tetramers found in the $wt_{SO4,SUC}$ model are visualised. Map boundaries are 3.5 Å away from the coordinates of the aligned $wt_{Br,SUC}$, or $W110P_{Cl,SUC}$ models. The expected position of a bound sulphate molecule is indicated by a dashed magenta line. Some density can belong to residues of the C-terminus of α3. The helix is drawn transparent as reference. The map around the transparent part was hidden, visible electron density is located behind the helix.

Foreground

Background

Figure 73: α3 and the active site loop around position 300 of wt$_{SO4,GOL}$.

As Figure 21A,B. The dimer around one chloride site consists of subunits AC, BD, and EF.

Foreground

A **B** **C**

D **E** **F**

Background

A **B** **E**

C **D** **F**
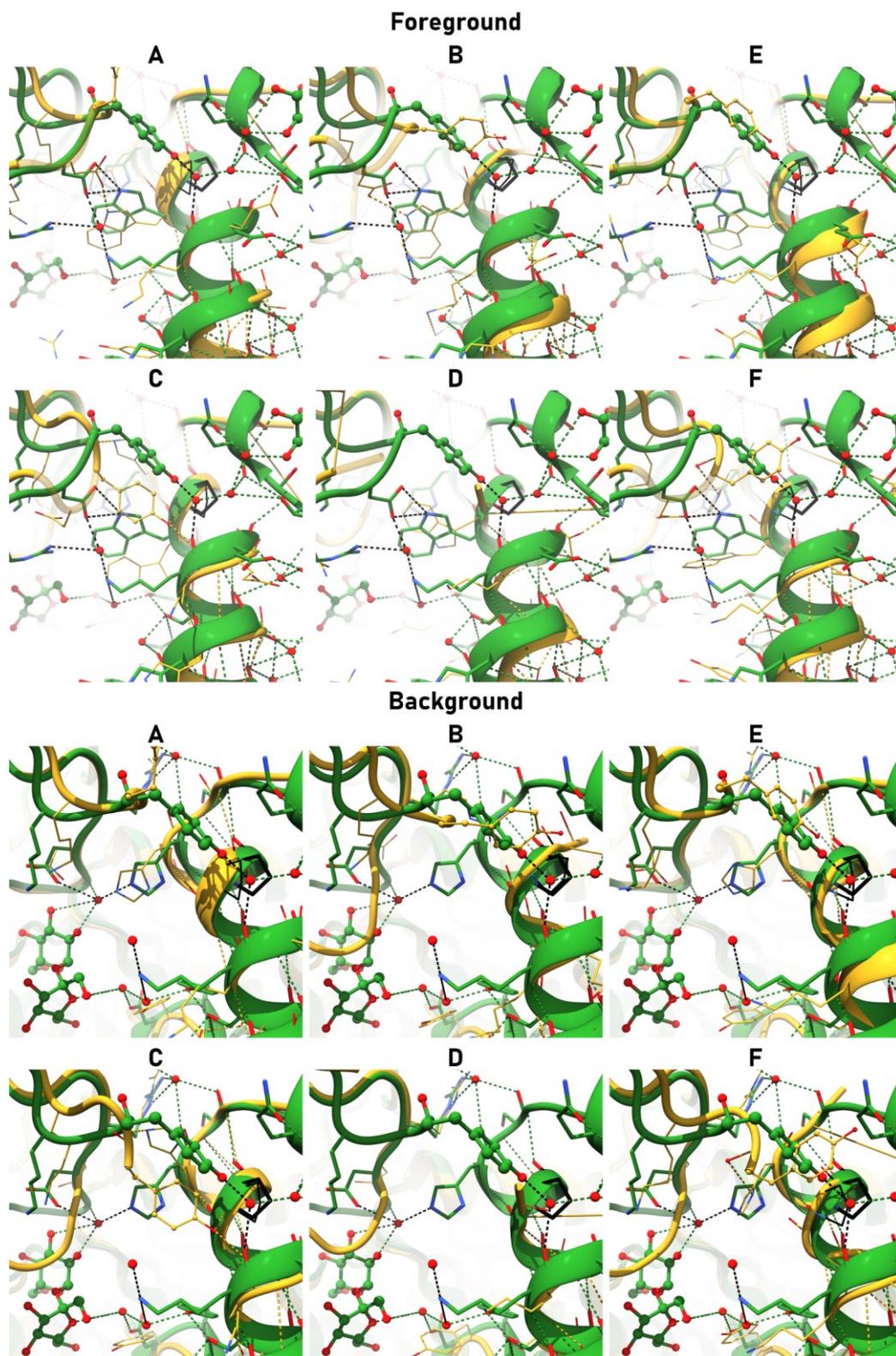
Figure 74: α3 and the active site loop around position 300 of wt$_{SO4,SUC}$.

As Figure 21A,B. The dimer around one chloride site consists of subunits AC, BD, and EF.

Figure 75: The active site loops of wt$_{SO4,GOL}$.

As Figure 21C,D. The dimer around one chloride site consists of subunits AC, BD, and EF. Background perspective: The loop preceding α12 is often disordered (straight thin tube on the left) or for example located at the bottom left in chain C—a different conformation as in the reference wt$_{Br,SUC}$.

Foreground

A   B   E

C   D   F

Background

A   B   E

C   D   F
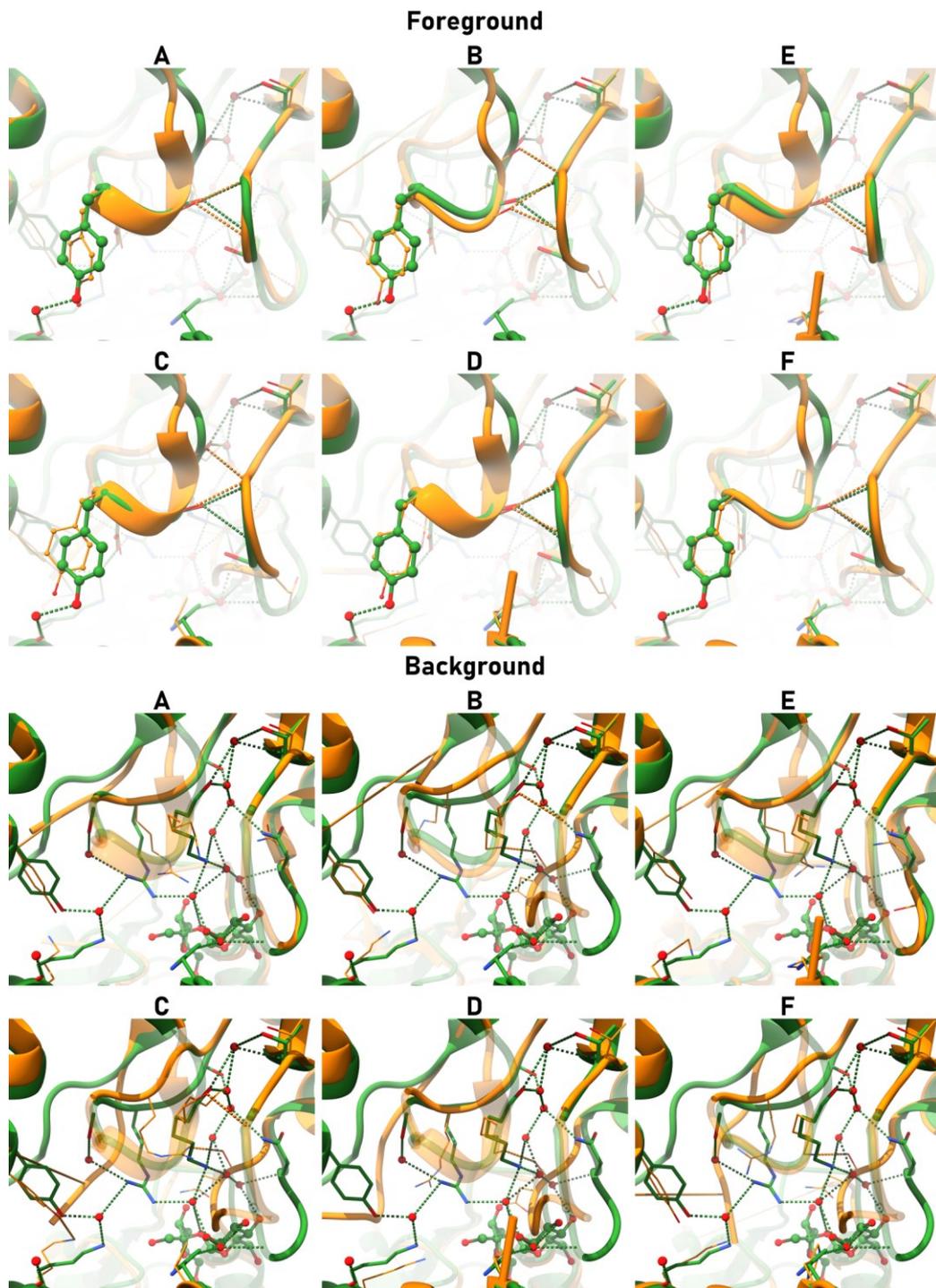
Figure 76: The active site loops of wt$_{SO4,SUC}$.

As Figure 21C,D. The dimer around one chloride site consists of subunits AC, BD, and EF. The loop preceding α12 is often disordered (i.e. straight thin tube). It is in background perspective images found on the left in the background.
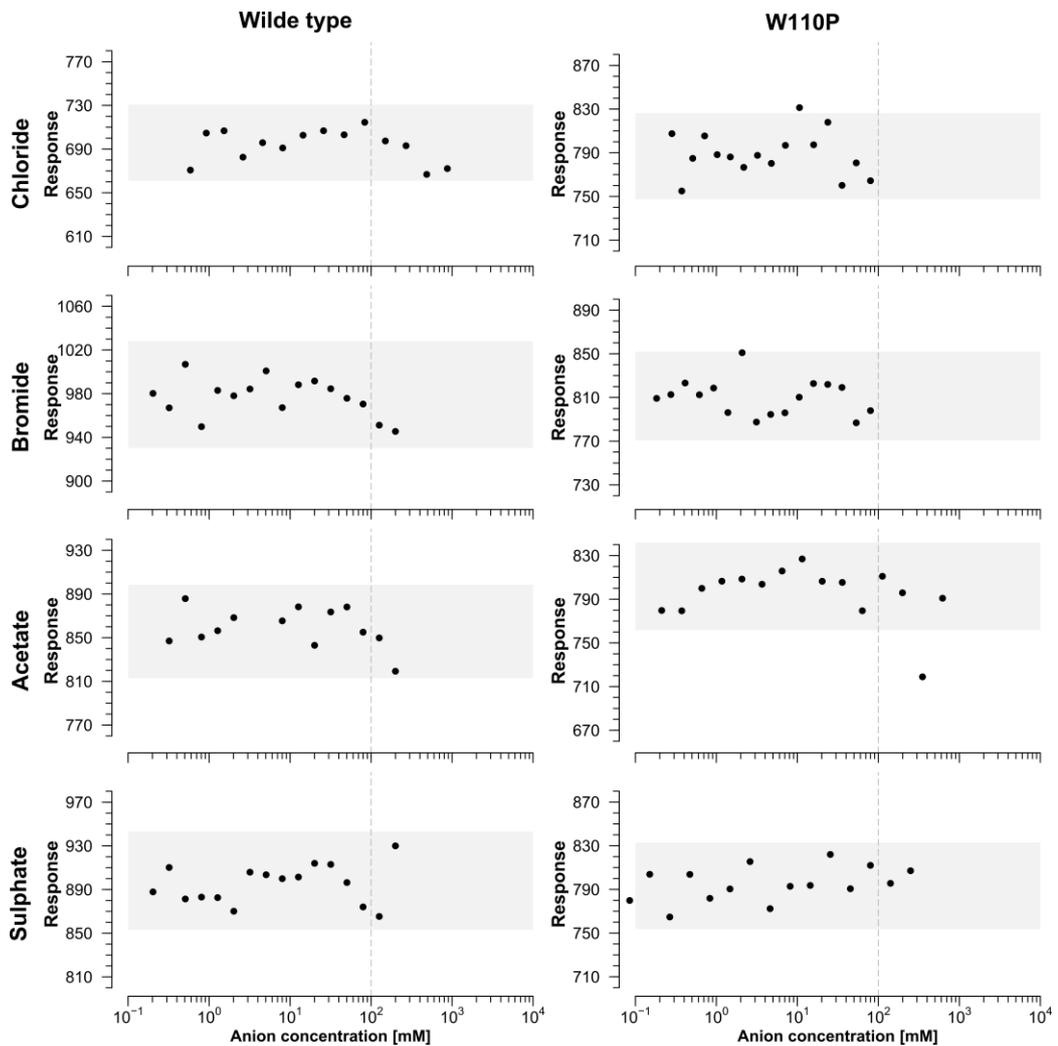
Figure 77: SD-test for the samples from the fluorescence-quenching interaction assay.

