UPPSALA
UNIVERSITET

UPTEC X 19003
Examensarbete 30 hp
Juni 2019

Purification, functional characterization and crystallization of the PerR peroxide sensor from *Saccharopolyspora erythraea*

Grim Elison Kalman

UPPSALA
UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

Purification, functional characterization and crystallization of the PerR peroxide sensor from *Saccharopolyspora erythraea*

Grim Elison Kalman

This report summarizes the work on the cloning, expression, and purification of PerR, a metal sensing regulator from *Saccharopolyspora erythraea* and the subsequent characterization using small angle X-ray scattering and other biochemical methods.

The report aims to provide an insight into prokaryotic metal homeostasis, provide a better understanding of how PerR works and provide valuable information for the continued work on the crystallization of PerR.

Handledare: Julia Griese
Ämnesgranskare: Robert Gustafsson
Examinator: Jan Andersson
ISSN: 1401-2138, UPTEC X19 003
Tryckt av: Uppsala

Sammanfattning

Vare sig det gäller fysik, kemi eller biologi är strukturen på beståndsdelarna en av de mest fundamentala aspekterna i ett givet system. Detta gäller från små system i jämnvikt, som mellan atomer i en molekyl, till stora system långt ifrån jämvikt som till exempel celler. Det är strukturen som tillsammans med de fyra fundamentala krafterna bestämmer hur saker och ting interagerar med varandra, och kunskap om detta ger därför en djupare förståelse för komplexa samspel mellan olika beståndsdelar i ett system.

Flertalet Nobelpris har, sedan dess instiftande, belönats för framsteg inom områden kopplade till strukturbestämning av makromolekyler. Detta vittnar om betydelsen av dessa metoder och den information de ger för mänskligheten, vetenskapen och inte minst för biologi. Det kanske inte är så konstigt då möjligheten att se mindre och mindre detaljer alltid fascinerat och eggat människors fantasi. Med kraftfulla spektroskopiska metoder som röntgenkristallografi, småvinkel-röntgenspridning och ljusmikroskopi kan vi idag se mer högupplöst än någonsin. Vi behöver inte längre förlita oss på våra högst begränsade sinnen och med dessa tekniker kan man ge svar på det tidslösa uttrycket ”jag tror på det först när jag ser det”.

I den här uppsatsen sammanfattas förstudien av proteinet PerR, en metallbindande transkriptionsfaktor, från den antibiotika-producerande jordartsbakterien *Saccharopolyspora erythraea*, en viktig del av den naturliga floran som återfinns i goda jordförhållanden. Transkriptionsfaktorer är proteiner som reglerar genuttrycket för de gener de kontrollerar och PerR är involverad i metalljons-balansen hos *S. erythraea* och viktig för dess känslighet mot oxidativ stress som uppstår av giftiga biprodukter från cellandningen eller syrerika miljöer.

Studien ämnar att slutligen leda till strukturbestämmandet och klarläggandet av dess mekanistiska funktion med hjälp av bland annat de ovannämnda teknikerna. För att kunna göra detta måste proteinet först produceras i tillräcklig stor mängd och sedan vara tillräckligt rent utan kontaminationer, en inte helt trivial uppgift. Röntgenkristallografi i synnerhet kräver höga proteinkoncentrationer med ofta över 95% renhet vilket ställer höga krav på den praktiserande studenten. Anledningen till detta är att kristallisering av protein är en delikat, närmast slumpmässig process, vilket inte är så konstigt med tanke på att proteiner ofta är stora flexibla molekyler som har svårt att bilda en stabil kristallstruktur.

Produktionen av antibiotika och andra potentiellt användbara ämnen i *S. erythraea* sätts i gång efter en så kallad metabolisk ”switch”, förmodligen som svar på svåra förhållanden såsom under oxidativ stress eller svält. Därför är det viktigt att studera de underliggande mekanismerna för denna process, som PerR har en betydande roll i.

Table of Contents

1 INTRODUCTION 1

2 BACKGROUND 4

- 2.1 Actinobacteria 4
- 2.2 Secondary Metabolites 5
  - *2.2.1 Erythromycin 5*
  - *2.2.2 Geosmin 7*
- 2.3 Metalloproteins 8
  - *2.3.1 Irving-Williams Series 11*
  - *2.3.2 Metal Regulation 12*
  - *2.3.3 Ferric Uptake Regulator Family 13*
  - *2.3.4 Reactive Oxygen Species 16*
  - *2.3.5 PerR 17*
- 2.4 Small-angle X-ray Scattering 19
  - *2.4.1 Theory 19*
  - *2.4.2 Shape Reconstruction 22*

3 METHODS 24

- 3.1 Cloning 24
  - *3.1.1 PCR 25*
  - *3.1.2 Restriction Digest 27*
  - *3.1.3 Dephosphorylation 28*
  - *3.1.4 Ligation 29*
  - *3.1.5 Transformation 30*
  - *3.1.6 Overnight Culture 31*
  - *3.1.7 Colony Plasmid Prep 32*
  - *3.1.8 Plasmid Fingerprinting 34*
- 3.2 Small Scale Expression Tests 36
  - *3.2.1 Transformation 36*
  - *3.2.2 Starter Culture 37*
  - *3.2.3 Expression & Cell Harvest 38*
- 3.3 Expression & Purification 40

3.3.1 Transformation 40
3.3.2 Overnight Culture 42
3.3.3 Expression & Cell Harvest 43
3.3.4 Cell Lysis 45
3.3.5 Tactin-affinity Chromatography 47
3.3.6 TEV Cleavage & Removal of Uncleaved Product 48
3.3.7 Size Exclusion Chromatography 50
3.3.8 Buffer Exchange, Concentration & Storage 51
3.4 Characterization 52
3.4.1 Solubility Screen 52
3.4.2 Solubility Screen with Additives 53
3.4.3 Buffer Optimization 54
3.4.4 Small-angle X-ray Scattering 56
3.5 Crystallization 61

4 RESULTS 62
4.1 Cloning 62
4.2 Expression & Purification 63
4.3 Characterization 66
4.3.1 Buffer Optimization 66
4.3.2 Small-angle X-ray Scattering 70
4.4 Crystallization 81

5 DISCUSSION 81
5.1 Cloning 81
5.2 Expression & Purification 82
5.3 Characterization 83
5.3.1 Small-angle X-ray Scattering 84
5.4 Crystallization 86

6 CONCLUSIONS 86

7 ACKNOWLEDGEMENTS 88

8 REFERENCES 89

1 Introduction

Life, as we know it, arose on Earth about 3.5 billion years ago. At that time, Earth's atmosphere and oceans were strongly reducing environments (Hong Enriquez & Do 2012). This meant that iron was freely available in its soluble, Fe^{2+}, form in the primordial soup for early life to utilize, while other metal ions that today are common in biological systems were tied up in insoluble salts. This put evolutionary pressure on ancient organisms and gave an advantage to those who could utilize iron in their metabolism. Phylogenetic studies and the fact that iron ions are found in many different processes in species from all three domains of life suggest that iron metabolism is an ancient strategy that arose multiple times and evolved convergently (Ilbert & Bonnefoy 2013).

Earth's geochemistry and life have been highly dependent on each other, which is evident in the fossil records throughout the world. For example, one can see that for primitive cyanobacteria the iron-reducing environment was a perfect milieu where the toxic oxygen produced via photosynthesis was quickly taken care of by the free iron ions which were oxidized and formed iron oxide that precipitated and sedimented at the bottom of the sea. When the levels of iron ions in the oceans decreased, the amount of oxygen dissolved in the sea increased, which cyclically led to mass die-offs of cyanobacteria. Even today, there are traces of these layers in the bedrock as dark red iron oxide bands interspersed by gray shale of biological origin, see figure 1 (Schopf 2012).

During the great oxygenation events around 2.4 and 0.7 billion years ago, oxygen levels in the oceans and the atmosphere rapidly increased. While the supply of bio-available iron fell, other previously insoluble metals began to become available. Suddenly, insoluble metallic sulfide salts oxidized into soluble sulfate salts and metal ions such as zinc, copper, and molybdenum became biologically available. These events pushed evolution to find new ways to handle the lack of iron ions and to use these newly available metals in different ways. This in conjunction with the increase in oxygen gave rise to an evolutionary growth spurt and resulted in an explosion of the biodiversity (Stigall 2017). The latter of these events is generally considered to be related to the emergence of multicellular life which speaks for how vital metal ions are for the evolution of life on Earth, see figure 1. Bacteria began to produce so-called siderophores that bind and prevent the oxidation of iron ions in the environment while facilitating their uptake, and some metals were exchanged for other more efficient metals in certain catalyst reactions (Hong Enriquez & Do 2012; Ilbert & Bonnefoy 2013).

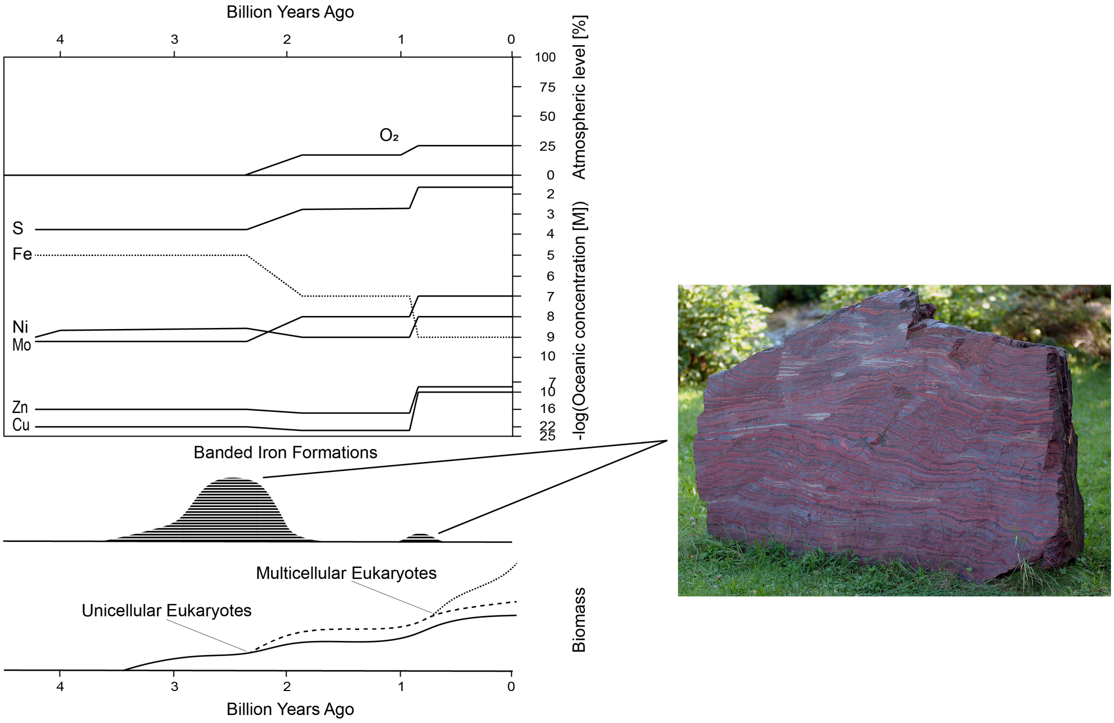

Figure 1: The abundance of elements throughout the evolution of Earth's geochemistry in the atmosphere and the ocean. The occurrences of banded iron formations and the relative biomass of prokaryotes, unicellular eukaryotes, and multicellular eukaryotes. The graphs are approximations based on average values from geological records as well as simple linear models. The graph was reworked and adapted from the article "Insight into the evolution of the iron oxidation pathways" by Ilbert & Bonnefoy (2013). Photo: André Karwath 2005.

Today, these metals and other members of the d-block in the periodic table are recognized as the essential building blocks of the metallome, which in contrast to the proteome, makes up all the metal ions and metal binding species in an organism (Szpunar 2005). The half-filled d-orbitals of these transition metals make their ions very useful within biological systems. Due to their strong bonds and interactions with charged groups and their different oxidation states, they can act as structural elements in proteins, help to catalyze reactions and transport charges. They are also important cofactors for many enzymes and can act as signaling molecules (Dudev & Lim 2014). It is therefore not surprising that metal ion homeostasis is strictly controlled in almost all living organisms known to date. It is so crucial for prokaryotes to control their intracellular metal ion concentrations that one of the defensive strategies used by the innate immune system in bacterial infections is to scavenge iron ions using lipocalins to prevent bacteria from using it in their metabolism (Nasioudis & Witkin 2015). The reverse strategy is also employed as a defense mechanism where the infecting cells are poisoned by metal intoxication by the host, causing mismetallation of proteins with severe toxic effects (Chandrangsu *et al.* 2017).

The key players in prokaryotic metal homeostasis are metal-sensing transcription factors. These metal sensors regulate the expression of target genes encoding, for example, metal transporters, as well as genes involved in the oxidative stress response. This allows the cells to control the uptake, efflux, and storage of metal ions in response to a changing environment (Ma *et al.* 2009).

The metal-sensing transcription factors are divided into three classes depending on their mechanism of action in sensing the concentrations of metal ions. The direct metal-sensing regulators functions by directly binding the metal ion in question, product-sensing regulators by binding to products dependent on the metal ion, and metal-sensing riboswitches with a secondary structure that changes in response to changes in the metal ion concentration, affecting the read-through of the mRNA (Chandrangsu *et al.* 2017).

In this project, the metal-sensing regulator PerR from *Saccharopolyspora erythraea* will be studied. PerR is a member of the ferric uptake regulator protein family which is involved in the regulation of iron and zinc metabolism. PerR belongs to the direct metal-sensing regulators and is involved in the oxidative stress response (Fillat 2014).

In *Bacillus subtilis*, PerR is a homodimer that is stabilized by two structural zinc ions which coordinates several parts of both monomers. The structure of PerR in its apo form is similar to a bumblebee with two thick wings (Traoré *et al.* 2006). When it binds its regulatory metal ion, the wings bend downward as in a wing stroke and act as a clamp that allows PerR to associate to the DNA target sequences readily (Jacquamet *et al.* 2009).

PerR can bind either iron or manganese because of the similarities between these ions and, therefore, can have two different regulatory roles in the cell. It mainly functions as a peroxide sensor, but it also serves as a Fe/Mn ratio sensor. When it binds manganese, PerR functions as a repressor for all PerR regulated genes and the ferric uptake regulation gene *fur*. This form of PerR is insensitive to peroxide, and the presence of it will not induce the genes. Conversely, when it binds iron ions, only peroxide stress response genes are repressed, and the presence of peroxide induces the genes (Helmann 2014). The peroxide reacts with this iron bound form of PerR and oxidizes two coordinating histidines catalyzed by the bound iron (Lee & Helmann 2006). This releases the metal ion, and PerR adopts a more open conformation leading to the dissociation from DNA and the activation of the peroxide stress response genes. This conformational change also exposes binding sites for the protease LonA, ultimately leading to the degradation of PerR (Ahn & Baker 2016).