Samples were split and one half was directly measured (Figure 31), the other half denatured by heat and additional SDS. Any effect here would be independent of the tertiary structure of the protein. These samples are diluted 1/2. The grey areas indicate a deviation by 5 % from the median of a series.
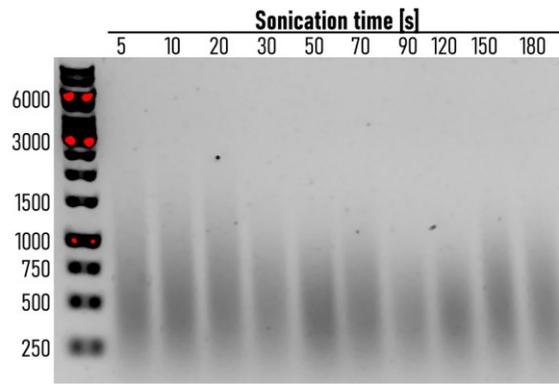
Figure 78: Herring sperm DNA used in the activity assay.

The DNA was solve in water and sonicated (2 s pulse, 4 s pause, 30 % amplitude, in ice) into pieces with 250–750 bp length.

Table 12: Statistics of datasets and their crystals of GgpS. All cell angles are in agreement with the hexagonal lattice: 90°, 90°, 120°.

Name	wt_{Cl} (7pdn)	$wt_{Cl,ADP}$ (7ptn)	$wt_{Cl,G3P}$ (7pto)	$wt_{Cl,SUC}$ (7pdl)	$wt_{Cl,SUC,Ano}$ (7ptn)
Protein	15 mg/ml R463C	10 mg/ml R463C	15 mg/ml R463C	10 mg/ml R463C	10 mg/ml R463C
$SEC_{crystallisation}$ buffer variation	NaCl, pH 8.0	NaCl, pH 8.0	NaCl, pH 8.0	NaCl, pH 8.0	NaCl, pH 8.0
Screen Condition	0.1 M Tris/HCl pH 8.5, 2 % Tacsimate pH 8.25, 14 % PEG 3350	0.1 M Tris/HCl pH 8.5, 2 % Tacsimate pH 8.00, 18 % PEG 3350	0.1 M Tris/HCl pH 8.5, 2 % Tacsimate pH 8.25, 14 % PEG 3350	0.1 M Tris/HCl pH 8.5, 2 % Tacsimate pH 7.50, 24 % PEG 3350	0.1 M Tris/HCl pH 8.5, 2 % Tacsimate pH 7.50, 24 % PEG 3350
Soaking	None	10 mM G3P, 10 mM ADPG	6 mM G3P	Cryo Condition	Cryo Condition
Cryo Condition	Paratone N	Paraffin	Paratone N	50 % sat. SUC in Reservoir	50 % saturated SUC
Beam line	ESRF – ID23-1, Pilatus 6M	SLS – PXI, Eiger X 16M	ESRF – ID23-1, Pilatus 6M	PETRA – P13, Pilatus 6M	PETRA – P13, Pilatus 6M
Wavelength [Å]	0.9724	1.7000	0.9724	0.9889	0.9889
Remarks	None	None	None	Merged	Merged, Anisotropic
Space Group	179	P 65 2 2	179	P 65 2 2	179	P 65 2 2	179	P 65 2 2	179	P 65 2 2
Cell Dimensions [Å]	101.71 101.71 344.32	102.30 102.30 349.39	101.68 101.68 336.91	103.68 103.68 364.85	103.68 103.68 364.85
Resolution	344.32–2.77 (2.93–2.77)	88.60–3.30 (3.50–3.30)	336.91–3.75 (3.98–3.75)	89.79–1.88 (1.93–1.88)	89.79–1.88 (1.93–1.88)
Completeness [%]	99.9 (99.5)	99.8 (99.0)	98.9 (93.1)	100.0 (100.0)	100.0 (100.0)
Multiplicity	18.94 (19.54)	10.63 (10.65)	13.36 (12.30)	22.02 (11.37)	11.75 (5.95)
$I/\sigma I$	20.02 (1.13)	18.14 (1.12)	9.77 (1.26)	19.18 (1.26)	14.08 (0.93)
R_{meas} [%]	8.7 (210.9)	7.4 (186.6)	16.3 (160.3)	9.3 (181.1)	9.3 (183.7)
$CC1/2$ [%]	99.9 (61.0)	99.9 (57.1)	59.8 (99.8)	99.9 (60.3)	99.9 (42.5)
Wilson B [Å²]	96.84	136.49	159.73	32.7	na
R_{work} / R_{free}	0.2602 / 0.2857	0.2309 / 0.2859	0.2441 / 0.2902	0.1802 / 0.2038	
Raman Favored / Outlier [%]	97.36 / 0.00	96.68 / 0.00	97.12 / 0.22	97.76 / 0.10	
Rotamer Outliers [%]	0.36	1.29	0.63	0.67	No model building and refinement
Clashscore	1.69	2.51	4.07	1.85	
Angle / Bond RMSD	0.611 / 0.003	0.635 / 0.003	0.987 / 0.004	0.802 / 0.004	
Proteins / Anions / Ligands per ASU	2 GgpS / 2*0.5 Cl⁻ / No ligand	2 GgpS / 2*0.5 Cl⁻ / 2 ADP	2 GgpS / 2*0.5 Cl⁻ / 2 G3P	2 GgpS / 2*0.5 Cl⁻ / 7 SUC	

Name	$wt_{Br,SUC}$ (7pdk)	$wt_{Br,SUC,Ano}$ (7pdk)	$wt_{Br,SUC,lowAno}$	$wt_{Br,SUC,highAno}$
Protein	10 mg/ml wt	10 mg/ml wt	10 mg/ml wt	10 mg/ml wt
$SEC_{crystallisation}$ buffer variation	NaBr, pH 8.5	NaBr, pH 8.5	NaBr, pH 8.5	NaBr, pH 8.5
Screen Condition	0.1 M Tris/malonic acid pH 8.5, 2 % Tacsimate pH 7.6, 19 % PEG 3350	0.1 M Tris/malonic acid pH 8.5, 2 % Tacsimate pH 7.6, 19 % PEG 3350	0.1 M Tris/malonic acid pH 8.5, 2 % Tacsimate pH 7.6, 19 % PEG 3350	0.1 M Tris/malonic acid pH 8.5, 2 % Tacsimate pH 7.6, 19 % PEG 3350
Soaking	Cryo Condition	Cryo Condition	Cryo Condition	Cryo Condition
Cryo Condition	50 % saturated SUC	50 % sat. SUC in Reservoir	50 % saturated SUC	50 % saturated SUC
Beam line, Detector	PETRA – P13, Pilatus 6M	PETRA – P13, Pilatus 6M	SLS – PXIII, Pilatus 2M-F	SLS – PXIII, Pilatus 2M-F
Wavelength [Å]	0.9763	0.9763	1.0000	0.9197
Remarks	Merged	Anomalous	Anomalous	Anomalous
Space Group	179 | P 65 2 2	179 | P 65 2 2	179 | P 65 2 2	179 | P 65 2 2
Cell Dimensions [Å]	103.68 103.68 363.27	103.61 103.61 363.71	103.88 103.88 364.89	103.78 103.78 364.72
Resolution	89.79–1.80 (1.85–1.80)	90.93–2.13 (2.19–2.13)	51.94–2.30 (2.44–2.30)	51.89–2.13 (2.26–2.13)
Completeness [%]	100.0 (100.0)	99.7 (100.0)	99.9 (99.5)	99.9 (99.3)
Multiplicity	23.23 (19.85)	5.59 (5.73)	7.86 (7.24)	10.42 (10.12)
$I/\sigma I$	23.21 (1.45)	19.93 (4.01)	13.13 (1.71)	14.31 (1.60)
R_{meas} [%]	8.3 (186.9)	5.4 (36.5)	14.4 (121.0)	15.9 (153.0)
CC1/2 [%]	100.0 (69.3)	99.9 (93.7)	99.8 (61.4)	99.8 (61.6)
Wilson B [Å^2]	32.31	na	35.64	31.07
R_{work} / R_{free}	0.1721 / 0.1948		0.1733 / 0.2129	0.1713 / 0.2064
Raman Favored / Outlier [%]	96.98 / 0.20		96.78 / 0.00	97.19 / 0.10
Rotamer Outliers [%]	1.22	No model building and refinement	0.44	0.34
Clashscore	1.77		1.66	1.42
Angle / Bond RMSD	0.798 / 0.003		0.809 / 0.006	0.722 / 0.004
Proteins / Anions / Ligands per ASU	2 GgpS / 2*0.5 Br$^-$ / 7 SUC		2 GgpS / 2*0.5 Br$^-$ / 7 SUC	2 GgpS / 2*0.5 Br$^-$ / 7 SUC

Name	W110P$_{Cl,SUC}$ (7pdm)	W110P$_{Cl,SUC,2}$	wt$_{SO4,GOL}$	wt$_{SO4,SUC}$
Protein	10 mg/ml W110P	10 mg/ml W110P	10 mg/ml wt	10 mg/ml wt
SEC$_{crystallisation}$ buffer variation	NaCl, pH7.9	NaCl, pH7.9	Na2SO4, pH 7.6	Na2SO4, pH 7.6
Screen Condition	0.20 M K/Na-tartrate, 22 % PEG 3350	0.16 M K/Na-tartrate, 26 % PEG 3350	0.23 M K/Na-tartrate, 20 % PEG 3350	0.23 M K/Na-tartrate, 20 % PEG 3350
Soaking	Cryo Condition	Cryo Condition	Cryo Condition	Cryo Condition
Cryo Condition	50 % saturated SUC	50 % saturated SUC	20 % Glycerol in FS2.1 C3	50 % saturated SUC
Beam line	PETRA – P13, Pilatus 6M	PETRA – P13, Pilatus 6M	PETRA – P13, Pilatus 6M	PETRA – P13, Pilatus 6M
Wavelength [Å]	1.0000	1.0000	0.9760	0.9760
Remarks	None	None	Anisotropic	Anisotropic
Space Group	179 | P 65 2 2	179 | P 65 2 2	178 | P 61 2 2	178 | P 61 2 2
Cell Dimensions [Å]	104.44 104.44 360.31	104.10 104.10 358.91	154.98 154.98 591.53	156.76 156.76 601.94
Resolution	90.45–3.35 (3.55–3.35)	59.82–3.30 (3.50–3.30)	601.93–3,20 (3.39–3.20)	591.53–3.43 (3.64–3.43)
Completeness [%]	99.0 (98.8)	99.7 (99.1)	100.0 (99.9)	99.9 (99.7)
Multiplicity	4.98 (5.21)	7.23 (7.38)	16.23 (16.78)	29.7 (25.99)
I/σI	7.88 (1.48)	12.10 (1.48)	9.77 (0.58)	10.04 (1.32)
R$_{meas}$ [%]	16.5 (108.0)	12.3 (120.3)	19.7 (311.8)	31.7 (185.5)
CC1/2 [%]	99.6 (61.6)	99.8 (61.5)	99.8 (61.1)	99.4 (59.3)
Wilson B [Å^2]	99.75	112.42	109.72	109.39
R$_{work}$ / R$_{free}$	0.2500 / 0.2631	0.2493 / 0.2752	0.2992 / 0.3213	0.3219 / 0.3397
Raman Favored / Outlier [%]	96.53 / 0.00	96.01	94.65 / 0.44	96.26 / 0.28
Rotamer Outliers [%]	1.81	0	5.68	2.84
Clashscore	1.92	1.93	3.99	3.66
Angle / Bond RMSD	0.665 / 0.004	0.667 / 0.003	0.876 / 0.004	0.701 / 0.003
Proteins / Anions / Ligands per ASU	2 GgpS / No anion / 3 SUC	2 GgpS / No anion / 3 SUC	6 GgpS / No anion / No ligand	6 GgpS / No anion / No ligand

Table 13: Statistics of datasets and their crystals of FtsH. All cell angles are in agreement with the rhombohedral lattice: 90°, 90°, 120°.

Name	$wt_{Aa,tr}$ (6gcn)	$F224A_{Aa,tr}$	$Q281A_{Aa,tr}$
Screen Condition	60 % Tacsimate pH 7.0, 0.1 M bis-tris-propane/HCl pH 7.0 20 °C	1.2 M trisodium–citrate, 0.1 M Tris/HCl pH 8.5, 6 °C	65 % Tacsimate pH 7.0, 0.1 M bis-tris-propane/HCl pH 7.0 6 °C
Beam line	SLS – X06DA, Pilatus 2M-F	SLS – X06DA, Pilatus 2M-F	SLS – X06DA, Pilatus 2M-F
Wavelength [Å]	1.0000	1.0000	1.0000
Remarks	Merged	No	No
Space Group	155 | R32	155 | R32	155 | R32
Cell Dimensions [Å]	197.55 197.55 323.92	197.99 197.99 328.46	198.66 196.66 324.37
Resolution	76.00–2.95 (3.03–2.95)	47.91–3.75 (3.98–3.75)	151–3.25 (3.45–3.25)
Completeness [%]	100.0 (99.8)	99.9 (99.9)	99.9 (99.4)
Multiplicity	20.6 (20.9)	10.2 (10.4)	10.2 (10.0)
$I/\sigma I$	24.73 (1.92)	13.54 (2.00)	17.06 (2.16)
R_{meas} [%]	10.5 (186.9)	14.5 (124.4)	12.6 (117.8)
$CC_{1/2}$ [%]	100.0 (71.6)	99.9 (69.6)	99.9 (74.3)
Wilson B [Å²]	88.71	131.70	94.11
R_{work} / R_{free}	0.213 / 0.256	0.241 / 0.287	0./0.213 / 255
Raman Favored / Outlier [%]	95.52 / 0.48	96.76 / 0.13	95.07 / 0.30
Rotamer Outliers [%]	0.79	1.71	4.21
Clashscore	3.9	4.86	6.45
Angle / Bond RMSD	0.67 / 0.004	0.755 / 0.004	1.573 / 0.016
Proteins / Ligands per ASU	4 FtsH / 4 ADP	4 FtsH / 4 ADP	4 FtsH / 4 ADP

10 ACKNOWLEDGMENTS

In the very beginning, then I was eleven or twelve years old, I wanted to do something with physics later in my life. However, biology was first the only, then the logical and eventually the right decision. I am now graduating in biochemistry, in some sense the absolute opposite from physics one might think. However, diffraction and synchrotrons, electron microscopy and programming, MST and DLS, energy conversion and domain movements – the biology and the chemistry is sometimes hard to see behind all the physics. I guess I found my place.

I want to thank you, Ulrich Baumann, for introducing me into the world of structural biology. The start (of the PhD) was probably financially hard, but we applied successfully for a 2-year fellowship from the *Association of the Chemical Industry e.V.* You teached me a lot about proteins, how they work and how to analyse them. Thanks to you I travelled the 'western' world and got in contact with scientist from all over the world. I started slow with London, Paris, Grenoble, Pavia and Barcelona but soon I was in San Francisco, Boston and Toronto. I always felt to be in good hands under you supervision and I will miss our chats in the evening after all the early riser left the lab.

Then there is Jan Gebauer, my second 'Doktorvater'. I am sure you sometimes wished me on a distant island. Yet I hope it was as much fun to work with me as it was for me to work with you. Your help, supervision and managing skills inspired me. I probably do not really know which things you organised and made possible, or endured behind the scenes so I (all of us) could do my (our) work. You definitely deserve a huge 'thank you'!

Karsten Niefind, you accompanied me during my time as bachelor, master and PhD. First as mentor, then as reviewer and eventually as adviser. Your questions were often great which means they let me rethink and gave me new impulses. I thank you for the last ten years.

Kay Hofmann, you let a bachelor student, you did not really know and who had no idea about programming, write a bachelor thesis about a little program for your everyday work. You told me on some occasions that the tool really helped and I was excited to use it myself 'in the future'. Well, know I worked usually on bacterial proteins which are somehow not covered by a program about co-expression in eukaryotes. However, even if the program itself had no use for me so far, creating it was certainly the most important skill I learned during my time as bachelor. A lot of the work here presented would have not been possible without the chance you gave me back then.

Cryo-electron microscopy is a great complementary tool to X-ray crystallography. Thank you, Elmar Behrmann, for the cooperation on FtsH and introducing me into this exciting field. I also thank you Gayathri for your work in the lab and the help in understanding cryo-EM. Unfortunately, FtsH did not want as we did but I am eagerly awaiting the processing of the new dataset. Hopefully, we can write a publication by this fall.

At this location, probably 199 pages too late, I thank you, Amy and Anubis. Your constant carping, niggling, moaning and crying was… no help at all! Nevertheless you were a purrfect deviation from the everyday stress. YES, you get your treats now!

I definitely have to thank my fellow PhD students. 'The old guard' consisting of Swantje, Anna, Sinan and Christian. Your warm welcoming helped me to find my way and to become the scientist I am now. You left me early, but being the 'senior' PhD student for so long time was great, even if Elena and Fabian ignored that entirely. Not to mention the adopted stepsisters Bine, Jacky and Magdalena. You all helped, gave advices and made life and work as 'senior' PhD student bright and sweet. Yes, I do mean the cookies and biscuits.

Many thanks also belong to Sandra, Doro and Elena. Similar to Jan, I probably do not really know what you all did so I could do my work. I enjoyed to work and chat with you in the past six years.

During the five years as PhD student I supervised Georg Rabbow, Seyma Bozkus, Monique Henschel and Leon Wehrhan. They worked on GgpS and helped me a lot. Special thanks to Monique Henschel who did her master's thesis on GgpS and established the first iteration of the pyruvate kinase/lactate dehydrogenase coupled assay. She also crystallised the bromide bound GgpS. In addition, I supervised Pavel Schweizer and Benjamin Chapple on FtsH related projects. I did not included Pavel's work about the crystallisation of the truncated FtsH from *Thermus thermophiles*. Benjamin helped me with the assays, especially the production of the GFP variants. I also have to thank Julia Charl. You want to start your master's thesis about FtsH soon. Yet, you already volunteered for proofreading of the introduction of the FtsH part for this dissertation. I wish you all best for your thesis and hope you can work in the lab!

Many thanks also go to the beam line or cryo-EM staffs at the SLS, ESRF and PETRA. None of this work would have been possible without them. I also thank Lutz Schmitz for SEC-MALS and Karl R. Schmitz from the Sauer lab for the $_{cp6}$GFP-ssrA.

I want to thank my friends outside the Baumann lab. Tobias you have been a great friend over the countless years we know each other. Special thanks for the enlightenment in regards to statistics used in this dissertation. Mostly, together with Alina and Nils you helped me escaping the work. During the time as PhD students, *our* five-year mission was to explore strange new videogames, to seek out new heists and new challenges, to boldly go where no player has gone before. I am glad no one had a red shirt. Thank you Jessi to keep Tobi under control and Christian for all your indirect support during the 'missions' with Jan-Erik and Heinrich in the Ding and the Old School.

Last but not least I want to thank my wife Schirin. You supported me all the time and endured my long stays in the institute. As technician in our lab you helped me with some mutageneses and especially with the FtsH from *E. coli*. I severely missed you when I was in San Francisco and was very happy when

you joined my trip to Boston and Toronto. I am excited about our next adventure: Our son Sam!

11 FIGURES AND TABLES

11.1 Figures

11.2 Tables

12 LITERATURE

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*, *66*(2), 213–221. https://doi.org/10.1107/S0907444909052925

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., Adams, P. D., & IUCr. (2012). Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallographica Section D Biological Crystallography*, *68*(4), 352–367. https://doi.org/10.1107/S0907444912001308

Afonine, P. V., Moriarty, N. W., Mustyakimov, M., Sobolev, O. V., Terwilliger, T. C., Turk, D., Urzhumtsev, A., & Adams, P. D. (2015). FEM: feature-enhanced map. *Acta Crystallographica Section D Biological Crystallography*, *71*(3), 646–666. https://doi.org/10.1107/S1399004714028132

Akiyama, Y. (2002). Proton-motive force stimulates the proteolytic activity of FtsH, a membrane-bound ATP- dependent protease in Escherichia coli. *Proceedings of the National Academy of Sciences*, *99*(12), 8066–8071. https://doi.org/10.1073/pnas.122616899

Akiyama, Yoshinori. (1999). Self-Processing of FtsH and Its Implication for the Cleavage Specificity of This Protease †. *Biochemistry*, *38*(36), 11693–11699. https://doi.org/10.1021/bi991177c

Akiyama, Yoshinori, Kihara, A., Mori, H., Ogura, T., & Ito, K. (1998). Roles of the Periplasmic Domain of Escherichia coli FtsH (HflB) in Protein Interactions and Activity Modulation. *Journal of Biological Chemistry*, *273*(35), 22326–22333. https://doi.org/10.1074/jbc.273.35.22326

Alber, A. B., & Suter, D. M. (2019). Dynamics of protein synthesis and degradation through the cell cycle. *Cell Cycle*, *18*(8), 784–794. https://doi.org/10.1080/15384101.2019.1598725

An, J. Y., Sharif, H., Kang, G. B., Park, K. J., Lee, J.-G., Lee, S., Jin, M. S., Song, J.-J., Wang, J., & Eom, S. H. (2017). Structural insights into the oligomerization of FtsH periplasmic domain from Thermotoga maritima. *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/j.bbrc.2017.11.158

Anand, R., Wai, T., Baker, M. J., Kladt, N., Schauss, A. C., Rugarli, E., & Langer, T. (2014). The i-AAA protease YME1L and OMA1 cleave OPA1 to balance mitochondrial fusion and fission. *Journal of Cell Biology*, *204*(6), 919–929. https://doi.org/10.1083/jcb.201308006

Anisimova, A. S., Alexandrov, A. I., Makarova, N. E., Gladyshev, V. N., & Dmitriev, S. E. (2018). Protein synthesis and quality control in aging. *Aging*, *10*(12), 4269–4288. https://doi.org/10.18632/aging.101721

Arlt, H., Steglich, G., Perryman, R., Guiard, B., Neupert, W., & Langer, T. (1998). The formation of respiratory chain complexes in mitochondria is under the proteolytic control of the m-AAA protease. *The EMBO Journal*, *17*(16), 4837–4847. https://doi.org/10.1093/emboj/17.16.4837

Arlt, H., Tauer, R., Feldmann, H., Neupert, W., & Langer, T. (1996). The YTA10–12 Complex, an AAA Protease with Chaperone-like Activity in the Inner Membrane of Mitochondria. *Cell*, *85*(6), 875–885. https://doi.org/10.1016/S0092-8674(00)81271-4

Asahara, Y., Atsuta, K., Motohashi, K., Taguchi, H., Yohda, M., & Yoshida, M. (2000). FtsH Recognizes Proteins with Unfolded Structure and Hydrolyzes the Carboxyl Side of Hydrophobic Residues. *Journal of Biochemistry*, *127*(5), 931–937. https://doi.org/10.1093/oxfordjournals.jbchem.a022689

Assmann, G., Brehm, W., & Diederichs, K. (2016). Identification of rogue datasets in serial crystallography. *Journal of Applied Crystallography*, *49*(3), 1021–1028. https://doi.org/10.1107/S1600576716005471

Augustin, S., Gerdes, F., Lee, S., Tsai, F. T. F., Langer, T., & Tatsuta, T. (2009). An Intersubunit Signaling Network Coordinates ATP Hydrolysis by m-AAA Proteases. *Molecular Cell*, *35*(5), 574–585. https://doi.org/10.1016/j.molcel.2009.07.018

Baek, Y. H., & Nowak, T. (1982). Kinetic evidence for a dual cation role for muscle pyruvate kinase. *Archives of Biochemistry and Biophysics*, *217*(2), 491–497. https://doi.org/10.1016/0003-9861(82)90529-X

Ball, P., & Hallsworth, J. E. (2015). Water structure and chaotropicity: Their uses, abuses and biological implications. In *Physical Chemistry Chemical Physics* (Vol. 17, Issue 13, pp. 8297–8305). Royal Society of Chemistry. https://doi.org/10.1039/c4cp04564e

Bandyopadhyay, K., Parua, P. K., Datta, A. B., & Parrack, P. (2010). Escherichia coli HflK and HflC can individually inhibit the HflB (FtsH)-mediated proteolysis of λCII in vitro. *Archives of Biochemistry and Biophysics*, *501*(2), 239–243. https://doi.org/10.1016/j.abb.2010.06.030