The exact mechanism behind how the choice of metal ion affects the peroxide sensitivity of PerR and which genes it regulates is not yet understood on a structural level and is the topic of current research. The specific role of PerR in *S. erythraea* and its mechanism of action is currently not determined, and this report has the aim to help shed some light on the matter.

Given the central role of metal ions in so many vital processes in biological systems, a deeper understanding of how metal ions are regulated in bacteria would potentially enhance the efficiency of fermentation in large-scale drug production, the development of healthy soil conditions for agriculture, increase the knowledge of pathogenic bacteria and their interaction with their host cells and unlock even more types of controllable operator and repressor systems to control the expression of genes.

2 Background

Saccharopolyspora erythraea is a common filamentous gram-positive soil bacterium which was first described by Nobel laureate Waksman (1919). It was isolated and described as part of the preliminary work that laid the foundation for his research on antibiotic-producing actinomycetes which he later received the Nobel Prize for.

2.1 Actinobacteria

The actinobacteria, which *S. erythraea* belongs to, are an interesting phylum of Gram-positive bacteria which are very important in the carbon cycle. Because of their saprophytic lifestyle, they play an essential role in humus formation and the release of nutrients in the soil through the decay of biological matter. Like fungi, these bacteria form mycelium networks and have a complex morphological differentiation throughout their life cycle (Barka *et al.* 2016). The majority of the species in this family live in complex symbiotic relationships with fungi, sponges, plants, and insects. Many also form microbial communities and compete over the highly limited resources in their ecological niches. There is a lot of evidence pointing to their unique lifestyle as being the source of the plethora of secondary metabolites that these bacteria produce. The actinomycetes together provide, for example, two-thirds of all known antibiotics, many immunosuppressants, and substances with antifungal and antiworm properties (van der Meij *et al.* 2017).

2.2 Secondary Metabolites

Secondary metabolites are substances not linked to the normal development or growth of an organism. Some of these substances are byproducts of the normal metabolism, but many have a function as signaling molecules or as a defense against competing microorganisms. In the symbiotic relationship between eukaryotes and actinobacteria, these antimicrobial substances protect the host against pathogenic microorganisms in exchange for complex sugars and other biomolecules (van der Meij *et al.* 2017). Some substances have also been shown to act as a growth hormone for host plants and stimulate root formation (Zhao *et al.* 2018). This reciprocal exchange appears to be established in the early stages of the development of plants, which already as a seedling leak up to 30-60% of all the products formed by photosynthesis into the surrounding soil. The specific mixture of amino acids, sugars, phenols, and secondary metabolites released from the plants forms a cocktail that attracts certain actinomycetes and it appears that many of these secondary metabolites are only produced in high amounts as a response to host specific signals (van der Meij *et al.* 2017; Barka *et al.* 2016). It has been shown that *Arabidopsis thaliana* produces and secretes the phytohormone salicylic acid in response to bacterial infections or pests which may act as a signal to attract actinobacteria to alleviate the symptoms by their antimicrobial secondary metabolites (Lebeis *et al.* 2015).

2.2.1 Erythromycin

S. erythraea produces many of these secondary metabolites and has at least 25 gene clusters linked to their production (Oliynyk *et al.* 2007). Perhaps the most famous of these is the antibiotic erythromycin, see figure 2, that is listed among the WHO Model List of Essential Medicines.

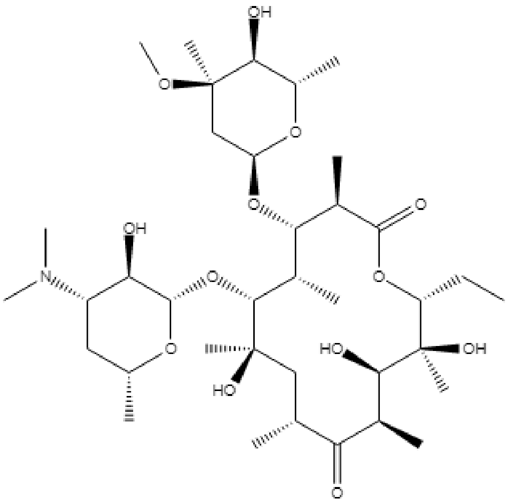

Figure 2: Erythromycin A. The lower right portion of the structure is the macrocyclic lactone ring. To this ring structure, cladinose and desosamine are attached via glycosidic bonds at C3 and C5 respectively. The macrocyclic lactone ring has been used as a starting point for the development of other macrolide antibiotics.

Erythromycin belongs to a class of substances called macrolides. These substances have a macrocyclic lactone ring with one or more sugar groups attached and exert their antibacterial effect by binding to the exit tunnel of the prokaryotic 50S ribosomal subunit, thus obstructing the newly produced peptide chain, which inhibits the protein synthesis, see figure 3. This is considered to be mainly bacteriostatic rather than bacteriocidal (Jelić & Antolović 2016).

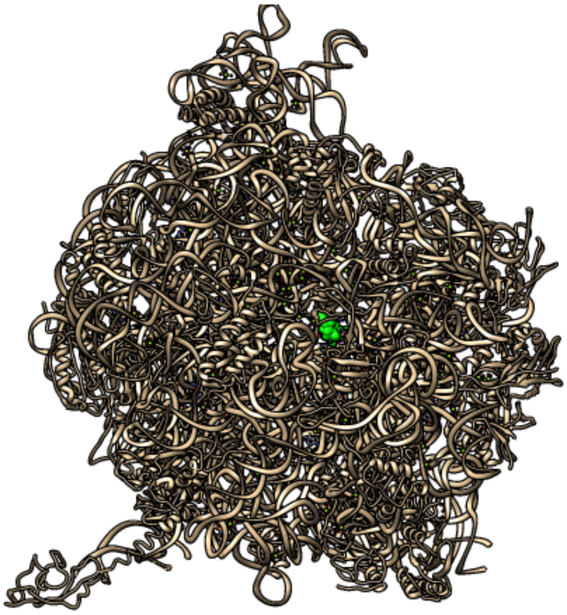

Figure 3: Crystal structure of the Escherichia coli *ribosome bound to erythromycin. The structure is oriented so that the ribosomal exit tunnel is along the normal of the plane. Erythromycin can be seen in the exit tunnel of the 50S subunit as a green space-fill model. PDB ID: 4V7U. Visualized with UCSF Chimera.*

Erythromycin has been known since the 1950s and generated many new second-generation semi-synthetic macrolides with better pharmacokinetics and stability. Unfortunately, these types of antibiotics have been shown to more easily cause antibiotic resistance due to their mechanism of action through the modification of the 23S rRNA in the 50S subunit (Jelić & Antolović 2016). A lot of resources were dedicated to finding new derivatives that were not as sensitive to antibiotic resistance, and a number of third-generation macrolides were produced. Unfortunately, they proved to have severe side effects, and the effort was abandoned. However, macrolides have recently received renewed interest from the research community after Seiple *et al.* (2016) succeeded in synthesizing macrolides completely *de novo* (Dinos 2017).

2.2.2 Geosmin

Another characteristic secondary metabolite produced by many actinomycetes, including *S. erythraea*, is geosmin, see figure 4. Geosmin is one of the substances giving rise to petrichor, the distinct, slightly metallic fragrance of soil that is exuded during rainfall or by newly plowed fields. No source could be found to support this, but it is evident that *S. erythraea* produces this metabolite, partly because of the strong soil odor that arises from the cultivation of *S. erythraea* and partly because *S. erythraea* has the entire metabolic pathway required to produce geosmin (Jiang *et al.* 2007; Oliynyk *et al.* 2007). It is not yet known why these bacteria produce geosmin, but it may serve as a signal molecule. Interestingly in the fruit fly *Drosophila melanogaster* a preserved olfactory nerve pathway dedicated exclusively to geosmin has been identified. When the receptors for geosmin are activated, the nerve signal completely overrides all other olfactory stimuli, warning the fly of potentially dangerous microorganisms causing the fly to avoid eating and laying its eggs there (Stensmyr *et al.* 2012).

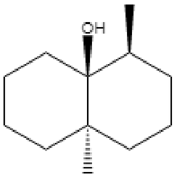

Figure 4: Geosmin. Humans are incredibly sensitive to the odor of geosmin and can detect it at values as low as only 5 parts per million almost rivaling sharks ability to detect blood in water. In terms of volume, this is like detecting a handball in the Ericsson Globe in Stockholm.

2.3 Metalloproteins

Almost half of all enzymes require at least one metal ion to function correctly. Metal ions can have several different roles within biological systems. Bound to proteins, metal ions increases the diversity and functionality of proteins that would otherwise have been limited by the chemistry of the residues from the naturally occuring amino acids. They can act as structural elements so that the enzyme folds properly, or help to catalyze reactions as Lewis acids or in redox-linked processes (Waldron *et al.* 2009).

A classic example where metal ions function as structural elements are so-called Zinc fingers. A Zinc-finger is a structural protein motif that often occurs in DNA-binding proteins where the zinc-ion coordinates two parts of the protein, usually two to three beta-strands, and an alpha-helix, so it folds over itself forming the shape of a “c”. Four amino acids, usually two cysteines and two histidines, bind the zinc from each side in a tetrahedron configuration, stabilizing the fold, see figure 5 (Razin *et al.* 2012).

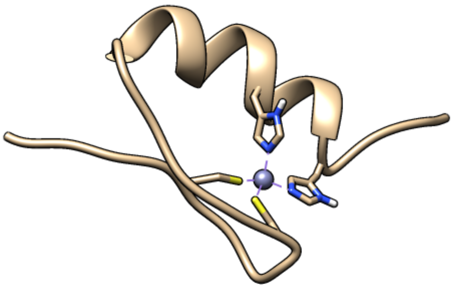

Figure 5: A zinc finger domain from transcription factor SP1F2, minimized average structure from Solution NMR. In addition to the four amino acids that coordinate the zinc atom in a tetrahedral manner, shown in purple, two hydrophobic residues at the bend are pointing inwards (not shown in the picture) which helps to stabilize the fold. The rest of the residues point outwards and are free to interact with other ligands, for example, DNA. PDB ID: 1SP2. Visualized with UCSF Chimera.

In the electron transport chain, which is the last step in the respiration, charges are transported in the inner membrane of the mitochondria by, among other proteins, cytochrome C. Cytochrome C is a heme-binding protein, see figure 6, that is capable of being reversibly oxidized and can thus transport charges. The iron in the heme-group is reduced to Fe^{2+} which is then oxidized back to Fe^{3+} when the electron is delivered to the next protein in the chain (Hine & Martin 2015).

Other important proteins involved in electron transport and several other processes are the so-called iron-sulfur proteins. They contain Fe-S clusters which can, in addition to transporting charges, reduce disulfides, donate sulfur, *et cetera*. Some also have a structural role in certain proteins. These proteins are relatively common and can be found in all branches of life. One example is Ferredoxin which transports charges in, for example, the photosynthesis (Johnson *et al.* 2005).

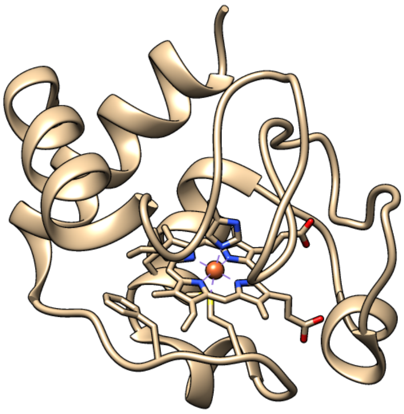

Figure 6: Crystal structure of cytochrome C from bovine heart. The heme group that is covalently bound via disulfide bonds can be seen as a disk surrounding the iron atom shown in orange which is coordinated in a octahedral fashion. PDB ID: 2B4Z. Visualized with UCSF Chimera.

Metal ions are also important catalysts in many enzymes. For example, in DNA polymerase, magnesium ions are needed as a cofactor to catalyze the elongation of DNA. In the catalytic center there are two magnesium ions. The first lowers the pKa value of the hydroxy group from the 3′-end which protonates an adjacent aspartic acid, the other positions the incoming nucleotide correctly and attracts the electron cloud of the triphosphate groups alleviating the steric hindrance and electronic effects on the incoming deprotonated hydroxy group. This allows a nucleophilic attack of the hydroxy group on the innermost phosphate group, see figure 7 (Perera *et al.* 2017).

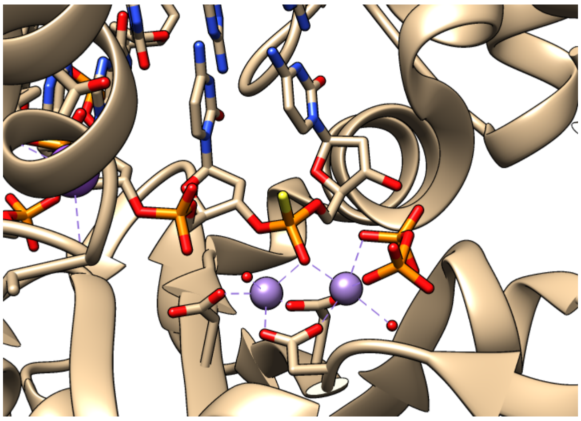

Figure 7: Crystal structure of DNA polymerase β in complex with DNA. Here the active center where a synthetic nucleotide has just been incorporated is shown. The catalytic magnesium ion to the left shown in purple lowered the pKa value of the 3'-hydroxy group which was directly above (now participating in a phosphodiester bond) making it more nucleophilic. The cleaved pyrophosphate group can still be seen coordinated to the magnesium ion in the lower right of the picture which correctly positioned the nucleotide phosphate tail to allow the nucleophilic attack from the 3'-end. PDB ID: 5U9H. Visualized with UCSF Chimera.

2.3.1 Irving-Williams Series

It has been found that the stability of metal ion complexes often follows the so-called Irving-Williams series. That is, the stability depends solely on the metal ion in complex with the following order Mn < Fe < Co < Ni < Cu > Zn, probably due to their ligand field stabilization energies (Irving & Williams 1948; Waldron *et al.* 2009). In the case of metal binding proteins, the binding pocket itself may give some preference to a particular metal ion. The metal ion ligands can selectively bind metal ions with a specific charge, and the ligand geometry can provide a higher affinity to specific metal ions, but due to the flexible nature of proteins, wrong metal ions can still be incorporated. This poses a major problem as it can lead to the mismetallation of proteins, and once a metal ion high up in the Irving-Williams series has been aquired, it can be difficult to displace it (Tottey *et al.* 2008). This can greatly affect the function of the protein with often severe consequences. In the case of zinc poisoning, for example, the electron transport chain is disrupted, probably due to the mismetallation of cytochrome oxidase. Mismetallation of metalloregulators can also cause an inappropriate response which, at worst, compromises the metal ion homeostasis of the cell (Chandrangsu *et al.* 2017). Many metalloproteins have, therefore, been under selective pressure to pick up correct metal ions in a specific cellular environment (Tottey *et al.* 2008).

2.3.2 Metal Regulation

It is vital for the cell to maintain correct metal ion concentrations and to control the availability of metal ions. This is done by metal sensors that control and regulate the metal ion homeostasis mainly by up and down-regulation of genes linked to the intake, efflux or storage of metal ions. These metal sensors can also control alternative metabolic pathways that do not require the missing metals or the release of stored metal ions. In this way, a cell can maintain a stable intracellular concentration of the various metal ions in a fluctuating environment (Frawley & Fang 2014). This is especially true for pathogenic bacteria whose virulence often depends on their metal ion homeostasis. This is partly due to the fact that the host initially constitutes an excellent environment and a rich source of metal ions, but as the infection continues, the bacteria are exposed to toxic amounts of metal ions and a war of attrition orchestrated by the host's immune system (Palmer & Skaar 2016).

One way to selectively regulate the metal ion acquisition of proteins is through carrier proteins. In the cytosol, this is achieved using metallochaperones which deliver the correct metal ions to the metalloproteins. However, for the vast majority of metalloproteins, metallochaperones are missing and metal ion specificity is instead governed by where the protein is folded. In the case of CucA and MncA, the most abundant copper and manganese binding proteins in the cyanobacterium *Synechocystis sp. PCC6803*, both bind copper *in vitro* in accordance with the Irving-Williams series. They bind their metal ion via the same amino acids in a cupin fold but have different export pathways out to the periplasm where the proteins are transported. *In vivo*, CucA is exported via the Sec-pathway which prevents the folding of the protein until it reaches the periplasm while MncA is allowed to fold in the cytoplasm before being transported. In this way, MncA and CucA can pick up the right metal ion when they fold in different parts of the cell (Tottey *et al.* 2008).

The observed affinities of the various metal sensors for their metal ions have been used to predict the intracellular concentration of the different metal ions, and consistently metals high up on the Irving-Williams series are suitably kept at low levels in the cytosol. In *E. coli*, for example, free copper ions are presumed to be maintained at a concentration corresponding to far less than one molecule per cell available in the rapidly exchanging, accessible pool otherwise known as the labile pool (Waldron *et al.* 2009; Changela *et al.* 2003).

How the metal sensors sense the metal ion concentrations can be divided into three main categories depending on whether they are direct sensing, product sensing, or metal binding riboswitches (Chandrangsu *et al.* 2017).

Direct metal sensing regulators, as the name suggests, react directly to the binding of the metal ion, which drives a conformational change that allows the metal regulator to exert its regulatory effect, usually by binding to an operator, blocking the transcription or distorting the secondary structure of the promoter so that its strength increases. An example of a direct metal sensing regulator is Zur. Zur detects the zinc concentration by binding two zinc ions sequentially, causing a conformational change that causes Zur to bind its DNA binding site. This represses the *znuACB genes*, which encode a zinc importer (Chandrangsu *et al.* 2017).

Indirect metal sensing regulators work by measuring the metal ion concentration by proxy, giving an indirect reading of the metal ion concentration. Irr is such a regulator that represses the storage of iron, the biosynthesis of heme and other iron binding proteins. Irr binds charged heme groups and under severe conditions when there is a shortage of iron and consequently uncharged heme, Irr is active, but when there is plenty of iron, Irr binds the charged heme groups which triggers the degradation of Irr (Chandrangsu *et al.* 2017).

The perhaps most interesting metal sensing regulators are the so-called metal sensing riboswitches. They are mRNAs that change their secondary structure in response to the binding of specific metal ions. In the case of the *MgtE* riboswitch in *B. subtilis*, the anti-terminator binding to the terminator is blocked when Mg^{2+} binds to the so-called "magnesium-box". This causes the terminator to become functional, and the transcription is stopped (Chandrangsu *et al.* 2017).

2.3.3 Ferric Uptake Regulator Family

The ferric uptake regulator family (FUR) is a family of proteins that control the metal metabolism in many bacteria. The family includes regulators that have evolved to sense different types of metals, such as Fur and Zur involved in the iron and zinc metabolism respectively, and PerR, involved in the peroxide stress response. It is a diverse family that occurs in at least 4,000 different species of bacteria and archaea. Commonly, they all have the conserved histidine-rich sequence HHHXHXXCXXC, involved in the binding of their regulatory metal ion (Fillat 2014), see figure 8.

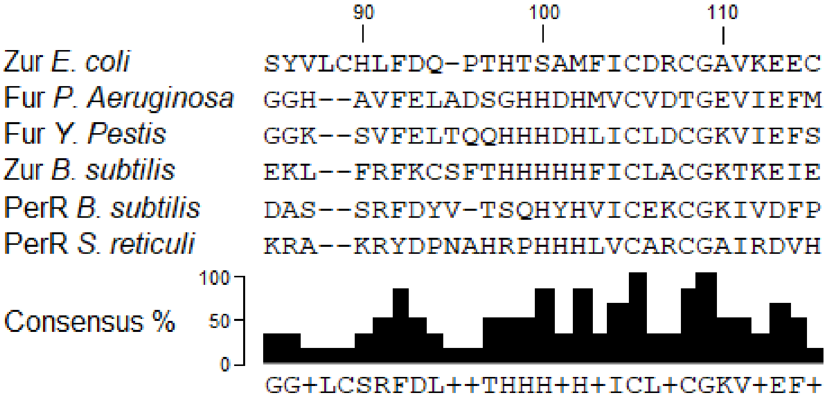
90
100
110
Zur E. coli
Fur P. Aeruginosa
Fur Y. Pestis
Zur B. subtilis
PerR B. subtilis
PerR S. reticuli
SYVLCHLFDQ-PTHTSAMFICDRCGAVKEEC
GGH--AVFELADSGHHDHMVCVDTGEVIEFM
GGK--SVFELTQQHHHDHLICLDCGKVIEFS
EKL--FRFKCSFTHHHHHFICLACGKTKEIE
DAS--SRFDYV-TSQHYHVICEKCGKIVDFP
KRA--KRYDPNAHRPHHHLVCARCGAIRDVH
Consensus %
100
50
0
GG+LCSRFDL++THHH+H+ICL+CGKV+EF+

Figure 8: Multiple sequence alignment of FUR proteins from different species. The conserved consensus sequence can be seen at the bottom and has the histidine-rich sequence HHHX-HXXCXXC. The amino acid position of the top sequence is indicated. Proteins were aligned with Clustal Omega and the sequences downloaded from the PFAM repository using Jalview, (Madeira et al. 2019; Waterhouse et al. 2009)

Structurally, they have two domains, one N-terminal DNA binding domain and one C-terminal dimerization domain linked via a flexible inter-domain region where the conserved sequence is located. This section acts as a hinge that closes when it binds a metal ion which causes the protein in its dimeric form to adopt a clamp-like structure that can “grasp” and make favourable contacts with the DNA, see figure 9.

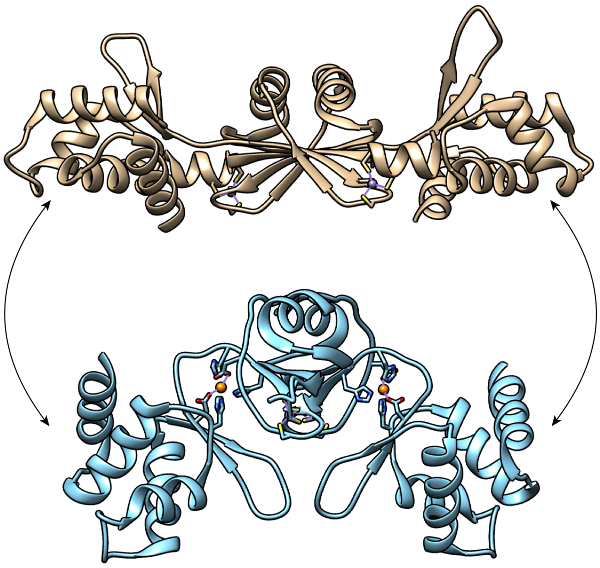

Figure 9: Crystal structures of PerR from B. subtilis *in its apo and holo conformation. The tan colored elongated version is the apo form of PerR with two structural zinc stabilizing the dimerization. The slightly bent blue version is the holo form and, in addition to the two structural zinc, has two manganese ions bound to the hinge region. This is where the conserved histidine-rich sequence HHHXHXXCXXC of the FUR-family is located which upon binding a metal ion brings the DNA-binding domain and the dimerization domain closer resulting in a downward wing flap. This conformational change allows it to associate with its DNA target sequence. PDB ID: 2FE3 (apo), 3F8N (holo). Visualized with UCSF Chimera.*

Very simplified, members of the FUR family binds to palindromic AT-rich sequences found in the promoters they regulate as a homodimer in the presence of their metal ion corepressors. The DNA interaction itself is facilitated by two winged-helix DNA binding domains, one from each monomer that interacts with the major groove through a helix-turn-helix motif (HTH). In some cases, for example, in the case with Fur from *Magnetospirillum gryphiswaldense*, it has been shown that DNA recognition is not only facilitated by the direct readout of the bases in the major groove but also by shape recognition by interacting with the minor groove through electrostatic potentials, see figure 10 (Fillat 2014; Deng *et al.* 2015).

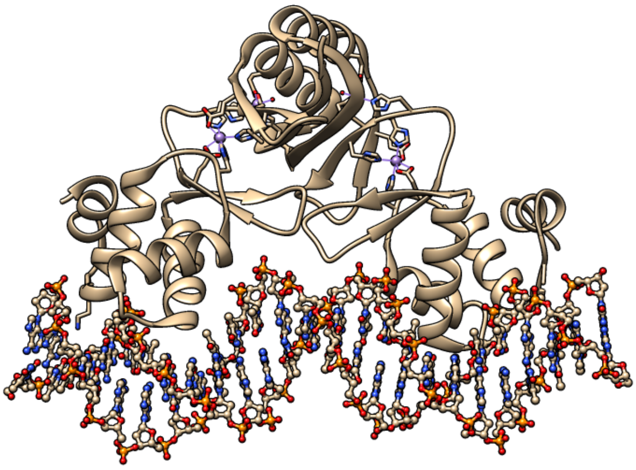

Figure 10: Crystal structure of Fur from Magnetospirillary Gryphiswaldense *associated with the feoAB1 operator. The protein-DNA interaction occurs mainly through base-specific interactions between the HTH motif and the major groove of the DNA. It has been suggested that a lysine, which can be seen in the left corner, which is lowered into the minor groove and interacts with the DNA through electrostatic potentials is involved in shape recognition. PDB ID: 4RB3. Visualized with UCSF Chimera.*

2.3.4 Reactive Oxygen Species

Since life developed for a long time under anaerobic conditions, there was no evolutionary pressure to evolve enzymes with little reactivity with oxygen. Many metabolic pathways were developed and already in place before photosynthesis occurred in the evolution of life, and because of this, many cellular systems are sensitive to oxygen and reactive oxygen species (ROS). To combat this, rather than reinventing new pathways, other methods for dealing with this evolved, usually with a mechanism of action involving the scavenging and sequestering of ROS. The source of oxidative damage in cells is primarily due to these ROS oxidizing redox cofactors in active sites, which deactivates them. ROS are mainly generated in the electron transport chain where exposed redox groups can donate electrons to intracellular oxygen. Oxidation events like this creates either superoxide or hydrogen peroxide which are both harmful to the cell (Imlay 2008).

Hydrogen peroxide can react with reduced iron ions in the so-called Fenton reaction to form hydroxyl radicals, see the reaction formula below, which are very reactive and have a short half-life because of their unpaired electron configuration. Since many of these iron ions are associated with DNA or bound to catalytic sites in enzymes, this can lead to mutations or the inactivation of enzymes with deleterious results (Imlay 2008).

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + HO \cdot + OH^-$$

$$Fe^{3+} + H_2O_2 \rightarrow Fe^{2+} + HOO \cdot + H^+$$

Interestingly, it has recently been reported that Cytochrome C peroxidase in *E. coli* uses hydrogen peroxide as an electron acceptor in the electron transport chain under anaerobic conditions. Hydrogen peroxide, which is otherwise considered harmful to the cell appears to be used under anoxic conditions for cell survival, proving that bacteria have found ingenious ways to make use of this otherwise unwanted byproduct (Kikhney & Svergun 2015).

2.3.5 PerR

PerR is a metal-dependent peroxide sensor from the FUR family which controls genes linked to the peroxide stress response. In *B. subtilis* PerR regulates in addition to itself the genes *katA*, *mrgA*, *ahpCF*, *zosA*, *hemAXCDBL* and *fur* (Faulkner *et al.* 2012).

KatA is a catalase, a type of enzyme that breaks down hydrogen peroxide into water and oxygen. Likewise, the alkyl hydroperoxide reductase ahpC catalyzes the reduction of hydrogen peroxide and organic peroxides into alcohols and water (Faulkner *et al.* 2012).

MrgA is involved in the storage of iron and Fur limits the iron uptake, which is essential during peroxide stress since free iron easily generates toxic radicals when reacting with peroxides causing oxidative cell damage (Bresgen & Eckl 2015). The genes encoded by *hemAXCDBL* are involved in the heme biosynthesis, further limiting the availability of iron ions in the labile pool (Faulkner *et al.* 2012).

ZosA is a zinc transporter in the zinc uptake system which during oxidative stress probably serves to protect thiols from oxidation, and prevents protein sulphydryl groups from forming disulfides (Gaballa & Helmann 2002).

Like the rest of the proteins in the FUR family, PerR has two domains, one DNA binding domain and one dimerization domain. In between is a hinge region with a binding site for a regulatory metal ion, either iron or manganese because of their chemical and physical similarities. At the C-terminal, there is another metal binding site with a preference for zinc. The zinc has a structural role and helps to stabilize the dimer, which is shown by the treatment with diamide or peroxide, which releases the zinc ion and causes the dissociation of the dimer (Fillat 2014).

Bound to iron, PerR associates with its DNA target sequences and represses the genes involved in the peroxide stress response system. This form is sensitive to hydrogen peroxide which can react with the bound iron in the active site causing an irreversible metal catalyzed histidine oxidation of the histidines coordinating the iron ion. This results in the formation of 2-oxo-histidine, rendering the metal site inactive. This causes PerR to dissociate from the DNA, and the peroxide stress genes are activated. When bound to manganese PerR is resistant to the oxidation of hydrogen peroxide and represses the peroxide stress genes regardless of the presence of it. This means that PerR is sensitive to the ratio between the intracellular iron and manganese concentrations, which dictates the metallation state of PerR, making PerR essential for the iron-manganese homeostasis (Fillat 2014).

Interestingly, PerR:Fe regulates only a subset of the genes that can be regulated by PerR, namely, the ones involved in the peroxide stress response system, reflecting the need to control the peroxide sensitivity while intracellular iron levels are high in the cell. At the same time, PerR:Mn suppresses all of the genes, including fur and PerR itself, leading to an increase in the iron uptake, see figure 11 (Helmann 2014).