Barth, S., Huhn, M., Matthey, B., Klimka, A., Galinski, E. A., & Engert, A. (2000). Compatible-solute-supported periplasmic expression of functional recombinant proteins under stress conditions. *Applied and Environmental Microbiology*, *66*(4), 1572–1579. https://doi.org/10.1128/AEM.66.4.1572-1579.2000

Bayburt, T. H., & Sligar, S. G. (2010). Membrane protein assembly into Nanodiscs. *FEBS Letters*, *584*(9), 1721–1727. https://doi.org/10.1016/j.febslet.2009.10.024

Beckwith, R., Estrin, E., Worden, E. J., & Martin, A. (2013). Reconstitution of the 26S proteasome reveals functional asymmetries in its AAA+ unfoldase. *Nature Structural & Molecular Biology*, *20*(10), 1164–1172. https://doi.org/10.1038/nsmb.2659

Bhabha, G., Johnson, G. T., Schroeder, C. M., & Vale, R. D. (2016). How Dynein Moves Along Microtubules. *Trends in Biochemical Sciences*, *41*(1), 94–105. https://doi.org/10.1016/j.tibs.2015.11.004

Bieniossek, C., Niederhauser, B., & Baumann, U. M. (2009). The crystal structure of apo-FtsH reveals domain movements necessary for substrate unfolding and translocation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(51), 21579–21584. https://doi.org/10.1073/pnas.0910708106

Bieniossek, C., Schalch, T., Bumann, M., Meister, M., Meier, R., & Baumann, U. (2006). The molecular architecture of the metalloprotease FtsH. *Proceedings of the National Academy of Sciences of the United States of*

America, 103(9), 3066–3071. https://doi.org/10.1073/pnas.0600031103

Bitop AG. (2020). *Products of bitop AG.* https://www.bitop.de/en/extremolytes

Bittner, L.-M., Arends, J., & Narberhaus, F. (2017). When, how and why? Regulated proteolysis by the essential FtsH protease in Escherichia coli. *Biological Chemistry, 398*(5–6), 625–635. https://doi.org/10.1515/hsz-2016-0302

Bittner, L.-M., Westphal, K., & Narberhaus, F. (2015). Conditional Proteolysis of the Membrane Protein YfgM by the FtsH Protease Depends on a Novel N-terminal Degron. *Journal of Biological Chemistry, 290*(31), 19367–19378. https://doi.org/10.1074/jbc.M115.648550

Bloch, M., Santiveri, M., & Taylor, N. M. I. (2020). Membrane Protein Cryo-EM: Cryo-Grid Optimization and Data Collection with Protein in Detergent. In C. Perez & T. Maier (Eds.), *Expression, Purification, and Structural Biology of Membrane Proteins, vol 2127* (pp. 227–244). Humana, New York, NY. https://doi.org/10.1007/978-1-0716-0373-4_16

Boehm, M., Yu, J., Krynicka, V., Barker, M., Tichy, M., Komenda, J., Nixon, P. J., & Nield, J. (2012). Subunit Organization of a Synechocystis Hetero-Oligomeric Thylakoid FtsH Complex Involved in Photosystem II Repair. *The Plant Cell, 24*(9), 3669–3683. https://doi.org/10.1105/tpc.112.100891

Bolivar, J. M., Luley-Goedl, C., Leitner, E., Sawangwan, T., & Nidetzky, B. (2017). Production of glucosyl glycerol by immobilized sucrose phosphorylase: Options for enzyme fixation on a solid support and application in microscale flow format. *Journal of Biotechnology, 257*, 131–138. https://doi.org/10.1016/j.jbiotec.2017.01.019

Bonn, F., Tatsuta, T., Petrungaro, C., Riemer, J., & Langer, T. (2011). Presequence-dependent folding ensures MrpL32 processing by the m -AAA protease in mitochondria. *The EMBO Journal, 30*(13), 2545–2556. https://doi.org/10.1038/emboj.2011.169

Bouchnak, I., & van Wijk, K. J. (2019). N-Degron Pathways in Plastids. *Trends in Plant Science, 24*(10), 917–926. https://doi.org/10.1016/j.tplants.2019.06.013

Bowie, J. U. (1997). Helix packing in membrane proteins. *Journal of Molecular Biology, 272*(5), 780–789. https://doi.org/10.1006/jmbi.1997.1279

Bünger, J., Degwert, J., & Driller, H. (2001). The Protective Function of Compatible Solute Ectoin on the Skin, Skin Cellsand its Biomolecules with Respect to UV-Radiation,Immunosupression and Membrane Damage. *International Federation of Societies of Cosmetic Chemists Magazine, 4*(2), 127–131. https://www.merckgroup.com/business-specifics/performance-materials/Cosmetics/publications-studies/Protective Function of compatible solute Ectoin.pdf

Campbell, N. A., & Reece, J. B. (2009). 36.2.2 Diffusion von Wasser. In *Biologie* (8th ed., pp. 1038–1039). Pearson Studium.

Carugo, O., JJR, F. da S., Williams, R., Bonar, P., Casey, J., Cant, N., Pollock, N., Ford, R., Buisson, G., Duee, E., Haser, R., Payan, F., Bricker, T., Roose, J., Fagerlund, R., Frankel, L., Eaton-Rye, J., Umena, Y., Kawakami, K., … Shannon, R. (2014). Buried chloride stereochemistry in the Protein Data Bank. *BMC Structural Biology, 14*(1), 19. https://doi.org/10.1186/s12900-014-0019-8

Carvalho, V., Kieffer, R., Lange, N. de, Engel, A., & Aubin-Tam, M.-E. (2017). Large conformational changes in FtsH create an opening for substrate entry. *BioRxiv*, 209445. https://doi.org/10.1101/209445

Carvalho, V., Prabudiansyah, I., Kovacik, L., Chami, M., Kieffer, R., van der Valk, R., de Lange, N., Engel, A., & Aubin-Tam, M.-E. (2020). The cytoplasmic domain of the AAA+ protease FtsH is tilted with respect to the membrane to facilitate substrate entry. *Journal of Biological Chemistry*, jbc.RA120.014739. https://doi.org/10.1074/jbc.RA120.014739

Chiba, S., Akiyama, Y., & Ito, K. (2002). Membrane Protein Degradation by FtsH Can Be Initiated from Either End. *Journal of Bacteriology*, *184*(17), 4775–4782. https://doi.org/10.1128/JB.184.17.4775-4782.2002

Chiba, S., Akiyama, Y., Mori, H., Matsuo, E., & Ito, K. (2000). Length recognition at the N-terminal tail for the initiation of FtsH-mediated proteolysis. *EMBO Reports*, *1*(1), 47–52. https://doi.org/10.1093/embo-reports/kvd005

Chiba, S., Ito, K., & Akiyama, Y. (2006). The Escherichia coli plasma membrane contains two PHB (prohibitin homology) domain protein complexes of opposite orientations. *Molecular Microbiology*, *60*(2), 448–457. https://doi.org/10.1111/j.1365-2958.2006.05104.x

Chuang, G. Y., Kozakov, D., Brenke, R., Comeau, S. R., & Vajda, S. (2008). DARS (Decoys As the Reference State) potentials for protein-protein docking. *Biophysical Journal*, *95*(9), 4217–4227. https://doi.org/10.1529/biophysj.108.135814

Cocchiaro, P., De Pasquale, V., Della Morte, R., Tafuri, S., Avallone, L., Pizard, A., Moles, A., & Pavone, L. M. (2017). The Multifaceted Role of the Lysosomal Protease Cathepsins in Kidney Disease. *Frontiers in Cell and Developmental Biology*, *5*(114). https://doi.org/10.3389/fcell.2017.00114

Cordova, J. C., Olivares, A. O., Shin, Y., Stinson, B. M., Calmat, S., Schmitz, K. R., Aubin-Tam, M.-E., Baker, T. A., Lang, M. J., & Sauer, R. T. (2014). Stochastic but Highly Coordinated Protein Unfolding and Translocation by the ClpXP Proteolytic Machine. *Cell*, *158*(3), 647–658. https://doi.org/10.1016/j.cell.2014.05.043

Coutinho, P. M., Deleury, E., Davies, G. J., & Henrissat, B. (2003). An evolving hierarchical family classification for glycosyltransferases. *Journal of Molecular Biology*, *328*(2), 307–317. https://doi.org/10.1016/S0022-2836(03)00307-3

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Research*, *14*(6), 1188–1190. https://doi.org/10.1101/gr.849004

Czech, L., Hermann, L., Stöveken, N., Richter, A. A., Höppner, A., Smits, S. H. J., Heider, J., & Bremer, E. (2018). Role of the extremolytes ectoine and hydroxyectoine as stress protectants and nutrients: Genetics, phylogenomics, biochemistry, and structural analysis. In *Genes* (Vol. 9, Issue 4, p. 177). MDPI AG. https://doi.org/10.3390/genes9040177

Dacks, J. B., & Field, M. C. (2007). Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. *Journal of Cell Science, 120*(17), 2977–2985. https://doi.org/10.1242/jcs.013250

Dasuri, K., Zhang, L., & Keller, J. N. (2013). Oxidative stress, neurodegeneration, and the balance of protein degradation and protein synthesis. *Free Radical Biology and Medicine*, *62*, 170–185.

https://doi.org/10.1016/j.freeradbiomed.2012.09.016

Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J., & Swanson, R. V. (1998). The complete genome of the hyperthermophilic bacterium Aquifex aeolicus. *Nature*, *392*(6674), 353–358. https://doi.org/10.1038/32831

Delcour, A. H. (2009). Outer membrane permeability and antibiotic resistance. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1794*(5), 808–816. https://doi.org/10.1016/j.bbapap.2008.11.005

Deller, M. C., Kong, L., & Rupp, B. (2016). Protein stability: a crystallographer's perspective. *Acta Crystallographica Section F Structural Biology Communications*, *72*(2), 72–95. https://doi.org/10.1107/S2053230X15024619

Deng, Z., Paknejad, N., Maksaev, G., Sala-Rabanal, M., Nichols, C. G., Hite, R. K., & Yuan, P. (2018). Cryo-EM and X-ray structures of TRPV4 reveal insight into ion permeation and gating mechanisms. *Nature Structural & Molecular Biology*, *25*(3), 252–260. https://doi.org/10.1038/s41594-018-0037-5

Denisov, I. G., & Sligar, S. G. (2017). Nanodiscs in Membrane Biochemistry and Biophysics. *Chemical Reviews*, acs.chemrev.6b00690. https://doi.org/10.1021/acs.chemrev.6b00690

Deville, C., Carroni, M., Franke, K. B., Topf, M., Bukau, B., Mogk, A., & Saibil, H. R. (2017). Structural pathway of regulated substrate transfer and threading through an Hsp100 disaggregase. *Science Advances*, *3*(8), e1701726. https://doi.org/10.1126/sciadv.1701726

Diederichs, K. (n.d.). *xdsstat*. ftp://strucbio.biologie.uni-konstanz.de/pub/linux_bin/xdsstat

Dong, Y., Zhang, S., Wu, Z., Li, X., Wang, W. L., Zhu, Y., Stoilova-McPhie, S., Lu, Y., Finley, D., & Mao, Y. (2019). Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature*, *565*(7737), 49–55. https://doi.org/10.1038/s41586-018-0736-4

Dörr, J. M., Scheidelaar, S., Koorengevel, M. C., Dominguez, J. J., Schäfer, M., van Walree, C. A., & Killian, J. A. (2016). The styrene-maleic acid copolymer: a versatile tool in membrane research. *European Biophysics Journal : EBJ*, *45*(1), 3–21. https://doi.org/10.1007/s00249-015-1093-y

Dougan, D. A., Micevski, D., & Truscott, K. N. (2012). The N-end rule pathway: From recognition by N-recognins, to destruction by AAA+proteases. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1823*(1), 83–91. https://doi.org/10.1016/j.bbamcr.2011.07.002

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 4), 486–501. https://doi.org/10.1107/S0907444910007493

Enemark, E. J., & Joshua-Tor, L. (2006). Mechanism of DNA translocation in a replicative hexameric helicase. *Nature*, *442*(7100), 270–275. https://doi.org/10.1038/nature04943

England, J. L., Pande, V. S., & Haran, G. (2008). Chemical denaturants inhibit the onset of dewetting. *Journal of the American Chemical Society*, *130*(36), 11854–11855. https://doi.org/10.1021/ja803972g