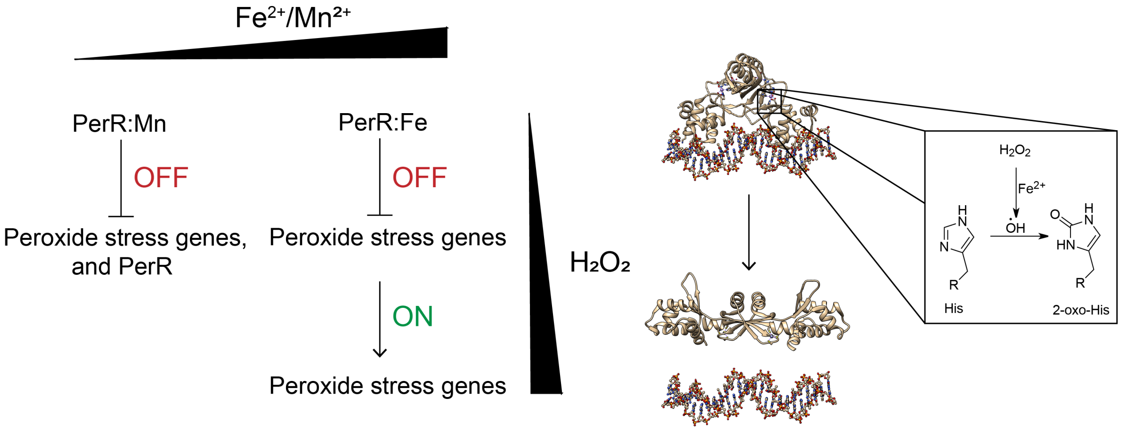

Figure 11: PerR's activity in response to differences in iron and manganese ion concentrations and hydrogen peroxide. When the concentration of manganese ions is high and iron ions low, PerR:Mn represses the peroxide stress genes and itself regardless of the level of hydrogen peroxide. When the conditions are reversed, PerR:Fe represses the peroxide stress genes and in response to increasing levels of hydrogen peroxide, is oxidized, deactivating PerR and derepressing the peroxide stress genes. PDB ID: 2FE3, 4RB3.

The dissociation constant of iron in complex with PerR in *B. subtilis* is roughly 28 times lower than for manganese (Helmann 2014). Cellular conditions reflected by a high concentration of PerR:Mn, therefore, imply low levels of iron and high levels of manganese in the labile pool. Since manganese is an antioxidant, cells under these conditions should consequently already be quite resistant to oxidative damage and therefore do not need to activate the peroxide stress genes. In fact, in response to oxidative stress, the manganese uptake is increased, and the iron ions in some enzymes are replaced by the less effective but oxidation resistant manganese as a cofactor (Helmann 2014; Joseph A. Cotruvo &

Stubbe 2012). Moreover, manganese can work as an effective scavenger of ROS when in complex with different metabolites, especially orthophosphates (Culotta & Daly 2013).

2.4 Small-angle X-ray Scattering

Small-angle X-ray scattering (SAXS) is a biophysical method that can be used to study macromolecules in solution. It relies on the elastic scattering of X-rays by electrons of single atoms in a sample. The scattered X-rays from the different atoms create a pattern that carries structural information about the sample, which can be recorded. Because of the random orientation of the molecules in the sample, this gives rise to an isotropic scattering pattern. Thus, the pattern intensities can be radially averaged and be represented as one-dimensional curves.

From this curve, several important features of the sample can be extracted like the radius of gyration R_{g}, maximum dimension D_{max}, overall flexibility, shape, and the useful distance distribution function $p(r)$ describing the pairwise distribution of inter-atom distances (Kikhney & Svergun 2015).

2.4.1 Theory

To understand how SAXS works, one needs to understand the basics of how a single atom scatters incoming X-rays. When an incoming X-ray beam hits the atom, the oscillating electromagnetic field rocks the electrons back and forth as a driven oscillator which in turn generates an electromagnetic wave; phase shifted with pi-radians in a process called Thomson scattering. When the incoming X-ray hits the atom, the electric field excites different parts of the electron density cloud as it passes over it, leading to a specific phase lag between the oscillating electrons that acts as point wave sources. Thus, the resulting wave in a particular scattering direction will be a super-positioned wave from all wave sources within the atom that oscillate in the same frequency but at different phases (Bernhard 2009).

It turns out that the resulting wave can be described as the Fourier transform of the electron density of the atom, $\rho(\mathbf{r})$, in a certain direction described by the scattering vector $\mathbf{q} = 4\pi sin(\theta)/\lambda$ where θ is scattering angle and λ the wavelength. Since $\rho(\mathbf{r})$ is a continuous function this becomes an integral known as the atomic scattering factor. Bold lettering will be used to indicate a vector throughout this text.

$$f(\mathbf{q}) = \int_V \rho(\mathbf{r}) e^{i\mathbf{qr}} d\mathbf{r}$$

Simply speaking this is the probability of a scattering event happening at the relative distance **r** times the partial wave, $e^{i\mathbf{qr}}$, with a certain phase lag, **qr**, resulting from this event. These partial waves weighted by their probability of occurring is then summed over all possible points in space spanned by the volume of the atom, V. Now taking all the atoms in a molecule into account one can repeat the argument above and sum all the resulting waves from each atom in the molecule, phase shifted relative to each other and obtain the structure factor or scattering amplitude, $F(\mathbf{q})$.

$$F(\mathbf{q}) = \sum_{i=1}^{N} f_i e^{i\mathbf{qr_i}}$$

The phase factor $e^{i\mathbf{qr}}$ describes each partial wave emanating from each point source within the molecule in its exponential vector form and f_i is the atomic scattering factor.

In simple terms, this equation can be understood as the scaling of the amplitude of each wave point source within a molecule with the strength of the scattering signal for each atom, summed over all atoms in the molecule.

The amplitude is unfortunately unavailable to us, but the intensity can be recorded on a detector. Intensity is proportional to the amplitude as $I = F(\mathbf{q})\overline{F}(\mathbf{q})$; therefore the intensity can be written in terms of the structure factor as

$$I(\mathbf{q}) = F(\mathbf{q})\overline{F}(\mathbf{q}) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(\mathbf{q}) f_j(\mathbf{q}) e^{i\mathbf{q}(r_i - r_j)}$$

and since the molecules are in solution and thus have random orientation the phase factor is spherically averaged, $\left\langle e^{iq(r_i - r_j)} \right\rangle = \frac{sin(qr_{ij})}{qr_{ij}}$, and the equation becomes Debye´s formula for scattered intensity

$$I(\mathbf{q}) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(\mathbf{q}) f_j(\mathbf{q}) \frac{sin(\mathbf{q}r_{ij})}{\mathbf{q}r_{ij}}$$

where $r_{ij} = |r_i - r_j|$ is the intramolecular distance between scatterers. This formula is useful to calculate low-resolution scattering curves from bead models and can be used in the shape reconstruction. In the formula above the sum over all atoms in a molecule can be substituted for the pair distribution function, $p(r)$, describing the distribution of the pairwise distances of the atoms within the molecule. This yields the equation

$$I(\mathbf{q}) = 4\pi \int_0^{D_{max}} p(r) \frac{sin(\mathbf{q}r)}{\mathbf{q}rdr}$$

Where D_{max} is the maximum intramolecular distance (Koch *et al.* 2003).

Different shapes will have different scattering pattern as seen in figure 12 below, generated by the FoXS tool using Debye's equation (Schneidman-Duhovny *et al.* 2016).

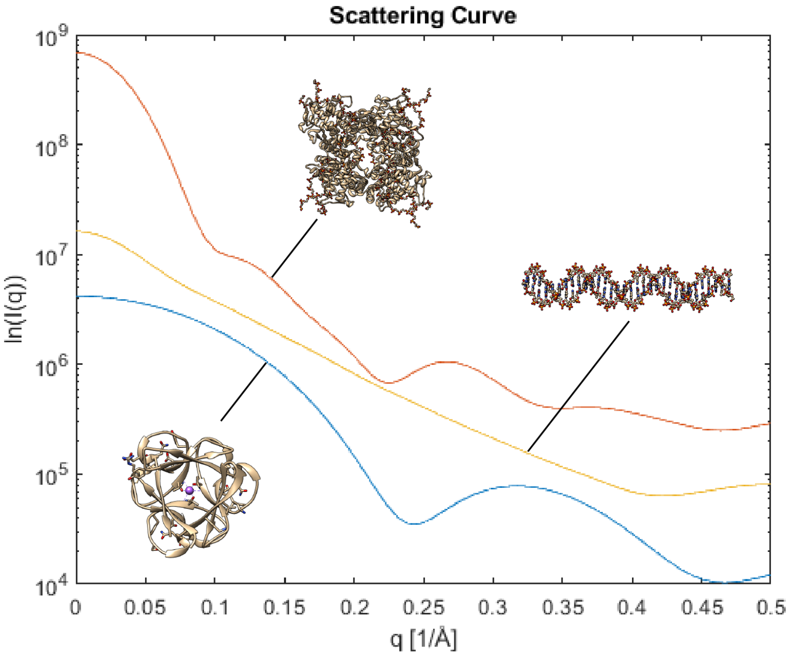

Figure 12: Calculated scattering curve for a generic DNA fragment generated in UCSF Chimera and PDB ID: 2HOI and 3PG0. Created with the FoXS web tool based on the Debye's equation for scattering intensity.

After the MacLaurin expansion (around $q = 0$) of the sinus term of the formula above, some simplification and rearrangement, substitution with R_g and $I(0)$, and the reverse Taylor expansion we arrive at Guinier's approximation for small angles.

$$lnI(\mathbf{q}) = lnI(0) - \frac{R_g^2}{3}\mathbf{q}^2$$

Where the zero angle scattering intensity $I(0) = 4\pi \int_0^{D_{max}} p(r)dr$ is used for calculating the size and mass of the molecule and the square of the radius of gyration $R_g^2 = \frac{1}{2}\frac{\int_0^{D_{max}} r^2 p(r)dr}{\int_0^{D_{max}} p(r)dr}$ is used to calculate the radius of gyration, which is basically the root mean square of the distances in the molecule weighted by their probability density curve, the pair distribution function $p(r)$ (Rambo 2019a).

From the Guinier plot where you plot $ln(I(\mathbf{q}))$ vs $\mathbf{q}^2$, one can extract R_g from the slope of the line and $I(0)$ from where the line intercepts the y-axis. It is also possible to see if the data follows the linear trend for small **q**-values. If it deviates from this, one or more of the assumptions inherently made in Guinier's approximation is false. This is especially true when there exist strong particle-particle interactions within the sample leading to the breakdown of the individual structure factors. If the Guinier plot shows a nonlinear decrease at the beginning of the curve before it becomes linear, there is a problem with aggregations in the sample and thus the scattering intensity positively deviates from the predicted intensity from the true size of the molecule. Conversely, if it shows a nonlinear increase, there are repulsive forces between the molecules and it negatively deviates from the predicted scattering intensity (Grant *et al.* 2015).

From the scattering pattern from the data collection, one can plot $I(\mathbf{q})\mathbf{q}^2$ vs **q** in a Kratky plot and thus divide out the decay of the scattering intensity as **q** increases. This allows other features of the sample to be more visible than in the scattering curve. From the Kratky plot one can evaluate the folded state of the sample. A well-folded protein should show a bell-shaped curve that tends to zero and an unfolded chain should show a plateau for high **q**-values (Putnam *et al.* 2007).

2.4.2 Shape Reconstruction

The shape reconstruction algorithm DAMMIF was used in this report and is based on the same principles as DAMMIN, which is an older and not as optimized algorithm. For example, in DAMMIF the initial search volume is populated by dummy atoms by spherical packing which increases the possible resolution compared to DAMMIN. It also does not commit to the more heavy calculations of the algorithm before checking all the requirements for a good solution like the compactness and connectivity of the new model.

DAMMIF first defines a random search volume, based on the experimentally determined R_g-value, filled by dummy atoms. At each iteration in the algorithm, a dummy atom is randomly selected and replaced by a bead representing the solution. If this change leads to a model that is disconnected, this change is rejected, and the next iteration is initiated. If this is not the case, the scattering pattern from the new model is calculated using spherical harmonics instead of the structure factor in Debye´s scattering formula, significantly lowering the time complexity of the algorithm. This pattern is then compared to the scattering pattern from the experiment. If the error value between the model and the experimental data decreases, the change is accepted, but if it increases, the new model can still be accepted with a certain probability as described by the probability distribution $e^{\frac{\Delta}{T}}$ where Δ is model error value and T temperature. If the temperature is high, or if the difference is low between the model and the experimental data, the likelihood is higher that the new model will be accepted. In the beginning, the temperature is high and decreases during the run. This is done to prevent the model from ending up in a local minimum and speeds up the model building in the beginning. At each temperature, 100 million iterations or 10 million iterations with successful changes are performed in the model before the temperature is lowered. This is repeated until no further improvements in error value between the model and the experimental data is observed (Franke & Svergun 2009).

Another algorithm used during the project was the ensemble optimization method (EOM). This method fits more than one model to the scattering pattern and can, therefore, distinguish between different conformational states present in the sample. If the protein assumes two different conformations, DAMMIF would give an averaged model of these while EOM would distinguish between these states. It also gives an indication of how much each state contributes to scattering the pattern, and thus gives an estimate of the content of the different conformations present in the sample (Bernadó *et al.* 2007).

This method, however, requires either a crystal structure or a homologous crystal structure of the protein and good knowledge of how the protein behaves, which parts are flexible and which parts make up rigid domains.

The method involves randomly generating 10,000 different conformations, and allowing the flexible links between the rigid domains to assume random but biological relevant values from a quasi-Ramachandran plot. These structures are then divided into subsets and screened using a genetic algorithm that selects the best subset explaining the experimental data. In each iteration, 20% of the structures in the subset are replaced with structures from other subsets of the same generation or from the generated pool. This is typically done over 5000 generations, which in the end generates the final model (Bernadó *et al.* 2007).

3 Methods

3.1 Cloning

This section describes the workflow of the cloning used to create the plasmid pET28-NStrep-TEV-PerR that was used in the rest of the project to express PerR in *E. coli*. The methods are presented in the order they were used in the experiment to facilitate readability and increase reproducibility. An overview of the final gene construct can be seen, in figure 13.

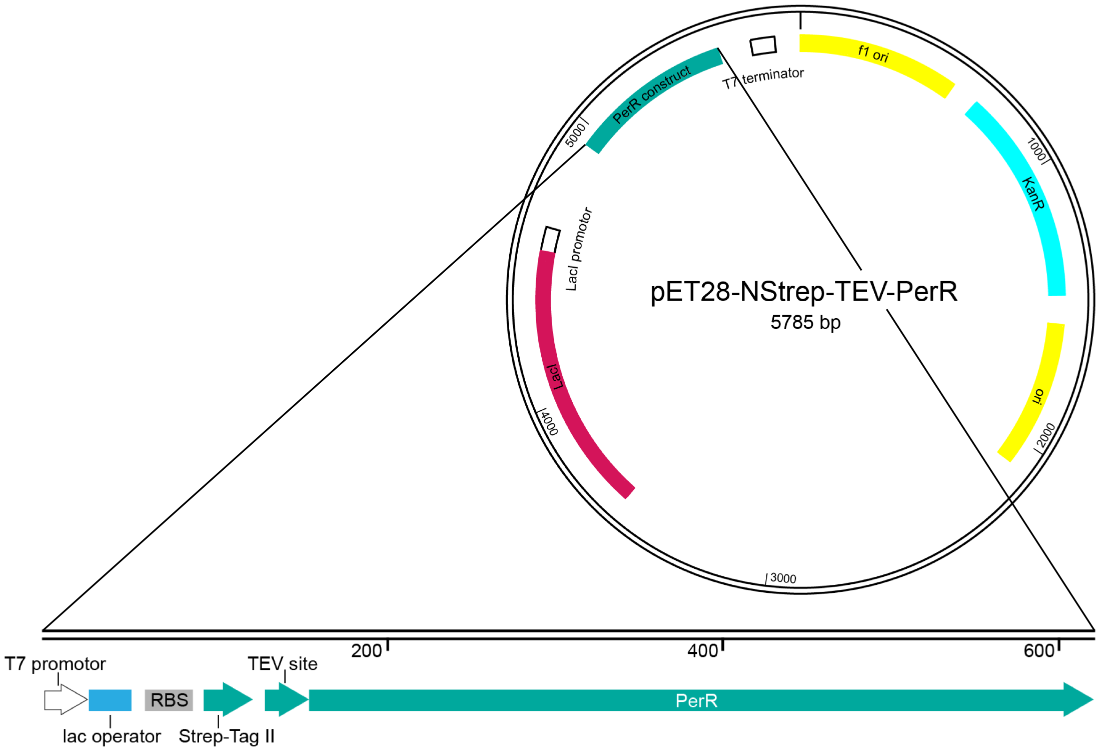
f1 ori
T7 terminator
PerR construct
5000
1000
KanR
LacI promotor
pET28-NStrep-TEV-PerR
5785 bp
LacI
ori
4000
2000
3000
200
400
600
T7 promotor
TEV site
RBS
PerR
lac operator
Strep-Tag II

Figure 13: The pET28-NStrep-TEV-PerR plasmid created during this project. At the top is the schematic map of the plasmid and underneath is the gene construct itself. The plasmid backbone carries a gene for kanamycin resistance, KanR, two origins of replications, f1 ori and ori, and the lac repressor, LacI. The gene construct is controlled by the lac operator making it inducible by lactose and IPTG, has a Strep-Tag II in the N-terminus for the capture of it on Tactin columns and a TEV site in between the tag and the gene itself allowing the removal of the tag.

3.1.1 PCR

The first step in the cloning was to extract the PerR gene from the genomic DNA by PCR using Phusion High-Fidelity DNA polymerase. This was done with two pre-ordered primers and genomic DNA from *S. erythraea* which were provided by Julia Griese's lab.

Materials:

- 182 ng/mL Template gDNA from *S. erythraea*
- 10 µM Forward primer (NdeI restriction site underlined):
  TAT CAT ATG CCA ACG ACG ACT GCG GAC TTC
- 10 µM Reverse primer (HindIII restriction site underlined):
  TAT AAG CTT TCA GTC ACC GGA ATT CCT TTC GGG CTC AG
- 2 U/µL Phusion High-Fidelity DNA polymerase
- 5X High-Fidelity Buffer
- 100% DMSO
- 10 µM dNTPs

Protocol:

1. The following reaction was prepared in a PCR tube, see table 1.

Table 1: PCR reaction composition.

Reagent	**Volume**
ddH$_2$O	30 µL
5X HF-buffer	10 µL
DMSO	2.5 µL
dNTPs	1 µL
Primers	2.5 µL
DNA template	1 µL
HF-polymerase	0.5 µL
Total	**50 µL**

2. The PCR was carried out in a thermocycler with a lid temperature of 105 ºC with the following program, see table 2.

Table 2: Program used for the thermocycler.

Step	Temperature/Action	Duration
1	98 ºC	0:30 s
2	98 ºC	0:20 s
3	68 ºC	0:15 s
4	72 ºC	0:45 s
5	GOTO Step 2, x 30	
6	72 ºC	5:00 min
7	4 ºC	hold

3. 5 µL of the finished PCR product was mixed with 1 µL 6X gel loading dye and run on a 1% agarose gel (8 cm) stained with SYBER safe at 120 V (15 V/cm) for 30 min in 1X TAE.

4. The PCR product was lastly purified using the QIAquick PCR purification kit following the accompanying instructions. The final concentration was measured using a NanoDrop 2000.

3.1.2 Restriction Digest

The next step after the amplification of *perR* was to trim the ends and to linearize the plasmid with compatible restriction enzymes so that the plasmid and *perR* could be joined together by ligase and re-circularized. FastDigest restriction enzymes were used, and the recommended protocol from the manufacturer was used.

Materials:

- 50 pmol PCR product
- 0.3 pmol pET28-NStrep-TEV plasmid
- 10X FastDigest Buffer
- 1 U/µL FastDigest HindIII Restriction enzyme
- 1 U/µL FastDigest NdeI Restriction enzyme

Protocol:

1. Two reactions were prepared — one for the plasmid backbone and one for the PCR product. The following components were mixed in two 1.5 mL Eppendorf tubes according to the table 3 below.

   Table 3: Digestion reaction composition.

Reagent	**Volume**
PCR product OR plasmid	85 µL
10X FD-buffer	10 µL
FD HindIII	2.5 µL
FD NdeI	2.5 µL
Total	**100 µL**

2. The reactions were left to incubate for 30 min in 37 °C.
3. When the incubation was done, the reaction was purified from enzymes and short DNA fragments using the QIAquick PCR purification kit following the accompanying instructions. The concentration for the purified and digested plasmid was measured using a NanoDrop 2000.

3.1.3 Dephosphorylation

To avoid intramolecular ligation of the plasmid, it was treated with phosphatase, dephosphorylating the 5′-end. Since the ligase needs at least one 5′-end phosphate to work this, in principle, ensures that the ligation only occurs in the presence of the insert, which is not dephosphorylated. FastAP Thermosensitive alkaline phosphatase was used in the reaction, and the protocol recommended by the manufacturer was followed.

Materials:

- Digested pET28-NStrep-TEV plasmid
- 1 U/µL FastAP Thermosensitive alkaline phosphatase
- 10X FastAP Buffer

Protocol:

1. The following reaction was prepared in a PCR tube, see table 4.

Table 4: Dephosphorylation reaction composition.

Reagent	**Volume**
10X AP-buffer	2 µL
Digested plasmid	0.1 µg
FastAP Phosphatase	1 µL
ddH_2O	to 20 µL
Total	**20 µL**

2. The reaction was mixed thoroughly, briefly, and then incubated for 10 min at 37 °C.
3. After the incubation was finished, the enzyme was heat inactivated at 75 °C for 5 min. The reaction mixture was then directly used in the ligation reaction.

3.1.4 Ligation

Following the dephosphorylation of the plasmid, the plasmid was ligated with the *perR* gene fragment and re-circularized using T4 DNA Ligase.

Materials:

- 12.5 fmol digested and dephosphorylated pET28-NStrep-TEV plasmid
- 125 fmol digested PCR product
- 1 U/µL T4 DNA Ligase
- 10X T4 Buffer

Protocol:

1. The reaction was prepared in a PCR tube according to table 5.

Table 5: Ligation reaction composition.

Reagent	**Volume**
ddH_20	6.5 µL
10X T4-buffer	2 µL
Digested/dephosphorylated plasmid	10 µL
Digested PCR product	0.5 µL
T4 DNA Ligase	1 µL
Total	**20 µL**

2. The reaction was left on the bench in room temperature for about 30 min.
3. Lastly, the enzyme was heat inactivated at 65 ºC for 10 min and then chilled on ice before the transformation.

3.1.5 Transformation

To screen for the plasmids that were successfully re-circularized with *perR*, the chemically competent Top10 *E. coli* cells were transformed on a heat bath with the ligase reaction from previous steps. The transformed cells were then allowed to grow on LB media containing antibiotics to select for the cells that had taken up functional plasmids.

Materials:

- Chemically competent Top10 *E. coli*
- Ligation reaction mix
- SOC media: 2% (v/w) tryptone, 0.5% (v/w) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose
- Agar plates supplemented with 50 µg/mL kanamycin

Protocol:

1. The competent cells were picked from the -80 ºC freezer and thawed on ice for about 20 min.
2. 10 µL of the ligation reaction was added to the Top10 cells. The samples were carefully mixed by gently tapping the vial on the bench, and then the cells were incubated on ice for 30 mins.
3. After the incubation was done, the cells were heat shocked at 42 ºC for 45 s and then incubated for 5 min on ice.
4. 450 µL of SOC media was added to each sample and then incubated on a shaker with 140 rpm at 37 ºC for 45 - 60 min.
5. Finally, 100 and 200 µL of the samples were spread on two agar plates and left overnight at 37 ºC.

3.1.6 Overnight Culture

To study which colonies carried the correct plasmid, more plasmid DNA was needed. This was done by expanding the colonies into small LB cultures for later plasmid extraction and purification.

Materials:

- Agarplates with colonies
- LB media: 1% (v/w) peptone, 0.5% (v/w) yeast extract, 0.5% (v/w) NaCl
- 1000X Kanamycin stock solution, 50 mg/mL

Protocol:

1. Nine cultures in total were prepared in 50 mL falcon tubes. 5 mL of LB media was pipetted into each falcon tube, and 5 µL of 1000x kanamycin stock solution was added to the working concentration 50 µg/mL.
2. Nine different colonies were then selected. With an inoculation loop, the colonies were scraped up and used to inoculate the cultures.
3. The cultures were left overnight at 37 ºC on a shaker set to 140 rpm.

3.1.7 Colony Plasmid Prep

The cultures from the previous step were harvested and the plasmids were isolated. This was done by first lysing the cells in an alkaline solution with RNase and SDS, which breaks down the RNA, and precipitates the proteins, and the DNA. The solution was then neutralized with potassium acetate, allowing only the small circular plasmid DNA to become soluble again, which was then extracted. Lastly, the isolated plasmids could be screened on an agarose gel for a size shift, which would indicate if it had the gene insert or not.

Materials:

- P1: 50 mM Tris, pH 8.0, 10 mM EDTA, and RNase 100 µg/mL
- P2: 200 mM NaOH, and 1% SDS
- P3: 3M Potassium acetate, pH 5.6

Protocol:

1. The overnight cultures from the previous step were harvested the day after by centrifugation at 4000 x rcf in 4 ºC.
2. The cell pellet was resuspended in 250 µL P1 solution and transferred to 1.5 mL Eppendorf tubes. 250 µL of P2 was added and the samples were inverted several times. Then 250 µL of P3 was added and again the tubes were inverted several times.
3. The tubes were centrifuged at 16100 x rcf for 5 min and the supernatant was carefully decanted into new 1.5 mL Eppendorf tubes containing 400 µL ice-cold isopropanol.
4. The tubes were centrifuged again at 16100 x rcf for 6 min in the cold room and the supernatant was discarded. The invisible DNA pellet that was left was washed with 400 µL ice-cold 75% EtOH.
5. Lastly the tubes were centrifuged for 10 min at the same speed as before. The supernatant was discarded and the tubes were placed upside down on a napkin in room temperature to let the pellets dry. After about 30 mins of drying 30 µL ddH$_2$O was added and the pellets were resuspended by vortexing. After the resuspension, the samples were briefly spun down.

6. 5 μL of the finished plasmid preps was mixed with 1 μL 6X gel loading dye and run on a 1% agarose gel (8 cm) stained with SYBER safe at 120 V (15 V/cm) in 1X TAE for 45 min.

3.1.8 Plasmid Fingerprinting

To be sure that the plasmid had picked up the gene, a plasmid fingerprinting experiment was performed. This involved cutting with a restriction enzyme in the middle of the gene that is not otherwise present on the plasmid and with another restriction enzyme that cut the plasmid independently whether the gene was inserted or not. If the plasmid was linearized, the gene had not been inserted, and if two bands of different size were present, it had been inserted. From the size of the band, one could tell if it had been taken up in the right direction. This was also confirmed by sequencing.

Materials:

- Positive colony plasmid
- 10X FastDigest Buffer
- 1 U/µL FastDigest HindIII Restriction enzyme
- 1 U/µL FastDigest NdeI Restriction enzyme
- 1 U/µL FastDigest BamHI Restriction enzyme
- 1 U/µL FastDigest Bg1II Restriction enzyme

Protocol:

1. The plasmid preps that were positive from the previous gel were chosen for further investigation.
2. Two different digestion reactions were prepared for each colony plasmid prep according to the table below. These reactions were designed so that the positive colonies would show a 600 bp long fragment for digestion reaction 1 and 500 bp for digestion reaction 2, see table 6.

Table 6: Restriction enzymes used in the different digestion reactions.

Digestion	Restriction enzymes
1	BamHI, Bg1II
2	HindIII, NdeI

3. The following reaction was prepared in each of the PCR tubes in two 4xPCR-strips, see table 7.

Table 7: Reaction composition for the digestion.

Reagent	**Volume**
Plasmid prep	8.5 µL
10X FD-buffer	1 µL
Digestion 1 OR 2 enzymes	2 x 0.25 µL
Total	**10 µL**

4. The reactions were left to incubate for 30 min in 37 ºC.
5. When the incubation was done 5 µL of each reaction was mixed with 2 µL 6X gel loading dye and run on a 1% agarose gel (8 cm) stained with SYBER safe at 120 V (15 V/cm) in 1X TAE for 30 min.
6. Since the experiments indicated that the gene had been correctly ligated, the plasmid was sent for sequencing with primers for the T7 promoter and T7 terminator flanking the insert, which confirmed it.

3.2 Small Scale Expression Tests

Below is the protocol for the small-scale expression test for determining the optimal conditions for the expression of PerR. Various parameters were tested such as strain, temperature, type of induction and time of harvest.

3.2.1 Transformation

Using the plasmid previously obtained from the cloning of *perR*, electrocompetent BL21-AI *E. coli* cells were transformed by electroporation for the heterologous expression of PerR from *S. erythraea*.

Materials:

- Electrocompetent BL21-AI and Rosetta *E. coli*
- pET28-6xHis-TEV-PerR plasmid
- SOC media: 2% (v/w) tryptone, 0.5% (v/w) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose
- Agar plates supplemented with 50 µg/mL kanamycin

Protocol:

1. The competent cells were picked from the -80 ºC freezer and thawed on ice for about 30 min.
2. 2 µL of plasmid was added to the cells. The samples were carefully mixed by gently tapping them on the table.
3. The cells were then incubated on ice for 30 mins.
4. The samples were electroporated in 1mm cuvettes at 1800 V with 25 µF and 200 Ω using an exponential decay wave pulse.
5. 450 µL of SOC media was added to the sample, and then the samples were incubated on a shaker with 140 rpm at 37 ºC for 45 min.
6. Lastly, 100 µL of the samples were spread on agar plates containing antibiotics for the selection of plasmids and left overnight at 37 ºC.