Errey, J. C., Lee, S. S., Gibson, R. P., Martinez Fleites, C., Barry, C. S., Jung, P. M. J., O'Sullivan, A. C., Davis, B. G., & Davies, G. J. (2010). Mechanistic Insight into Enzymatic Glycosyl Transfer with Retention of Configuration through Analysis of Glycomimetic Inhibitors. *Angewandte Chemie International Edition*, *49*(7), 1234–1237. https://doi.org/10.1002/anie.200905096

Erwin, A. L. (2016). Antibacterial Drug Discovery Targeting the Lipopolysaccharide Biosynthetic Enzyme LpxC. *Cold Spring Harbor Perspectives in Medicine*, *6*(7), a025304. https://doi.org/10.1101/cshperspect.a025304

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallographica Section D Biological Crystallography*, *62*(1), 72–82. https://doi.org/10.1107/S0907444905036693

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D Biological Crystallography*, *69*(7), 1204–1214. https://doi.org/10.1107/S0907444913000061

Faria, T. Q., Mingote, A., Siopa, F., Ventura, R., Maycock, C., & Santos, H. (2008). Design of new enzyme stabilizers inspired by glycosides of hyperthermophilic microorganisms. *Carbohydrate Research*, *343*(18), 3025–3033. https://doi.org/10.1016/j.carres.2008.08.030

Fei, X., Bell, T. A., Jenni, S., Stinson, B. M., Baker, T. A., Harrison, S. C., & Sauer, R. T. (2020). Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. *ELife*, *9*. https://doi.org/10.7554/eLife.52774

Frauenfeld, J., Löving, R., Armache, J.-P., Sonnen, A. F.-P., Guettou, F., Moberg, P., Zhu, L., Jegerschöld, C., Flayhan, A., Briggs, J. A. G., Garoff, H., Löw, C., Cheng, Y., & Nordlund, P. (2016). A saposin-lipoprotein nanoparticle system for membrane proteins. *Nature Methods*, *13*(4), 345–351. https://doi.org/10.1038/nmeth.3801

Fritze, J., Zhang, M., Luo, Q., & Lu, X. (2020). An overview of the bacterial SsrA system modulating intracellular protein levels and activities. *Applied Microbiology and Biotechnology*, *104*(12), 5229–5241. https://doi.org/10.1007/s00253-020-10623-x

Führer, F., Müller, A., Baumann, H., Langklotz, S., Kutscher, B., & Narberhaus, F. (2007). Sequence and Length Recognition of the C-terminal Turnover Element of LpxC, a Soluble Substrate of the Membrane-bound FtsH Protease. *Journal of Molecular Biology*, *372*(2), 485–496. https://doi.org/10.1016/j.jmb.2007.06.083

Fuller, D. N., Raymer, D. M., Kottadiel, V. I., Rao, V. B., & Smith, D. E. (2007). Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. *Proceedings of the National Academy of Sciences*, *104*(43), 16868–16873. https://doi.org/10.1073/pnas.0704008104

Galinski, E. A., & Trüper, H. G. (1994). Microbial behaviour in salt-stressed ecosystems. *FEMS Microbiology Reviews*, *15*(2–3), 95–108. https://doi.org/10.1111/j.1574-6976.1994.tb00128.x

Gao, Y., Cui, Y., Fox, T., Lin, S., Wang, H., de Val, N., Zhou, Z. H., & Yang, W. (2019). Structures and operating principles of the replisome. *Science*, *363*(6429), eaav7003. https://doi.org/10.1126/science.aav7003

Gloster, T. M. (2014). Advances in understanding glycosyltransferases from a structural perspective. In *Current Opinion in Structural Biology* (Vol. 28, pp.

131–141). Elsevier Ltd. https://doi.org/10.1016/j.sbi.2014.08.012

Götzke, H., Muheim, C., Altelaar, A. F. M., Heck, A. J. R., Maddalo, G., & Daley, D. O. (2015). Identification of Putative Substrates for the Periplasmic Chaperone YfgM in Escherichia coli Using Quantitative Proteomics. *Molecular & Cellular Proteomics*, *14*(1), 216–226. https://doi.org/10.1074/mcp.M114.043216

Götzke, H., Palombo, I., Muheim, C., Perrody, E., Genevaux, P., Kudva, R., Müller, M., & Daley, D. O. (2014). YfgM Is an Ancillary Subunit of the SecYEG Translocon in Escherichia coli. *Journal of Biological Chemistry*, *289*(27), 19089–19097. https://doi.org/10.1074/jbc.M113.541672

Griffith, K. L., Shah, I. M., & E. Wolf, R. (2004). Proteolytic degradation of Escherichia coli transcription activators SoxS and MarA as the mechanism for reversing the induction of the superoxide (SoxRS) and multiple antibiotic resistance (Mar) regulons. *Molecular Microbiology*, *51*(6), 1801–1816. https://doi.org/10.1046/j.1365-2958.2003.03952.x

Guest, R. L., Samé Guerra, D., Wissler, M., Grimm, J., & Silhavy, T. J. (2020). YejM Modulates Activity of the YciM/FtsH Protease Complex To Prevent Lethal Accumulation of Lipopolysaccharide. *MBio*, *11*(2). https://doi.org/10.1128/mBio.00598-20

Hagemann, M., Effmert, U., Kerstan, T., Schoor, A., & Erdmann, N. (2001). Biochemical characterization of glucosylglycerol-phosphate synthase of Synechocystis sp. strain PCC 6803: comparison of crude, purified, and recombinant enzymes. *Current Microbiology*, *43*(4), 278–283. http://www.ncbi.nlm.nih.gov/pubmed/11683363

Hagemann, M., & Erdmann, N. (1994). Activation and pathway of glucosylglycerol synthesis in the cyanobacterium Synechocystis sp. PCC 6803. *Microbiology*, *140*(6), 1427–1431. https://doi.org/10.1099/00221287-140-6-1427

Hagemann, Martin. (2011). Molecular biology of cyanobacterial salt acclimation. In *FEMS Microbiology Reviews* (Vol. 35, Issue 1, pp. 87–123). Oxford Academic. https://doi.org/10.1111/j.1574-6976.2010.00234.x

Hagemann, Martin, Schoor, A., & Erdmann, N. (1996). NaCl acts as a direct modulator in the salt adaptive response: Salt-dependent activation of glucosylglycerol synthesis in vivo and in vitro. *Journal of Plant Physiology*, *149*(6), 746–752. https://doi.org/10.1016/S0176-1617(96)80101-5

Hagn, F., Etzkorn, M., Raschle, T., & Wagner, G. (2013). Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins. *Journal of the American Chemical Society*, *135*(5), 1919–1925. https://doi.org/10.1021/ja310901f

Haines, T. H. (1994). Water transport across biological membranes. *FEBS Letters*, *346*(1), 115–122. https://doi.org/10.1016/0014-5793(94)00470-6

Hanna, A., Berg, M., Stout, V., & Razatos, A. (2003). Role of Capsular Colanic Acid in Adhesion of Uropathogenic Escherichia coli. *Applied and Environmental Microbiology*, *69*(8), 4474–4481. https://doi.org/10.1128/AEM.69.8.4474-4481.2003

Hanna, J., Guerra-Moreno, A., Ang, J., & Micoogullari, Y. (2019). Protein Degradation and the Pathologic Basis of Disease. *The American Journal of Pathology*, *189*(1), 94–103. https://doi.org/10.1016/j.ajpath.2018.09.004

Hanson, P. I., & Whiteheart, S. W. (2005). AAA+ proteins: have engine, will work. *Nature Reviews Molecular Cell Biology*, *6*(7), 519–529. https://doi.org/10.1038/nrm1684

Harjes, S., Scheidig, A., & Bayer, P. (2004). Expression, purification and crystallization of human 3′-phosphoadenosine-5′-phosphosulfate synthetase 1. *Acta Crystallographica Section D: Biological Crystallography*, *60*(2), 350–352. https://doi.org/10.1107/S0907444903027628

Hayward, S., & Berendsen, H. J. (1998). Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins*, *30*(2), 144–154. http://www.ncbi.nlm.nih.gov/pubmed/9489922

Herman, C., Thevenet, D., Bouloc, P., Walker, G. C., & D'Ari, R. (1998). Degradation of carboxy-terminal-tagged cytoplasmic proteins by the Escherichia coli protease HflB (FtsH). *Genes & Development*, *12*(9), 1348–1355. https://doi.org/10.1101/gad.12.9.1348

Herman, C, Thévenet, D., D'Ari, R., & Bouloc, P. (1997). The HflB protease of Escherichia coli degrades its inhibitor lambda cIII. *Journal of Bacteriology*, *179*(2), 358–363. https://doi.org/10.1128/JB.179.2.358-363.1997

Herman, Christophe, Prakash, S., Lu, C. Z., Matouschek, A., & Gross, C. A. (2003). Lack of a Robust Unfoldase Activity Confers a Unique Level of Substrate Specificity to the Universal AAA Protease FtsH. *Molecular Cell*, *11*(3), 659–669. https://doi.org/10.1016/S1097-2765(03)00068-6

Hofmeister, F. (1888). Zur Lehre von der Wirkung der Salze - Zweite Mittheilung. *Archiv Für Experimentelle Pathologie Und Pharmakologie*, *24*(4–5), 247–260. https://doi.org/10.1007/BF01918191

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure on JSTOR. *Scandinavian Journal of Statistics*, *6*(2), 65–70. https://doi.org/10.2307/4615733

Holthauzen, L. M. F., Rösgen, J., & Bolen, D. W. (2010). Hydrogen bonding progressively strengthens upon transfer of the protein urea-denatured state to water and protecting osmolytes. *Biochemistry*, *49*(6), 1310–1318. https://doi.org/10.1021/bi9015499

Iadanza, M. G., Higgins, A. J., Schiffrin, B., Calabrese, A. N., Brockwell, D. J., Ashcroft, A. E., Radford, S. E., & Ranson, N. A. (2016). Lateral opening in the intact β-barrel assembly machinery captured by cryo-EM. *Nature Communications*, *7*(1), 12865. https://doi.org/10.1038/ncomms12865

Iyer, L. M., Leipe, D. D., Koonin, E. V, & Aravind, L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. *Journal of Structural Biology*, *146*(1–2), 11–31. https://doi.org/10.1016/j.jsb.2003.10.010

Jackson, M. P., & Hewitt, E. W. (2016). Cellular proteostasis: degradation of misfolded proteins by lysosomes. *Essays in Biochemistry*, *60*(2), 173–180. https://doi.org/10.1042/EBC20160005

Janska, H., Kwasniak, M., & Szczepanowska, J. (2013). Protein quality control in organelles - AAA/FtsH story. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1833*(2). https://doi.org/10.1016/j.bbamcr.2012.03.016

Jorge, C. D., Borges, N., Bagyan, I., Bilstein, A., & Santos, H. (2016). Potential applications of stress solutes from extremophiles in protein folding diseases and healthcare. In *Extremophiles* (Vol. 20, Issue 3, pp. 251–259). Springer

Tokyo. https://doi.org/10.1007/s00792-016-0828-8

Kabsch, W. (2010). XDS. *Acta Crystallographica Section D Biological Crystallography*, *66*(2), 125–132. https://doi.org/10.1107/S0907444909047337

Kachmar, J. F., & Boyer, P. D. (1953). Kinetic analysis of enzyme reactions. II. The potassium activation and calcium inhibition of pyruvic phosphoferase. *The Journal of Biological Chemistry*, *200*(2), 669–682.