3.2.2 Starter Culture

Materials:

- Agarplates with colonies
- LB media: 1% (v/w) peptone, 0.5% (v/w) yeast extract, 0.5% (v/w) NaCl
- 1000X Kanamycin stock solution, 50 mg/mL

Protocol:

1. The next day two starter culture were prepared in 5 mL of LB supplemented with antibiotics in two 50 mL Falcon tubes.
2. Two colonies of medium size were picked from the plate from the previous day and used to inoculate the starter cultures.
3. The cultures were incubated on a shaker at 140 rpm in 37 ºC for 4 h.

3.2.3 Expression & Cell Harvest

Materials:

- Starter culture
- LB media: 1% (v/w) peptone, 0.5% (v/w) yeast extract, 0.5% (v/w) NaCl
- ZYM-5052 media: 1% (v/w) peptone, 0.5% (v/w) yeast extract, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 0.5% glycerol, 0.05% glucose, 0.2% α-lactose, 2 mM $MgSO_4$, 10 µM $FeCl_3$, 4 µM $CaCl_2$, 2 µM $MnCl_2$, 2 µM $ZnSO_4$, 0.4 µM $CoCl_2$, 0.4 µM $CuCl_2$, 0.4 µM $NiCl_2$, 0.4 µM Na_2MoO_4, 0.4 µM Na_2SeO_3, 0.4 µM H_3BO_3
- 1000X Kanamycin stock solution, 50 mg/mL
- 1 M IPTG
- 20% L-arabinose

Protocol:

1. Ten cultures were prepared in total. The cultures were prepared in 50 mL falcon tubes with 5 mL of media supplemented with antibiotics, induced and incubated as follows, see table 8 below.

Table 8: Experimental setup for the expression test.

Nr.	Strain	Media	Temperature	OD_{600} of induction	Induced with
1	Rosetta	ZYM-5052	20 ºC	Auto-induced	Auto-induced
2	BL21-AI	ZYM-5052	20 ºC	Auto-induced	Auto-induced
3	BL21-AI	LB	20 ºC	0.4	0.1 mM IPTG 0.1% L-arabinose
4	Rosetta	ZYM-5052	22 ºC	Auto-induced	Auto-induced
5	BL21-AI	ZYM-5052	22 ºC	Auto-induced	Auto-induced
6	BL21-AI	LB	22 ºC	0.4	0.1 mM IPTG 0.1% L-arabinose
7	Rosetta	ZYM-5052	37 ºC	Auto-induced	Auto-induced
8	BL21-AI	ZYM-5052	37 ºC	Auto-induced	Auto-induced
9	BL21-AI	LB	37 ºC	0.4	0.1 mM IPTG 0.1% L-arabinose
10	BL21-AI	LB	37 ºC	0.4	0.1 mM IPTG 0.1% L-arabinose

2. The cultures were set on a shaker at 140 rpm and incubated at their corresponding temperatures overnight; Culture Nr. 10 was harvested 3 h after induction.

3. The following day, after roughly 21 h of incubation for the auto-induced cultures and 16 h for the manually induced cultures, the cells were harvested by centrifugation. The cell cultures were centrifuged at 4000 x rcf for 15 min at 4 ºC. The supernatant was decanted, and the cell pellets were kept on ice. The remaining pellet was dissolved in 250 µL of ddH_2O by pipetting up and down and analyzed on SDS-PAGE.

3.3 Expression & Purification

Here is the workflow used to express PerR on a large scale in *E. coli*. This protocol was developed with the support of previous attempts by *Julia Griese* and much experimentation. Furthermore, each step of the protocol was fine-tune after many useful tips from colleagues at the structural biology lab. This method is not yet complete and may need to be optimized in the future with regards to buffers and chromatographic steps to increase the yield of the dimeric form of PerR.

3.3.1 Transformation

Using the plasmid previously obtained from the cloning of *perR*, electrocompetent BL21-AI *E. coli* cells were transformed by electroporation for the heterologous expression of PerR from *S. erythraea*.

Materials:

- Electrocompetent BL21-AI *E. coli*
- pET28-NStrep-TEV-PerR plasmid
- SOC media: 2% (v/w) tryptone, 0.5% (v/w) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose
- Agar plates supplemented with 50 µg/mL kanamycin

Protocol:

1. The competent cells were picked from the -80 ºC freezer and thawed on ice for about 10 min.
2. 2 µL of plasmid was added to the cells. The samples were carefully mixed by gently tapping them on the table.
3. The cells were then incubated on ice for 30 mins.
4. The samples were electroporated in 1mm cuvettes at 1800 V with 25 µF and 200 Ω using an exponential decay wave pulse.
5. 450 µL of SOC media was added to the sample, and then the samples were incubated on a shaker with 140 rpm at 37 ºC for 45 min.

6. Lastly, 100 μL of the samples were spread on agar plates containing antibiotics for the selection of plasmids and left overnight at 37 ºC.

3.3.2 Overnight Culture

To minimize the risk of contamination and speed up the colonization rate, a colony was first expanded into a starter culture in LB, which was then used in the inoculation of the larger cultures.

Materials:

- Agarplates with colonies
- LB media: 1% (v/w) peptone, 0.5% (v/w) yeast extract, 0.5% (v/w) NaCl
- 1000X Kanamycin stock solution, 50 mg/mL

Protocol:

1. The next day one overnight culture was prepared in 100 mL of LB supplemented with antibiotics in a 500 mL Erlenmeyer flask.
2. One colony of medium size was picked from the plate from the previous day and used to inoculate the overnight culture.
3. The culture was incubated on a shaker at 120 rpm in 37 ºC overnight.

3.3.3 Expression & Cell Harvest

PerR was determined to be slightly toxic to the *E. coli* cells, probably due to its DNA binding properties. Since PerR was from another organism, this should lead to off target effects in the *E.coli* cell genome, and significantly impair the growth rate. Therefore, the BL21-AI strain was used, which allowed a tighter control over the expression of PerR than the usual BL21 strain, which is otherwise the gold standard for the expression of proteins. In BL21-AI, the T7 RNA polymerase is under the control of the arabinose-inducible araBAD promoter. Because pET plasmids use the T7 system, both arabinose and IPTG had to be used for the induction. Because of the toxicity of PerR, the cells were grown in LB media, induced at a late stage, and harvested only 3 hours later.

Materials:

- 8 x 0.8 L LB media: 1% (v/w) peptone, 0.5% (v/w) yeast extract, 0.5% (v/w) NaCl
- 100 mL Overnight culture
- 1000X Kanamycin stock solution, 50 mg/mL
- 8 x 8.8 mL Induction solution: 0.8 mL 1 M IPTG, 8 mL 20% L-arabinose

Protocol:

1. Eight cultures were prepared in 2.8 L baffled Fernbach-flasks with prewarm 800 mL of LB media supplemented with 800 µL of 1000X kanamycin stock solution in each.
2. 16 mL of the overnight culture was added to each of the cultures and then incubated on a shaker at 120 rpm in 37 ºC for about 2 h, until the OD_{600} was measured to 1.2.
3. The cultures were induced with the premixed and prewarmed IPTG and L-arabinose solution to a final concentration of 1 mM and 0.2% respectively and incubated on a shaker at the same speed as before in 37 ºC for 3 h before they were harvested.
4. The cell cultures were transferred from the Fernbach-flasks to 1 L centrifuge tubes, balanced and then centrifuged at 4500 x rcf for 12 min at 20 ºC.
5. The supernatant was decanted, and the cell pellet was transferred to two 50 mL Falcon tube on ice using a plastic spatula.

6. The wet cell weight was measured, and the pellets were stored in -80 ºC.

3.3.4 Cell Lysis

As the Tactin-affinity chromatography step required specific conditions to function optimally, the protocol provided by the manufacturers was followed, and their recommended buffer was used as the lysis buffer. Furthermore, the binding capacity of these beads were not as high as that of the Ni-Sepharose beads, which made it easy to overload them. Therefore, sonication was used which does not require large volumes of lysis buffer for the resuspension of the cells as compared to a high pressure cell disrupter, which is otherwise considered to be a better alternative for lysing cells.

Materials:

- Buffer W: 100 mM Tris, pH 8.0, 150 mM NaCl
- cOmplete, EDTA-free Protease Inhibitor Cocktail
- Frozen *E. coli* cells from previous step

Protocol:

1. One tablet of cOmplete, EDTA-free Protease Inhibitor Cocktail for 50 mL was added and dissolved in 48 mL Buffer W in a 100 mL beaker with a magnetic stirrer in the 8 ºC cold room.
2. One Falcon tube containing roughly 14 g of cells (wet weight) from the previous step was resuspended with the 48 mL pre-chilled Buffer W. The pellet was dissolved by vortexing and occasionally manually shaking the tubes vigorously to make sure the solution was thoroughly homogenized.
3. Once the cell pellet was dissolved, the resuspension was poured into two Falcon tubes. The Falcon tubes were fitted in the mouth of a 500 mL media-bottle filled with ice and some water to increase the heat conduction. The cells were then lysed with a 130 W sonicator with 10 s ON 30 s OFF pulses for a total ON time of 12 min with 20 kHz, at 50% amplitude.
4. The lysed cells were then centrifuged at 16000 x rcf for 15 min at 4 ºC. The reason for this short and gentle centrifugation was due to the observation that PerR sedimented in weaker centrifugal fields when concentrating it. This means that much is lost in the decanting step after the centrifugation when PerR gets stuck in the surface tension of the liquid layer that remains.

5. The supernatant was carefully aspirated into another Falcon tube without disturbing the loose top-layer of lipids using a pipette. The supernatant was then filtered through 0.45 μM filters using a 50 mL lauer syringe, changing the filters when needed as the filters became clogged.

3.3.5 Tactin-affinity Chromatography

The Strep-tag:Strep-Tactin system is based on the binding of biotin to streptavidin, which is one of the strongest non-covalent bonds known. Once biotin has bound, it remains in place despite harsh treatment with detergents, extreme pH or temperatures. This property was exploited by developing the Strep-tag, a short, 8-amino acid long peptide that can bind to the same pocket that biotin binds to. After the sequence was found, streptavidin was optimized with regards to binding properties leading to the creation of Strep-Tactin. By immobilizing Strep-Tactin on gel beads, it is possible to capture proteins that carry a Strep-tag in a chromatographic step when purifying proteins. To elute the proteins, desthiobiotin is added, which displaces the bound protein competitively and reversibly binds to the Strep-Tactin (IBA Lifesciences 2019). Since the expressed PerR carried a Strep-tag it could be captured to a high degree of purity in a single Tactin-affinity chromatography step.

Materials:

- Buffer W: 100 mM Tris, pH 8.0, 150 mM NaCl
- Buffer E: 100 mM Tris, pH 8.0, 150 mM NaCl, 2 mM desthiobiotin
- 5 mL Strep-Tactin beads

Protocol:

1. All the following steps were performed in the 8 ºC cold room. The Strep-Tactin column was equilibrated with at least two CVs (column bed volumes) of Buffer W. The column bed volume was 5 mL.
2. The supernatant of the cleared lysate was added to the column.
3. After the cell extract had completely entered the column, the column was washed 5 times with one CV Buffer W. The flow-through was collected in one CV fraction.
4. 6 times of 0.5 CVs Buffer E was added, and the eluate was collected in 0.5 CV fractions.
5. The fractions were analyzed on an SDS-PAGE and the concentrations measured using a NanoDrop at A_{280}. The fractions which had the highest concentration of PerR were pooled together.

3.3.6 TEV Cleavage & Removal of Uncleaved Product

Because the Strep-tag may affect PerR's natural function, TEV protease was added, and the tag was cleaved off. TEV is a cysteine protease from the Tobacco etch virus that cleaves at the specific peptide sequence ENLYFQ | G inserted between the Strep-tag and PerR. The TEV protease, therefore, leaves a glycine on the original protein and cleaves off the rest of the Strep-tag. After the removal of the Strep-tag, uncleaved PerR together with cleaved Strep-tag was removed in small spin columns containing Strep-Tactin.

Materials:

- Buffer W: 100 mM Tris, pH 8.0, 150 mM NaCl
- TEV protease
- 14.3 M β-mercaptoethanol
- 0.5 M EDTA, pH 8.0

Protocol:

1. The pooled fractions were first diluted to 1-2 mg/mL with Buffer W
2. β-mercaptoethanol was then added to a final concentration of 10 mM and EDTA to a final concentration of 0.5 mM. In the end 1:50 (w/w) TEV to target protein was added.
3. The sample was incubated for 1 h at 30 ºC and then put on ice.
4. The sample was first concentrated down to a total volume of 3 mL by centrifugation at 4000 x rcf in 4 ºC for 15 min using a Vivaspin Turbo 15 column equilibrated with Buffer W.
5. All the following steps were performed in the 8 ºC cold room. Six Strep-Tactin Spin Column were equilibrated with 500 µL Buffer W. The columns were centrifuged for 30 seconds at 700 x rcf (approx. 2000 rpm), and the flowthrough was discarded. This was repeated one more time.
6. 500 µL of the concentrated sample was loaded onto the pre-equilibrated Strep-Tactin Spin Columns, and the columns were centrifuged for 30 seconds at 700 x rcf.

7. The columns were then washed two times with 100 µL Buffer W. For each washing step the column was centrifuged for 30 seconds at 16000 x rcf (approx. 13000 rpm).
8. The flowthrough was poured into a 10 mL falcon tube and kept in the 4 ºC fridge overnight.

3.3.7 Size Exclusion Chromatography

Size exclusion chromatography is based on the principle that smaller molecules can penetrate porous gel beads to a greater extent than large molecules. This means that the average distance for smaller molecules is longer than for molecules of larger size when migrating through the gel. This is usually done in a column packed with gel beads which the sample is allowed to pass through leading to a separation of the molecules depending on size (GE Healthcare 2010). This was done to separate PerR from TEV, but unfortunately, the dimeric form of PerR has the same size as TEV, therefore, only PerR monomers could be resolved in this purification step. This problem can, however, be circumvented using an immobilized metal ion affinity chromatography step since TEV carries a 6xHis-tag or by putting a Strep-tag on TEV.

Materials:

- Buffer W: 100 mM Tris, pH 8.0, 150 mM NaCl

Protocol:

1. The size exclusion chromatography was carried out with a ÄKTApurifier (GE healthcare) on a HiLoad 16/60 Superdex 75 column. First the pump was flushed with 10 mL of Buffer W, and then the column was equilibrated with 1.2 CV of Buffer W with a flow rate of 1.5 mL/s.
2. When the column was equilibrated, the injection loop was flushed with some buffer W, and the sample was loaded and then run on the gel column with a flow rate of 1.2 mL/s. Beginning from 40 mL, 2 mL fractions were collected until 120 mL.

3.3.8 Buffer Exchange, Concentration & Storage

To store PerR, it was stabilized in a optimized buffer solution, see subsections 3.4.1, 3.4.2 & 3.4.3. It was concentrated to a sufficiently high concentration for later crystallization experiments and then rapidly cooled down by plunging the protein sample into liquid nitrogen to prevent ice crystal formation which can damage the protein by shearing.

Materials:

- Buffer W: 100 mM Tris, pH 8.0, 150 mM NaCl
- Buffer S: 100 mM MES-NaOH, pH 6.5, 300 mM NaCl, 15% (v/v) glycerol

Protocol:

1. Fractions free of TEV were pooled together and concentrated in a Vivaspin Turbo 15 column equilibrated with Buffer W by centrifugation at 4000 x rcf in 4 ºC for about 15 min.
2. Once the centrifugation was complete, the flowthrough was discarded. The cup was partially filled with Buffer S, and the sample in the concentrator cup was pipetted up and down to ensure that the protein did not adsorb to the column, the container was then filled to the 15 mL mark with Buffer S.
3. This was repeated three times.
4. In the last run, the protein sample was centrifuged for 45 min at the same speed and temperature as before and concentrated to 13.6 mg/mL in a total volume of 300 µL.
5. The sample was pipetted into PCR tubes in 50 µL aliquots and flash-frozen in liquid nitrogen.
6. The protein sample was stored in -80 ºC.

3.4 Characterization

Here are the PerR characterization experiments that were done during the project. The first part, section 3.4.1, 3.4.2 & 3.4.3, aimed to find a buffer that stabilized the protein to increase the likelihood of crystal formation in the crystallization attempts and to find a suitable buffer for storage. The second part, section 3.4.4, describes the preparations, the collection, and the data analysis of the SAXS data of PerR in solution.

3.4.1 Solubility Screen

To find the best buffer for the stabilization of PerR for storage and crystallization, a thermal shift assay was carried out using the commercial JBScreen Solubility HTS kit in a CFXConnect Real-Time PCR Detection System (BioRad). The idea is that the more stable the protein is in a solution, the higher the melting point. When the protein unfolds at the melting temperature, the hydrophobic portions of the protein to which the dye SYPRO Orange binds are exposed. Because water quenches the fluorescence from SYPRO Orange, the binding causes a signal due to the exclusion of water, and thus the stability of the protein can be monitored over a temperature range.

Materials:

- 5000X SYPRO Orange
- JBScreen Solubility HTS kit
- PerR in 100 mM Tris-HCl, pH 8.0, 100 mM NaCl

Protocol:

1. The reactions were set up according to the instructions for the kit on a polycarbonate 96 well plate using a protein concentration of 140 µM in 100 mM Tris-HCl, pH 8.0, 100 mM NaCl.
2. The thermal shift assay was done on a CFX Connect Real-Time PCR Detection System (BioRad) with a temperature range between 15 - 95 ºC with increments of 1 ºC.

3.4.2 Solubility Screen with Additives

In an effort to increase the stability of the protein in the buffer found in the previous solubility screen another one was performed with additives using the commercial RUBIC Additive Screen.

Materials:

- 5000X SYPRO Orange
- RUBIC Additive Screen kit
- PerR in 100 mM MES-NaOH, pH 6.5, 150 mM NaCl, 5% (v/v) glycerol.

Protocol:

1. The reactions were set up according to the instructions for the kit on a polycarbonate 96 well plate using a total protein concentration of 140 µM PerR in 100 mM MES-NaOH, pH 6.5, 150 mM NaCl, 5% (v/v) glycerol.
2. The thermal shift assay was done on a CFX Connect Real-Time PCR Detection System (BioRad) with a temperature range between 15 - 95 ºC with increments of 1 ºC.

3.4.3 Buffer Optimization

Since the additive screen showed a preference for glycerol, a protein stability analysis was carried out to gauge the effect of different salt and glycerol concentrations. The experiment was set up as a full factorial experiment since the response was not expected to be a simple linear relationship with respect to the variables. The sample was run on a Tycho NT.6. The principle is similar to the thermal shift assay, but instead of measuring the binding of a dye, the intrinsic fluorescence of tryptophan and tyrosine is measured at 330 and 350 nm as the sample is heated up. This gives a rough idea about the folding state of the protein, as unfolded protein exposes more tryptophans than folded proteins.

Materials:

- PerR in 100 mM MES-NaOH, pH 6.5, 150 mM NaCl, 5% glycerol

Protocol:

1. 16 different buffer conditions with 100 mM MES-NaOH, pH 6.5 but with varying salt and glycerol concentrations were prepared with a protein concentration of 100 µM as follows, see figure 14.

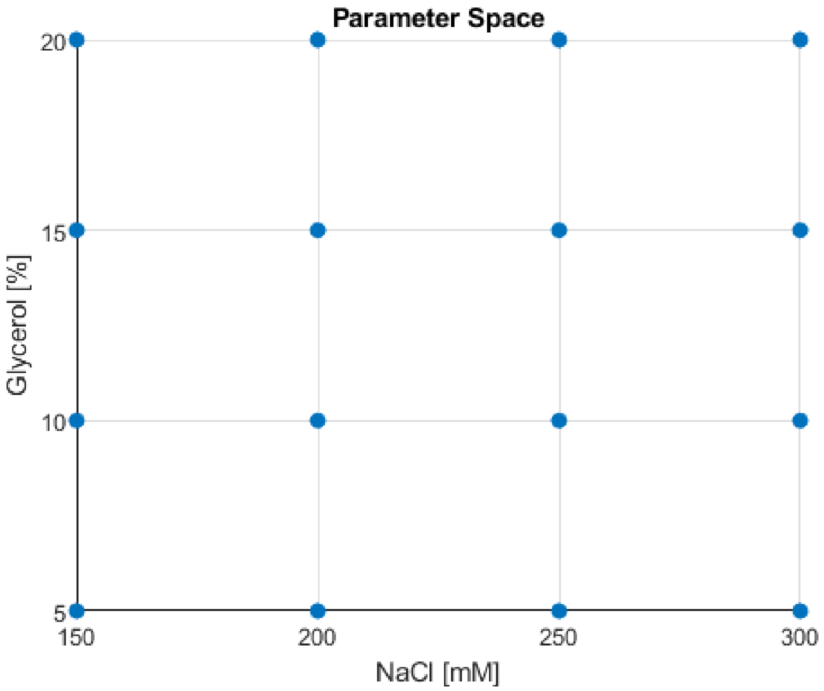

Figure 14: Sampled 100 mM MES-NaOH, pH 6.5 buffer conditions with varying salt and glycerol concentrations indicated as blue dots in the parameter space plot.

2. The samples were run on Tycho NT.6 (NanoTemper) using small capillaries to load the samples. The sample was heated from 35 to 95 ºC with a speed of 30 ºC/min.

3.4.4 Small-angle X-ray Scattering

To study how PerR behaved in solution and in what form PerR was in depending on factors such as ligands, metal ions and the presence of metal chelator, samples were sent for SAXS measurements at the B21 beam-line at Diamond Light Source Ltd. in Oxford. The samples were flash-frozen and stored on dry ice until its arrival at the beam-line where they were stored in 4 ºC until data collection. The SEC and the SAXS data collection was performed in room temperature. The SEC was carried out on a Shodex KW-403 column with 25 mM MES pH 6, 300 mM NaCl, 5% glycerol, and 1% sucrose. Beamsize: <75 µm, Detector: Eiger 4M, Energy: 13.1 keV.

The final concentration of PerR used in the SAXS experiment was 6 mg/mL. The four different samples were setup according to table 9.

Table 9: The ratio of PerR to different additives in the samples.

Sample	**1**	**2**	**3**	**4**
PerR:Zn	1:2	1:2	1:2	
PerR:Mn		1:3	1:3	
PerR:DNA			25:1	
PerR:EDTA				1:4

The DNA used in this experiment was found by *Francisco Javier Marcos Torres* from Julia Griese's lab by bioinformatical analysis of the genome of *S. erythraea* and upstream regions of genes regulated by PerR.

Materials:

- PerR in 100 mM MES-NaOH, pH 6.5, 300 mM NaCl, 15% glycerol (9 mg/mL)
- Buffer S: 100 mM MES-NaOH, pH 6.5, 300 mM NaCl, 15% glycerol
- 10 mM $ZnCl_2$
- 10 mM $MnCl_2$
- 0.5 M EDTA, pH 8.0
- 375 µM DNA (The putative PerR binding site is underlined):
  CTC TTG ATT CAT TCC AGA AAA TCA GCC AGA

Protocol:

1. Four different samples with a final protein concentration of 6 mg/mL was prepared in PCR tubes according to table 10 and 11.

Table 10: The composition of sample 1 & 2.

Sample 1		**Sample 2**	
Component	**Volume**	**Component**	**Volume**
PerR	36.36 µL	PerR	36.36 µL
10 mM Zn	3.9 µL	10 mM Zn	3.9 µL
10 mM Mn		10 mM Mn	5.86 µL
375 µM DNA		375 µM DNA	
0.5 M EDTA, pH 8.0		0.5 M EDTA, pH 8.0	
Buffer S	14.74 µL	Buffer S	8.88 µL
Total	**55 µL**	**Total**	**55 µL**

Table 11: The composition of sample 3 & 4.

Sample 3		**Sample 4**	
Component	**Volume**	**Component**	**Volume**
PerR	36.36 µL	PerR	36.36 µL
10 mM Zn	3.9 µL	10 mM Zn	
10 mM Mn	5.86 µL	10 mM Mn	
375 µM DNA	2.08 µL	375 µM DNA	
0.5 M EDTA, pH 8.0		0.5 M EDTA, pH 8.0	0.156 µL
Buffer S	6.8 µL	Buffer S	18.484 µL
Total	**55 µL**	**Total**	**55 µL**

2. The samples were flash frozen in liquid nitrogen, packed with dry ice and sent for data collection at beamline B21 at the Diamond Light Source Ltd. in Oxford. The samples were stored in 4 ºC before the data collection. The SEC and the data collection was carried out at room temperature.

Data Analysis:

1. ScÅtter (Rambo 2019b) was used for the initial processing of the data. The data for the samples was uploaded and traced in the graphical user interface, and a suitable region for the buffer subtraction was selected. A range slightly bigger than the frames of interest was chosen based on the elution profile and estimates of the radius of gyration, R_g.

   This region was then trimmed based on the similarity of the images in the similarity plot, usually resulting in the selection of the second half of the peaks of interest.

   Once the region to be analyzed was chosen, the buffer was subtracted, and the radially averaged curves were analyzed.

   First, the size-scale for the sample was extracted from a Guinier plot by linear regression at small **q**-values, and the quality of the fit was checked in the Guinier peak analysis.

   The Porod exponent was determined in the flexibility analysis, and the q_{max} was noted. Then the volume was determined using the volume analysis tool by fitting a line to the linear range in the Porod-Debye plot.

   The pair-distance distribution function, $p(r)$, was determined by typically trimming the first 15-30 data points and the endpoints down to about 900-1200. Then the $p(r)$ function was fitted to the data by trying to match the R_g with the predetermined value from the Guinier analysis at the same time as trying to achieve a low χ^2-value and a smooth $p(r)$ function with a reasonable end-slope that did not abruptly end with a sharp decline nor elongated for a considerable distance with unrealistically high D_{max}.

   Once this was achieved, the curve was refined, and the file was saved for further data processing.

   Before continuing, the data quality was checked once more with the volume of correlation analysis tool, indicating if any problems with the buffer subtraction, the presence of aggregation or interparticle interference was present.

2. With the generated data from ScÅtter the DAMMIF tool (Franke & Svergun 2009) from the ATSAS online provided by EMBL Hamburg was used to generate *ab initio* bead shape models fitted to the SAXS data with the following parameters, see table 12 below. For the dimeric peaks both P1 and P2 were used in two different runs to make sure that the imposed P2 symmetry did not lead to unbiased results.

Table 12: Parameters used for the DAMMIF modelling.

Dimeric PerR		**Monomeric PerR**	
Option	**Input**	**Option**	**Input**
Angular units:	1/Å	Angular units:	1/Å
Symmetry:	P1 AND P2	Symmetry:	P1
Anisometry:	Unknown	Anisometry:	Unknown
Run DAMMIF:	20 times	Run DAMMIF:	20 times
Postprocessing:	Yes	Postprocessing:	Yes

3. Using the EOM tool (Bernadó *et al.* 2007) from the ATSAS online provided by EMBL Hamburg, the data were fitted to an ensemble model. This was done using the crystal structure of the apo form of PerR from *B. subtilis*, PDB ID: 2FE3. First, the principal inertia was aligned with the coordinate axes in ALPRAXIN (Kozin & Svergun 2001) in the ATSAS package, and the new PDB was saved. One monomer was deleted, and the one remaining was divided into two domains in UCSF Chimera. These domains were chosen based on previous knowledge of PerR's different forms. In both structures, the flexible link between the domains was omitted. Since the amino acid sequence of the PDB started at 4, this was fixed so that it started from 1 and the other domain's amino acid sequence was shifted with 3. The domains were saved in two different PDB files. A .seq file containing the amino acid sequence in UPPERCASE was created in a text editor that corresponded to the sequences in the PDB files including the missing flexible link. To model the dimer, the core domain was needed in dimeric form. This was done in MASSHA (Konarev *et al.* 2001) in the ATSAS package by applying p2 symmetry to the core domain and then saving the PDB. When this was done, the files were uploaded to the web server and run with the following parameters, see table 13.

Table 13: Parameters used for the EOM modelling.

Option	**Input**	**Alternative**
Angular units:	1/Å	
Symmetry:	P2	
Sequence:	file.seq	
Generate:	random coil models	
Number of domains:	2	
Domain 1:	file.pdb	monomer, free
Domain 2:	file.pdb	multimer, fixed

3.5 Crystallization

Four different sitting drop vapor diffusion crystallization experiments were set up. Two commercial screens were used, and PerR from two different constructs were tried in different buffers and concentrations.