Karata, K., Inagawa, T., Wilkinson, A. J., Tatsuta, T., & Ogura, T. (1999). Dissecting the Role of a Conserved Motif (the Second Region of Homology) in the AAA Family of ATPases. *Journal of Biological Chemistry*, *274*(37), 26225–26232. https://doi.org/10.1074/jbc.274.37.26225

Karata, K., Verma, C. S., Wilkinson, A. J., & Ogura, T. (2001). Probing the mechanism of ATP hydrolysis and substrate translocation in the AAA protease FtsH by modelling and mutagenesis. *Molecular Microbiology*, *39*(4), 890–903. https://doi.org/10.1046/j.1365-2958.2001.02301.x

Kelch, B. A., Makino, D. L., O'Donnell, M., & Kuriyan, J. (2012). Clamp loader ATPases and the evolution of DNA replication machinery. *BMC Biology*, *10*(1), 34. https://doi.org/10.1186/1741-7007-10-34

Kihara, A., Akiyama, Y., & Ito, K. (1995). FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proceedings of the National Academy of Sciences*, *92*(10), 4532–4536. https://doi.org/10.1073/pnas.92.10.4532

Kihara, A., Akiyama, Y., & Ito, K. (1996). A protease complex in the Escherichia coli plasma membrane: HflKC (HflA) forms a complex with FtsH (HflB), regulating its proteolytic activity against SecY. *The EMBO Journal*, *15*(22), 6122–6131. https://doi.org/10.1002/j.1460-2075.1996.tb01000.x

Kihara, A., Akiyama, Y., & Ito, K. (1997). Host regulation of lysogenic decision in bacteriophage : Transmembrane modulation of FtsH (HflB), the cII degrading protease, by HflKC (HflA). *Proceedings of the National Academy of Sciences*, *94*(11), 5544–5549. https://doi.org/10.1073/pnas.94.11.5544

Klein, G., Kobylak, N., Lindner, B., Stupak, A., & Raina, S. (2014). Assembly of Lipopolysaccharide in Escherichia coli Requires the Essential LapB Heat Shock Protein. *Journal of Biological Chemistry*, *289*(21), 14829–14853. https://doi.org/10.1074/jbc.M113.539494

Kobiler, O., Koby, S., Teff, D., Court, D., & Oppenheim, A. B. (2002). The phage λ CII transcriptional activator carries a C-terminal domain signaling for rapid proteolysis. *Proceedings of the National Academy of Sciences*, *99*(23), 14964–14969. https://doi.org/10.1073/pnas.222172499

Koodathingal, P., Jaffe, N. E., Kraut, D. A., Prakash, S., Fishbain, S., Herman, C., & Matouschek, A. (2009). ATP-dependent Proteases Differ Substantially in Their Ability to Unfold Globular Proteins. *Journal of Biological Chemistry*, *284*(28), 18674–18684. https://doi.org/10.1074/jbc.M900783200

Koppen, M., Metodiev, M. D., Casari, G., Rugarli, E. I., & Langer, T. (2007). Variable and Tissue-Specific Subunit Composition of Mitochondrial m-AAA Protease Complexes Linked to Hereditary Spastic Paraplegia. *Molecular and Cellular Biology*, *27*(2), 758–767. https://doi.org/10.1128/MCB.01470-06

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (2017). The ClusPro web server for protein-protein docking.

Nature Protocols, *12*(2), 255–278. https://doi.org/10.1038/nprot.2016.169

Krissinel, E., & Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*, *372*(3), 774–797. https://doi.org/10.1016/j.jmb.2007.05.022

Kühnle, N., Dederer, V., & Lemberg, M. K. (2019). Intramembrane proteolysis at a glance: from signalling to protein degradation. *Journal of Cell Science*, *132*(16), jcs217745. https://doi.org/10.1242/jcs.217745

Kunte, H., Lentzen, G., & Galinski, E. (2014). Industrial Production of the Cell Protectant Ectoine: Protection Mechanisms, Processes, and Products. *Current Biotechnology*, *3*(1), 10–25. https://doi.org/10.2174/22115501113026660037

Lairson, L. L., Henrissat, B., Davies, G. J., & Withers, S. G. (2008). Glycosyltransferases: Structures, Functions, and Mechanisms. *Annual Review of Biochemistry*, *77*(1), 521–555. https://doi.org/10.1146/annurev.biochem.76.061005.092322

Laloux, G., & Collet, J.-F. (2017). Major Tom to Ground Control: How Lipoproteins Communicate Extracytoplasmic Stress to the Decision Center of the Cell. *Journal of Bacteriology*, *199*(21). https://doi.org/10.1128/JB.00216-17

Lanyi, J. K. (1974). Salt dependent properties of proteins from extremely halophilic bacteria. *Bacteriological Reviews*, *38*(3), 272–290. https://doi.org/10.1128/mmbr.38.3.272-290.1974

Larsen, T. M., Benning, M. M., Wesenberg, G. E., Rayment, I., & Reed, G. H. (1997). Ligand-Induced Domain Movement in Pyruvate Kinase: Structure of the Enzyme from Rabbit Muscle with Mg2+, K+, andl-Phospholactate at 2.7 Å Resolution. *Archives of Biochemistry and Biophysics*, *345*(2), 199–206. https://doi.org/10.1006/abbi.1997.0257

Lee, B., & Richards, F. M. (1971). The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology*, *55*(3), 379-IN4. https://doi.org/10.1016/0022-2836(71)90324-X

Lee, S., Augustin, S., Tatsuta, T., Gerdes, F., Langer, T., & Tsai, F. T. F. (2011). Electron Cryomicroscopy Structure of a Membrane-anchored Mitochondrial AAA Protease. *Journal of Biological Chemistry*, *286*(6), 4404–4411. https://doi.org/10.1074/jbc.M110.158741

Lee, S. C., & Pollock, N. L. (2016). Membrane proteins: is the future disc shaped? *Biochemical Society Transactions*, *44*(4). http://www.biochemsoctrans.org/content/44/4/1011

Lee, S., Roh, S. H., Lee, J., Sung, N., Liu, J., & Tsai, F. T. F. (2019). Cryo-EM Structures of the Hsp104 Protein Disaggregase Captured in the ATP Conformation. *Cell Reports*, *26*(1), 29-36.e3. https://doi.org/10.1016/j.celrep.2018.12.037

Lee, S. S., Hong, S. Y., Errey, J. C., Izumi, A., Davies, G. J., & Davis, B. G. (2011). Mechanistic evidence for a front-side, SNi-type reaction in a retaining glycosyltransferase. *Nature Chemical Biology*, *7*(9), 631–638. https://doi.org/10.1038/nchembio.628

Lentzen, G., & Schwarz, T. (2006). Extremolytes: Natural compounds from extremophiles for versatile applications. In *Applied Microbiology and Biotechnology* (Vol. 72, Issue 4, pp. 623–634). Springer.

https://doi.org/10.1007/s00253-006-0553-9

Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B.-K., & Gao, N. (2015). Structure of the eukaryotic MCM complex at 3.8 Å. *Nature*, *524*(7564), 186–191. https://doi.org/10.1038/nature14685

Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., & Adams, P. D. (2017). Polder maps: Improving OMIT maps by excluding bulk solvent. *Acta Crystallographica Section D: Structural Biology*, *73*(2), 148–157. https://doi.org/10.1107/S2059798316018210

Lo Nostro, P., & Ninham, B. W. (2012). Hofmeister Phenomena: An Update on Ion Specificity in Biology. *Chemical Reviews*, *112*(4), 2286–2322. https://doi.org/10.1021/cr200271j

Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, *42*(D1), D490–D495. https://doi.org/10.1093/nar/gkt1178

Lonhienne, T. G. A., Sagulenko, E., Webb, R. I., Lee, K.-C., Franke, J., Devos, D. P., Nouwens, A., Carroll, B. J., & Fuerst, J. A. (2010). Endocytosis-like protein uptake in the bacterium Gemmata obscuriglobus. *Proceedings of the National Academy of Sciences*, *107*(29), 12883–12888. https://doi.org/10.1073/pnas.1001085107

Mahalakshmi, S., Sunayana, M. R., SaiSree, L., & Reddy, M. (2014). yciM is an essential gene required for regulation of lipopolysaccharide synthesis in Escherichia coli. *Molecular Microbiology*, *91*(1), 145–157. https://doi.org/10.1111/mmi.12452

Mahmoud, S. A., & Chien, P. (2018). Regulated Proteolysis in Bacteria. *Annual Review of Biochemistry*, *87*(1), 677–696. https://doi.org/10.1146/annurev-biochem-062917-012848

Marin, K., Huckauf, J., Fulda, S., & Hagemann, M. (2002). Salt-dependent expression of glucosylglycerol-phosphate synthase, involved in osmolyte synthesis in the cyanobacterium synechocystis sp. strain PCC 6803. *Journal of Bacteriology*, *184*(11), 2870–2877. https://doi.org/10.1128/JB.184.11.2870-2877.2002

Marin, K., Stirnberg, M., Eisenhut, M., Krämer, R., & Hagemann, M. (2006). Osmotic stress in Synechocystis sp. PCC 6803: Low tolerance towards nonionic osmotic stress results from lacking activation of glucosylglycerol accumulation. *Microbiology*, *152*(7), 2023–2030. https://doi.org/10.1099/mic.0.28771-0

Marmé, N., Knemeyer, J.-P., Sauer, M., & Wolfrum, J. (2003). Inter- and Intramolecular Fluorescence Quenching of Organic Dyes by Tryptophan. *Bioconjugate Chemistry*, *14*(6), 1133–1139. https://doi.org/10.1021/bc0341324

Marta, K., Marta, G., Adam, U., & Hanna, J. (2007). The significance of Arabidopsis AAA proteases for activity and assembly/stability of mitochondrial OXPHOS complexes. *Physiologia Plantarum*, *129*(1), 135–142. https://doi.org/10.1111/j.1399-3054.2006.00835.x

Martin, A., Baker, T. A., & Sauer, R. T. (2008). Protein unfolding by a AAA+ protease is dependent on ATP-hydrolysis rates and substrate energy landscapes. *Nature Structural & Molecular Biology*, *15*(2), 139–145. https://doi.org/10.1038/nsmb.1380

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., Read, R. J., & IUCr. (2007). *Phaser* crystallographic software. *Journal of Applied Crystallography*, *40*(4), 658–674. https://doi.org/10.1107/S0021889807021206

Miller, J. M., & Enemark, E. J. (2016). Fundamental Characteristics of AAA+ Protein Family Structure and Function. *Archaea*, *2016*, 1–12. https://doi.org/10.1155/2016/9294307

Monroe, N., Han, H., Shen, P. S., Sundquist, W. I., & Hill, C. P. (2017). Structural basis of protein translocation by the Vps4-Vta1 AAA ATPase. *ELife*, *6*. https://doi.org/10.7554/eLife.24487

Moremen, K. W., & Haltiwanger, R. S. (2019). Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nature Chemical Biology*, *15*(9), 853–864. https://doi.org/10.1038/s41589-019-0350-2

Nager, A. R., Baker, T. A., & Sauer, R. T. (2011). Stepwise unfolding of a β barrel protein by the AAA+ ClpXP protease. *Journal of Molecular Biology*, *413*(1), 4–16. https://doi.org/10.1016/j.jmb.2011.07.041

Nandi, D., Tahiliani, P., Kumar, A., & Chandu, D. (2006). The ubiquitin-proteasome system. *Journal of Biosciences*, *31*(1), 137–155. https://doi.org/10.1007/BF02705243

Nishimura, K., Kato, Y., & Sakamoto, W. (2016). Chloroplast Proteases: Updates on Proteolysis within and across Suborganellar Compartments. *Plant Physiology*, *171*(4), 2280–2293. https://doi.org/10.1104/pp.16.00330

Nolden, M., Ehses, S., Koppen, M., Bernacchia, A., Rugarli, E. I., & Langer, T. (2005). The m-AAA Protease Defective in Hereditary Spastic Paraplegia Controls Ribosome Assembly in Mitochondria. *Cell*, *123*(2), 277–289. https://doi.org/10.1016/j.cell.2005.08.003

Novak, Jens F, Stirnberg, M., Roenneke, B. M. E., & Marin, K. (2011). A novel mechanism of osmosensing, a salt-dependent protein-nucleic acid interaction in the cyanobacterium Synechocystis Species PCC 6803. *The Journal of Biological Chemistry*, *286*(5), 3235–3241. https://doi.org/10.1074/jbc.M110.157032

Novak, Jens Florian. (2010). *Regulation der Glukosylglycerol-Phosphat-Synthase aus dem Cyanobakterium Synechocystis sp. PCC 6803*. University of Cologne.

Oberdörfer, Y., Schrot, S., Fuchs, H., Galinski, E., & Janshoff, A. (2003). Impact of compatible solutes on the mechanical properties of fibronectin: A single molecule analysis. *Physical Chemistry Chemical Physics*, *5*(9), 1876–1881. https://doi.org/10.1039/b301022h

Ogura, T., Inoue, K., Tatsuta, T., Suzaki, T., Karata, K., Young, K., Su, L.-H., Fierke, C. A., Jackman, J. E., Raetz, C. R. H., Coleman, J., Tomoyasu, T., & Matsuzawa, H. (1999). Balanced biosynthesis of major membrane components through regulated degradation of the committed enzyme of lipid A biosynthesis by the AAA protease FtsH (HflB) in Escherichia coli. *Molecular Microbiology*, *31*(3), 833–844. https://doi.org/10.1046/j.1365-2958.1999.01221.x

Ogura, T., Whiteheart, S. W., & Wilkinson, A. J. (2004). Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. *Journal of Structural Biology*, *146*(1–2), 106–112. https://doi.org/10.1016/j.jsb.2003.11.008

Okuno, T., Yamanaka, K., & Ogura, T. (2006). An AAA protease FtsH can initiate proteolysis from internal sites of a model substrate, apo-flavodoxin. *Genes to Cells*, *11*(3), 261–268. https://doi.org/10.1111/j.1365-2443.2006.00940.x

Ollivier, B., Caumette, P., Garcia, J. L., & Mah, R. A. (1994). Anaerobic bacteria from hypersaline environments. In *Microbiological Reviews* (Vol. 58, Issue 1, pp. 27–38). American Society for Microbiology. https://doi.org/10.1128/mmbr.58.1.27-38.1994

Opalińska, M., & Jańska, H. (2018). AAA Proteases: Guardians of Mitochondrial Function and Homeostasis. *Cells*, *7*(10), 163. https://doi.org/10.3390/cells7100163

Oren, A. (2011). Thermodynamic limits to microbial life at high salt concentrations. In *Environmental Microbiology* (Vol. 13, Issue 8, pp. 1908–1923). John Wiley & Sons, Ltd. https://doi.org/10.1111/j.1462-2920.2010.02365.x

Oren, A. (2013). Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Frontiers in Microbiology*, *4*(NOV), 315. https://doi.org/10.3389/fmicb.2013.00315

Padfield, D., & Matheson, G. (2020). *nls.multstart: Robust Non-Linear Regression using AIC Scores* (1.1.0). https://cran.r-project.org/package=nls.multstart

Parey, K., Brandt, U., Xie, H., Mills, D. J., Siegmund, K., Vonck, J., Kühlbrandt, W., & Zickermann, V. (2018). Cryo-EM structure of respiratory complex I at work. *ELife*, *7*, e39213. https://doi.org/10.7554/eLife.39213

Park, E., Rho, Y. M., Koh, O., Ahn, S. W., Seong, I. S., Song, J.-J., Bang, O., Seol, J. H., Wang, J., Eom, S. H., & Chung, C. H. (2005). Role of the GYVG Pore Motif of HslU ATPase in Protein Unfolding and Translocation for Degradation by HslV Peptidase. *Journal of Biological Chemistry*, *280*(24), 22892–22898. https://doi.org/10.1074/jbc.M500035200

Perez-Jimenez, R., Garcia-Manyes, S., Ainavarapu, S. R. K., & Fernandez, J. M. (2006). Mechanical unfolding pathways of the enhanced yellow fluorescent protein revealed by single molecule force spectroscopy. *The Journal of Biological Chemistry*, *281*(52), 40010–40014. https://doi.org/10.1074/jbc.M609890200

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera – A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Ponsford, A. H., Ryan, T. A., Raimondi, A., Cocucci, E., Wycislo, S. A., Fröhlich, F., Swan, L. E., & Stagi, M. (2020). Live imaging of intra-lysosome pH in cell lines and primary neuronal culture using a novel genetically encoded biosensor. *Autophagy*, 1–19. https://doi.org/10.1080/15548627.2020.1771858

Poornam, G. P., Matsumoto, A., Ishida, H., & Hayward, S. (2009). A method for the analysis of domain movements in large biomolecular complexes. *Proteins: Structure, Function, and Bioinformatics*, *76*(1), 201–212. https://doi.org/10.1002/prot.22339

Puchades, C., Ding, B., Song, A., Wiseman, R. L., Lander, G. C., & Glynn, S. E. (2019). Unique Structural Features of the Mitochondrial AAA+ Protease AFG3L2 Reveal the Molecular Basis for Activity in Health and Disease. *Molecular Cell*, *75*(5), 1073-1085.e6.

https://doi.org/10.1016/j.molcel.2019.06.016

Puchades, C., Rampello, A. J., Shin, M., Giuliano, C. J., Wiseman, R. L., Glynn, S. E., & Lander, G. C. (2017). Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. *Science*, *358*(6363), eaao0464. https://doi.org/10.1126/science.aao0464

R Core Team. (2020). *R: A language and environment for statistical computing* (4.0.2). R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Ramelot, T. A., Yang, Y., Sahu, I. D., Lee, H.-W., Xiao, R., Lorigan, G. A., Montelione, G. T., & Kennedy, M. A. (2013). NMR structure and MD simulations of the AAA protease intermembrane space domain indicates peripheral membrane localization within the hexaoligomer. *FEBS Letters*, *587*(21), 3522–3528. https://doi.org/10.1016/j.febslet.2013.09.009

Reed, R. H., Warr, S. R. C., Richardson, D. L., Moore, D. J., & Stewart, W. D. P. (1985). Multiphasic osmotic adjustment in a euryhaline cyanobacterium. *FEMS Microbiology Letters*, *28*(3), 225–229. https://doi.org/10.1111/j.1574-6968.1985.tb00796.x

Roenneke, B. M. E. (2014). *Die Glukosylglycerolphosphatsynthase aus dem Cyanobakterium Synechocystis sp. PCC 6803: Mechanismus der Aktivitätsregulation* [University of Cologne]. https://kups.ub.uni-koeln.de/5488/

Roenneke, B., Rosenfeldt, N., Derya, S. M., Novak, J. F., Marin, K., Krämer, R., & Seibold, G. M. (2018). Production of the compatible solute α-d-glucosylglycerol by metabolically engineered Corynebacterium glutamicum. *Microbial Cell Factories*, *17*(1), 94. https://doi.org/10.1186/s12934-018-0939-2

Rotanova, T. V., Melnikov, E. E., Khalatova, A. G., Makhovskaya, O. V., Botos, I., Wlodawer, A., & Gustchina, A. (2004). Classification of ATP-dependent proteases Lon and comparison of the active sites of their proteolytic domains. *European Journal of Biochemistry*, *271*(23–24), 4865–4871. https://doi.org/10.1111/j.1432-1033.2004.04452.x

Sacharz, J., Bryan, S. J., Yu, J., Burroughs, N. J., Spence, E. M., Nixon, P. J., & Mullineaux, C. W. (2015). Sub-cellular location of FtsH proteases in the cyanobacterium Synechocystis sp. PCC 6803 suggests localised PSII repair zones in the thylakoid membranes. *Molecular Microbiology*, *96*(3), 448–462. https://doi.org/10.1111/mmi.12940

Saha, I., & Shamala, N. (2012). Investigating diproline segments in proteins: Occurrences, conformation and classification. *Biopolymers*, *97*(1), 54–64. https://doi.org/10.1002/bip.21703

Saikawa, N., Akiyama, Y., & Ito, K. (2004). FtsH exists as an exceptionally large complex containing HflKC in the plasma membrane of Escherichia coli. *Journal of Structural Biology*, *146*(1–2), 123–129. https://doi.org/10.1016/j.jsb.2003.09.020

Santos, J., Iglesias, V., Santos-Suárez, J., Mangiagalli, M., Brocca, S., Pallarès, I., & Ventura, S. (2020). pH-Dependent Aggregation in Intrinsically Disordered Proteins Is Determined by Charge and Lipophilicity. *Cells*, *9*(1), 145. https://doi.org/10.3390/cells9010145

Sauer, R. T., & Baker, T. A. (2011). AAA+ Proteases: ATP-Fueled Machines of Protein Destruction. *Annual Review of Biochemistry*, *80*(1), 587–612.

https://doi.org/10.1146/annurev-biochem-060408-172623

Sawangwan, T. (2015). Glucosylglycerol on performance of prebiotic potential. *Functional Foods in Health and Disease*, *5*(12), 427. https://doi.org/10.31989/ffhd.v5i12.222

Schäkermann, M., Langklotz, S., & Narberhaus, F. (2013). FtsH-Mediated Coordination of Lipopolysaccharide Biosynthesis in Escherichia coli Correlates with the Growth Rate and the Alarmone (p)ppGpp. *Journal of Bacteriology*, *195*(9), 1912–1919. https://doi.org/10.1128/JB.02134-12

Scharfenberg, F., Serek-Heuberger, J., Coles, M., Hartmann, M. D., Habeck, M., Martin, J., Lupas, A. N., & Alva, V. (2015). Structure and evolution of N-domains in AAA metalloproteases. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2014.12.024

Scheffzek, K., Ahmadian, M. R., Kabsch, W., Wiesmüller, L., Lautwein, A., Schmitz, F., & Wittinghofer, A. (1997). The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science*, *277*(5324), 333–338. https://doi.org/10.1126/science.277.5324.333

Schlieker, C., Weibezahn, J., Patzelt, H., Tessarz, P., Strub, C., Zeth, K., Erbse, A., Schneider-Mergener, J., Chin, J. W., Schultz, P. G., Bukau, B., & Mogk, A. (2004). Substrate recognition by the AAA+ chaperone ClpB. *Nature Structural & Molecular Biology*, *11*(7), 607–615. https://doi.org/10.1038/nsmb787

Schoor, Hagemann, & Erdmann. (1999). Glucosylglycerol-phosphate synthase: target for ion-mediated regulation of osmolyte synthesis in the cyanobacterium synechocystis sp. strain PCC 6803. *Archives of Microbiology*, *171*(2), 101–106. http://www.ncbi.nlm.nih.gov/pubmed/9914306

Schrader, A., Siefken, W., Kueper, T., Breitenbach, U., Gatermann, C., Sperling, G., Biernoth, T., Scherner, C., Stäb, F., Wenck, H., Wittern, K.-P., & Blatt, T. (2012). Effects of Glyceryl Glucoside on AQP3 Expression, Barrier Function and Hydration of Human Skin. *Skin Pharmacology and Physiology*, *25*(4), 192–199. https://doi.org/10.1159/000338190

Schröder, B. A., Wrocklage, C., Hasilik, A., & Saftig, P. (2010). The proteome of lysosomes. *PROTEOMICS*, *10*(22), 4053–4076. https://doi.org/10.1002/pmic.201000196

Schrödinger, L. (2010). *PyMOL - The PyMOL Molecular Graphics System* (1.7). https://pymol.org

Sen, M., Maillard, R. A., Nyquist, K., Rodriguez-Aliaga, P., Pressé, S., Martin, A., & Bustamante, C. (2013). The ClpXP Protease Unfolds Substrates Using a Constant Rate of Pulling but Different Gears. *Cell*, *155*(3), 636–646. https://doi.org/10.1016/j.cell.2013.09.022

Shivanand, P., & Mugeraya, G. (2011). Halophilic bacteria and their compatible solutes - osmoregulation and potential applications. *Current Science*, *100*(10), 1516–1521. https://www.currentscience.ac.in/Volumes/100/10/1516.pdf

Signorile, A., Sgaramella, G., Bellomo, F., & De Rasmo, D. (2019). Prohibitins: A Critical Role in Mitochondrial Functions and Implication in Diseases. *Cells*, *8*(1), 71. https://doi.org/10.3390/cells8010071

Simoes, M. C., Hughes, K. J., Ingham, D. B., Ma, L., & Pourkashanian, M. (2017).

Estimation of the Thermochemical Radii and Ionic Volumes of Complex Ions. *Inorganic Chemistry*, *56*(13), 7566–7573. https://doi.org/10.1021/acs.inorgchem.7b01205

Singh, S., & Darwin, A. J. (2011). FtsH-Dependent Degradation of Phage Shock Protein C in Yersinia enterocolitica and Escherichia coli. *Journal of Bacteriology*, *193*(23), 6436–6442. https://doi.org/10.1128/JB.05942-11

Sleator, R. D., & Hill, C. (2002). Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiology Reviews*, *26*(1), 49–71. https://doi.org/10.1111/j.1574-6976.2002.tb00598.x

Smith, G. R., Contreras-Moreira, B., Zhang, X., & Bates, P. A. (2004). A link between sequence conservation and domain motion within the AAA+ family. *Journal of Structural Biology*, *146*(1–2), 189–204. https://doi.org/10.1016/j.jsb.2003.11.022

Sokratous, K., Hadjisavvas, A., Diamandis, E. P., & Kyriacou, K. (2014). The role of ubiquitin-binding domains in human pathophysiology. *Critical Reviews in Clinical Laboratory Sciences*, *51*(5), 280–290. https://doi.org/10.3109/10408363.2014.915287

Stirnberg, M., Fulda, S., Huckauf, J., Hagemann, M., Krämer, R., & Marin, K. (2007). A membrane-bound FtsH protease is involved in osmoregulation in Synechocystis sp. PCC 6803: the compatible solute synthesizing enzyme GgpS is one of the targets for proteolysis. *Molecular Microbiology*, *63*(1), 86–102. https://doi.org/10.1111/j.1365-2958.2006.05495.x

Story, R. M., & Steitz, T. A. (1992). Structure of the recA protein–ADP complex. *Nature*, *355*(6358), 374–376. https://doi.org/10.1038/355374a0

Strop, P., & Brunger, A. T. (2005). Refractive index-based determination of detergent concentration and its application to the study of membrane proteins. *Protein Science*, *14*(8), 2207–2211. https://doi.org/10.1110/ps.051543805

Suno, R., Niwa, H., Tsuchiya, D., Zhang, X., Yoshida, M., & Morikawa, K. (2006). Structure of the Whole Cytosolic Region of ATP-Dependent Protease FtsH. *Molecular Cell*, *22*(5), 575–585. https://doi.org/10.1016/j.molcel.2006.04.020

Szathmáry, E. (2007). Coevolution of metabolic networks and membranes: the scenario of progressive sequestration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1486), 1781–1787. https://doi.org/10.1098/rstb.2007.2070

Takenaka, F., & Uchiyama, H. (2000). Synthesis of α-D-Glucosylglycerol by α-Glucosidase and Some of Its Characteristics. *Bioscience, Biotechnology, and Biochemistry*, *64*(9), 1821–1826. https://doi.org/10.1271/bbb.64.1821

Takenaka, F., & Uchiyama, H. (2001). Effects of α-D-glucosylglycerol on the in vitro digestion of disaccharides by rat intestinal enzymes. *Bioscience, Biotechnology and Biochemistry*, *65*(7), 1458–1463. https://doi.org/10.1271/bbb.65.1458

Tatsuta, T., Augustin, S., Nolden, M., Friedrichs, B., & Langer, T. (2007). m-AAA protease-driven membrane dislocation allows intramembrane cleavage by rhomboid in mitochondria. *The EMBO Journal*, *26*(2), 325–335. https://doi.org/10.1038/sj.emboj.7601514

Taujale, R., Venkat, A., Huang, L. C., Zhou, Z., Yeung, W., Rasheed, K. M., Li, S.,

Edison, A. S., Moremen, K. W., & Kannan, N. (2020). Deep evolutionary analysis reveals the design principles of fold A glycosyltransferases. *ELife*, *9*. https://doi.org/10.7554/eLife.54532

Tekirdag, K., & Cuervo, A. M. (2018). Chaperone-mediated autophagy and endosomal microautophagy: Jointed by a chaperone. *Journal of Biological Chemistry*, *293*(15), 5414–5424. https://doi.org/10.1074/jbc.R117.818237

Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, NA-NA. https://doi.org/10.1002/jcc.21334

Urbani, A., & Warne, T. (2005). A colorimetric determination for glycosidic and bile salt-based detergents: applications in membrane protein research. *Analytical Biochemistry*, *336*(1), 117–124. https://doi.org/10.1016/j.ab.2004.09.040

Uthoff, M., & Baumann, U. (2018). Conformational flexibility of pore loop-1 gives insights into substrate translocation by the AAA+ protease FtsH. *Journal of Structural Biology*. https://doi.org/10.1016/j.jsb.2018.08.009

van Bloois, E., Dekker, H. L., Fröderberg, L., Houben, E. N. G., Urbanus, M. L., de Koster, C. G., de Gier, J.-W., & Luirink, J. (2008). Detection of cross-links between FtsH, YidC, HflK/C suggests a linked role for these proteins in quality control upon insertion of bacterial inner membrane proteins. *FEBS Letters*, *582*(10), 1419–1424. https://doi.org/10.1016/j.febslet.2008.02.082

Vostrukhina, M., Popov, A., Brunstein, E., Lanz, M. A., Baumgartner, R., Bieniossek, C., Schacherl, M., & Baumann, U. (2015). The structure of Aquifex aeolicus FtsH in the ADP-bound state reveals a C2-symmetric hexamer. *Acta Crystallographica Section D Biological Crystallography*, *71*(6), 1307–1318. https://doi.org/10.1107/S1399004715005945

Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., Quentin, D., Roderer, D., Tacke, S., Siebolds, B., Schubert, E., Shaikh, T. R., Lill, P., Gatsogiannis, C., & Raunser, S. (2019). SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology*, *2*(1), 218. https://doi.org/10.1038/s42003-019-0437-z

Walker, J. E., Saraste, M., Runswick, M. J., & Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO Journal*, *1*(8), 945–951. http://www.ncbi.nlm.nih.gov/pubmed/6329717

Wall, E., Majdalani, N., & Gottesman, S. (2018). The Complex Rcs Regulatory Cascade. *Annual Review of Microbiology*, *72*(1), 111–139. https://doi.org/10.1146/annurev-micro-090817-062640

Wang, J., Song, J. J. J., Seong, I. S. S., Franklin, M. C. C., Kamtekar, S., Eom, S. H. H., & Chung, C. H. H. (2001). Nucleotide-Dependent Conformational Changes in a Protease-Associated ATPase HslU. *Structure*, *9*(11), 1107–1116. https://doi.org/10.1016/S0969-2126(01)00670-0

Wang, K., Jin, M., Liu, X., & Klionsky, D. J. (2013). Proteolytic processing of Atg32 by the mitochondrial i-AAA protease Yme1 regulates mitophagy. *Autophagy*, *9*(11), 1828–1836. https://doi.org/10.4161/auto.26281

Weber, E. R., Hanekamp, T., & Thorsness, P. E. (1996). Biochemical and

functional analysis of the YME1 gene product, an ATP and zinc-dependent mitochondrial protease from S. cerevisiae. *Molecular Biology of the Cell*, *7*(2), 307–317. https://doi.org/10.1091/mbc.7.2.307

Welsh, D. T. (2000). Ecological significance of compatible solute accumulation by micro-organisms: from single cells to global climate. *FEMS Microbiology Reviews*, *24*(3), 263–290. https://doi.org/10.1111/j.1574-6976.2000.tb00542.x

Westphal, K., Langklotz, S., Thomanek, N., & Narberhaus, F. (2012). A Trapping Approach Reveals Novel Substrates and Physiological Functions of the Essential Protease FtsH in Escherichia coli. *Journal of Biological Chemistry*, *287*(51), 42962–42971. https://doi.org/10.1074/jbc.M112.388470

Wikipedia. (2020). *Life*. Wikimedia Foundation. https://en.wikipedia.org/wiki/Life

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., Wilson, K. S., & IUCr. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography*, *67*(4), 235–242. https://doi.org/10.1107/S0907444910045749

Woellhaf, M. W., Hansen, K. G., Garth, C., & Herrmann, J. M. (2014). Import of ribosomal proteins into yeast mitochondria. *Biochemistry and Cell Biology*, *92*(6), 489–498. https://doi.org/10.1139/bcb-2014-0029

Wohlever, M. L., Nager, A. R., Baker, T. A., & Sauer, R. T. (2013). Engineering fluorescent protein substrates for the AAA+ Lon protease. *Protein Engineering, Design & Selection : PEDS*, *26*(4), 299–305. https://doi.org/10.1093/protein/gzs105

Xu, H., & Ren, D. (2015). Lysosomal Physiology. *Annual Review of Physiology*, *77*(1), 57–80. https://doi.org/10.1146/annurev-physiol-021014-071649

Yamada-Inagawa, T., Okuno, T., Karata, K., Yamanaka, K., & Ogura, T. (2003). Conserved Pore Residues in the AAA Protease FtsH Are Important for Proteolysis and Its Coupling to ATP Hydrolysis. *Journal of Biological Chemistry*, *278*(50), 50182–50187. https://doi.org/10.1074/jbc.M308327200

Yang, Y., Gunasekara, M., Muhammednazaar, S., Li, Z., & Hong, H. (2019). Proteolysis mediated by the membrane-integrated ATP-dependent protease FtsH has a unique nonlinear dependence on ATP hydrolysis rates. *Protein Science*, *28*(7), 1262–1275. https://doi.org/10.1002/pro.3629

Yang, Y., Guo, R., Gaffney, K., Kim, M., Muhammednazaar, S., Tian, W., Wang, B., Liang, J., & Hong, H. (2018). Folding-Degradation Relationship of a Membrane Protein Mediated by the Universally Conserved ATP-Dependent Protease FtsH. *Journal of the American Chemical Society*, *140*(13), 4656–4665. https://doi.org/10.1021/jacs.8b00832

Yeung, H. O., Förster, A., Bebeacua, C., Niwa, H., Ewens, C., McKeown, C., Zhang, X., & Freemont, P. S. (2014). Inter-ring rotations of AAA ATPase p97 revealed by electron cryomicroscopy. *Open Biology*, *4*(3), 130142. https://doi.org/10.1098/rsob.130142

Yim, W. W.-Y., & Mizushima, N. (2020). Lysosome biology in autophagy. *Cell Discovery*, *6*(1), 6. https://doi.org/10.1038/s41421-020-0141-7

Yura, T., & Nakahigashi, K. (1999). Regulation of the heat-shock response. *Current Opinion in Microbiology*, *2*, 153–158. https://doi.org/10.1016/S1369-5274(99)80027-7

Zhang, F., Wu, Z., Zhang, P., Tian, G., Finley, D., & Shi, Y. (2009). Mechanism of Substrate Unfolding and Translocation by the Regulatory Particle of the Proteasome from Methanocaldococcus jannaschii. *Molecular Cell*, *34*(4), 485–496. https://doi.org/10.1016/j.molcel.2009.04.022

Zhang, S., & Mao, Y. (2020). AAA+ ATPases in Protein Degradation: Structures, Functions and Mechanisms. *Biomolecules*, *10*(4), 629. https://doi.org/10.3390/biom10040629

Zhang, Y., & Cremer, P. S. (2006). Interactions between macromolecules and ions: the Hofmeister series. In *Current Opinion in Chemical Biology* (Vol. 10, Issue 6, pp. 658–663). Elsevier Current Trends. https://doi.org/10.1016/j.cbpa.2006.09.020

Zhang, Y., Furyk, S., Bergbreiter, D. E., & Cremer, P. S. (2005). Specific ion effects on the water solubility of macromolecules: PNIPAM and the Hofmeister series. *Journal of the American Chemical Society*, *127*(41), 14505–14510. https://doi.org/10.1021/ja0546424

13 DECLARATION

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Ulrich Baumann betreut worden.

Integrierte Teilpublikation:

Uthoff, M., & Baumann, U. (2018). Conformational flexibility of pore loop-1 gives insights into substrate translocation by the AAA+ protease FtsH. *Journal of Structural Biology*. https://doi.org/10.1016/j.jsb.2018.08.009

Matthias Uthoff

14 CV

BIRTHDAY	28.01.1990
ADDRESS	Im Fuchsbau 4 D-51109 Cologne
NATIONALITY	German
ORCID	0000-0002-6477-5505
LINKEDIN	linkedin.com/in/matthias-uthoff
MAIL	matthias.uthoff@alumni.uni-koeln.de

DOCTORATE

Since 2015	Group of Ulrich Baumann, Institute for Biochemistry, University of Cologne *2-year PhD fellowship from the Association of the Chemical Industry e.V.*

EDUCATION

2013 until 2015	M.Sc. Biology at the University of Cologne (1.4) *On the way towards the structure of full-length FtsH*
2010 until 2013	B.Sc. Biology at the University of Cologne Best graduate of the year (1.4) *Prediction of protein localization by conserved co-expression of compartment marker*
2006 until 2009	Carl-Severing-Berufskolleg in Bielefeld Biological Technical Assistant and Abitur (2.0)

PRACTICAL EXPERIENCE – INTERNSHIPS

Winter 2015 (8 weeks)	Structural biochemistry, University Bielefeld; Prof Dr Niemann *Expression and Purification of Helicobacter pylori CagI-159ff*
Winter 2013 (8 weeks)	Institute for Virology, University Hospital Bonn; Dr Müller *Cloning of coronavirus-Spike proteins and study of seroprevalence on clades of coronaviruses*
Sommer 2009 (10 weeks)	Miltenyi Biotec GmbH, Division: R&D in Bergisch Gladbach, Germany *Expression and isolation of recombinant proteins from E. coli*

SKILLS

Languages	German (native), English (Professional working proficiency)
IT	Windows, Linux, MS Word/Excel/PowerPoint, Python, X-ray crystallography software: Phenix, CCP4, Coot, Chimera, Basic knowledge in Relion-3 and Sphire
Softskills	Skills for Project Leaders/Managers (2018), §15 GenTSV (2018), GLP/GMP (2009)

Matthias Uthoff