Materials:

- PerR in 100 mM MES-NaOH, pH 6.5, 150 NaCl, 5% glycerol (8.5 mg/mL)
- PerR in 100 mM MES-NaOH, pH 6.5, 150 NaCl, 15% glycerol (13.6 mg/mL)
- PerR in 50 mM Tris-HCl, pH 8.0, 200 NaCl, 1 mM $ZnSO_4$, 10% glycerol (12.9 mg/mL)
- Morpheus II HT-96
- SG1 Screen

Protocol:

1. Four different plates were set up in room temperature (ca. 20 ºC). Two plates with PerR in 50 mM Tris-HCl, pH 8.0, 200 NaCl, 1 mM $ZnSO_4$, 10% glycerol (12.9 mg/mL) using the SG1 and Morpheus screening conditions. One plate with PerR in 100 mM MES-NaOH, pH 6.5, 150 NaCl, 5% glycerol (8.5 mg/mL) and one plate with PerR in 100 mM MES-NaOH, pH 6.5, 150 NaCl, 15% glycerol (13.6 mg/mL) in SG1 screening conditions.
2. One 50 µL aliquot per plate with PerR in respective buffer was quickly thawed between the fingers and put on ice.
3. One 96-well 3-drop Swissci plate per sample was set up with 30 µL of the SG1 **OR** Morpheus screening conditions and the protein was pipetted onto the plate with a mosquito crystallization robot as a sitting drop experiment with three different concentrations: 1:3, 1:1, and 3:1 protein to precipitant for each condition in 200 nl drops. When the plate was done it was stored in a plate hotel at 20 ºC.

4 Results

The most important results of the project are presented here. From the cloning work to the characterization of PerR, the results are shown in chronological order and are briefly described.

4.1 Cloning

The cloning attempt was successful. In the plasmid fingerprinting experiment restriction enzymes were chosen so that if the cloning was succesful either a 500 or 600 bp fragment would show on a agarose gel. An agarose gel on the plasmid fingerprinting showed that all four colonies tested were positive, see figure 15, and had picked up the right gene construct. Two of these plasmids were sent for sequencing, confirming that the cloning had succeeded.

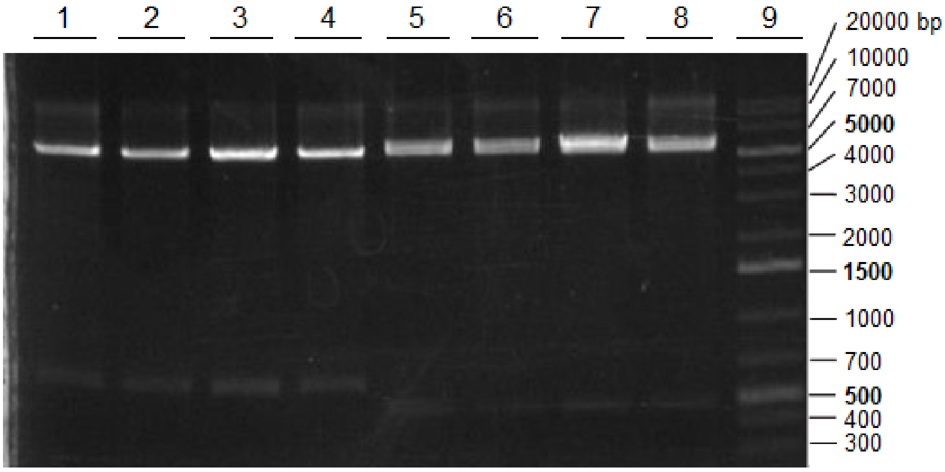

Figure 15: Gel electrophoresis on the digested colony plasmids. The samples were run on a 1 % agarose gel (8 cm) at 120 V (15 V/cm) in 1x TAE for 30 min. GeneRuler 1 Kb Plus DNA Ladder was used as a reference, and the scale is in nr. of base pairs. From left to right, lane 1-4 is digestion reaction 1 which should show a band corresponding to 600 bp if positive, lane 5-8 digestion reaction 2 which should show a band corresponding to 500 bp if positive, and the last lane, lane 9 is the ladder. All tested colonies were positive.

4.2 Expression & Purification

Several tests were performed regarding the expression of PerR. This was done with a slightly different construct. This construct had a 6xHis-tag instead of a Strep-tag, but the results should still be comparable. From the results, the best strain was determined to be BL21-AI. Autoinduction seemed to give the highest yield. There was only a small increase in the yield by letting the culture express overnight and the increase in the yield was mostly explained by a greater cell culture density, see figure 16.

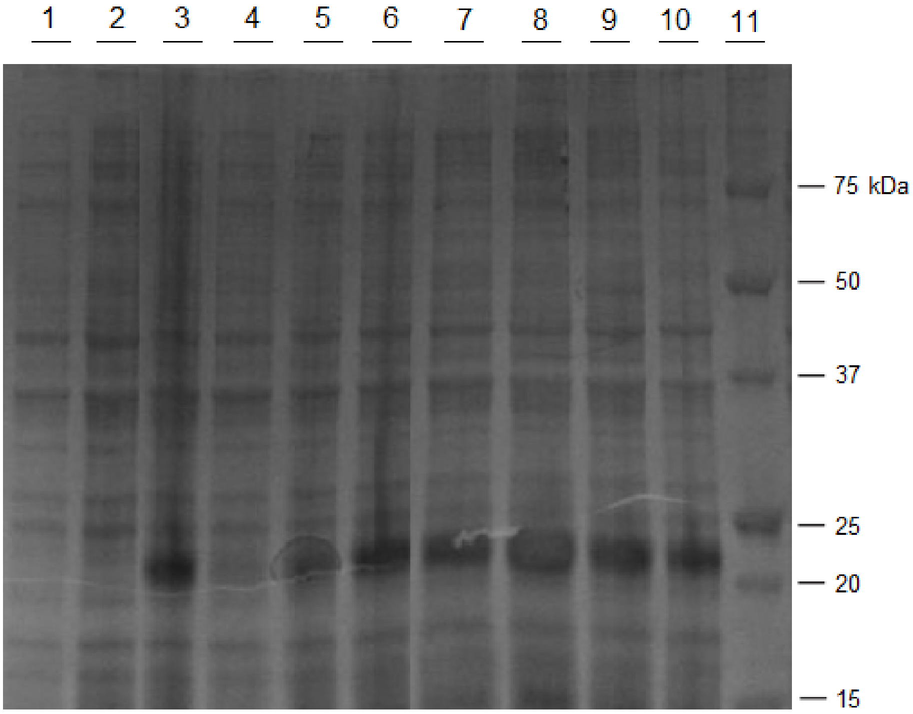

Figure 16: SDS-PAGE on the different expression tests. Lane 1: Rosetta strain, autoinduced at 20 °C. Lane 2: BL21-AI strain, autoinduced at 20 °C. Lane 3: BL21-AI strain, induced with 0.1 mM IPTG;0.1% L-arabinose at 20 °C. Lane 4: Rosetta strain, autoinduced at 22 °C. Lane 5: BL21-AI strain, autoinduced at 22 °C. Lane 6: BL21-AI strain, induced with 0.1 mM IPTG;0.1% L-arabinose at 22 °C. Lane 7: Rosetta strain, autoinduced at 37 °C. Lane 8: BL21-AI strain, autoinduced at 37 °C. Lane 9: BL21-AI strain, induced with 0.1 mM IPTG;0.1% L-arabinose at 37 °C. Lane 10: BL21-AI strain, induced with 0.1 mM IPTG;0.1% L-arabinose at 37 °C (harvested 3h later). Each sample was mixed with 4X SDS loading dye in a 1:3 ratio and boiled on a heating block at 95 °C for 5 min before 2 µL of each sample was loaded on a Mini-PROTEAN TGX Precast Gels. The gel was run for 35 min at 200 V.

PerR could be expressed in large quantities and be purified by a combination of chromatography methods, described in detail in the method section. After the first purification step with Tactin affinity chromatography, PerR was obtained with high purity. PerR was expressed in such large quantities that the column became supersaturated by the protein, and much was therefore eluted in the washing steps, see figure 17.

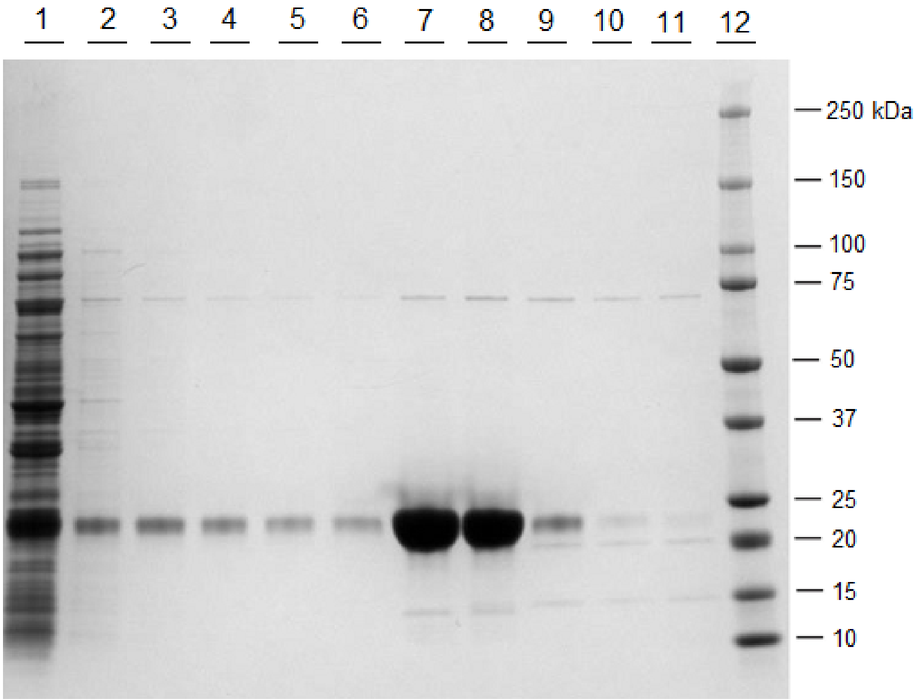

Figure 17: SDS-PAGE on the fractions from the Tactin-affinity chromatography step. Lane 1-5 is the washing fractions, lane 6-11 the elution fractions and lane 12 the ladder. Precision Plus Protein standard was used as a ladder. Each sample was mixed with 4X SDS loading dye in a 1:3 ratio and boiled on a heating block at 95 ºC for 5 min before 2 µL of each sample was loaded on a Mini-PROTEAN TGX Precast Gels. The gel was run for 35 min at 200 V.

Size exclusion after the addition of TEV protease revealed several forms of PerR. The size exclusion chromatogram also shows that the dimeric form of PerR is impossible to separate based on size from the TEV protease. The majority of PerR was in the monomeric fraction that could be resolved and seperated from the TEV protease, see figure 18.

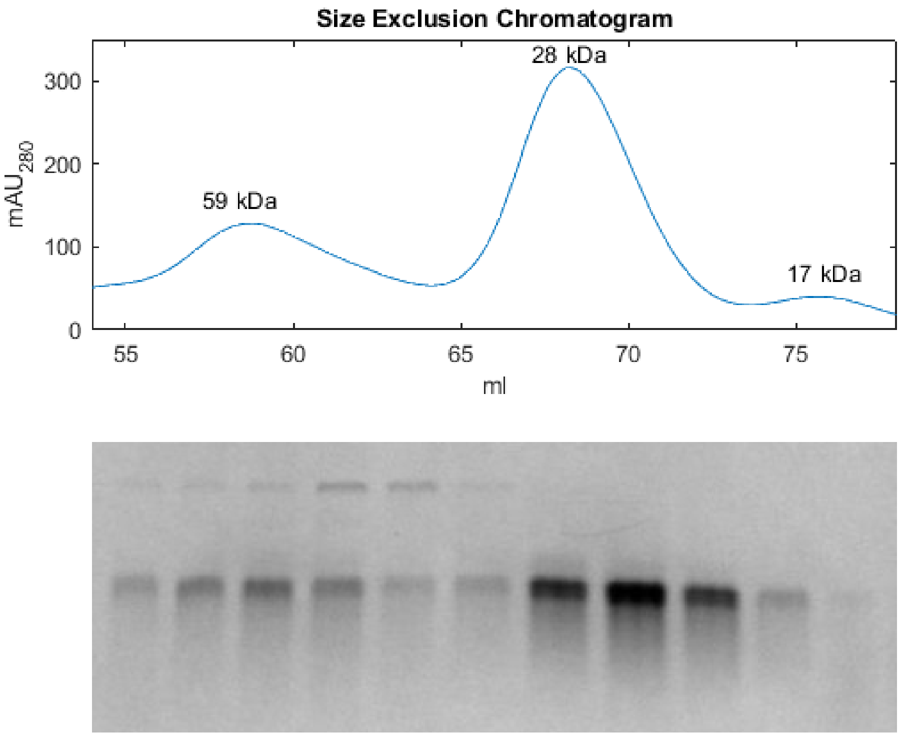

Figure 18: Size exclusion chromatogram with a SDS-PAGE underneath corresponding to the fractions collected during the run. The size corresponding to the peak is indicated. The size was determined from linear interpolation of a calibration run that was done 080817 in 25 mM Tris, pH7.5, 150 mM NaCl on bovine serum albumin (66 kDa) and α-chymotrypsin (25 kDa). The size exclusion chromatography was carried out with a ÄKTApurifier (GE Healthcare) on a HiLoad 16/60 Superdex 75 column with Buffer W: 100 mM Tris, pH 8.0, 150 mM NaCl. Each sample was mixed with 4X SDS loading dye in a 1:3 ratio and boiled on a heating block at 95 °C for 5 min before 2 µL of each sample was loaded on a Mini-PROTEAN TGX Precast Gel. The gel was run for 35 min at 200 V.

After the SEC and the subsequent concentration of PerR, it had been purified to > 95% degree purity as analyzed by SDS-PAGE, enough for crystallization screening and other experiments, see figure 19.

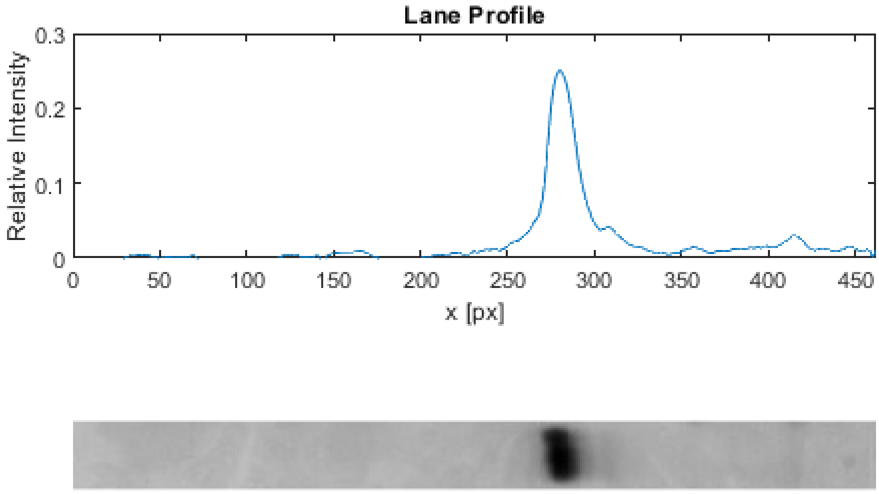

Figure 19: Lane profile from a SDS-PAGE. The mean background intensity was subtracted from the picture and the intensities from each pixel in the lane were summed in the y-directon of the picture, normalized and plotted against the x-axis of the picture in pixels. Each sample was mixed with 4X SDS loading dye in a 1:3 ratio and boiled on a heating block at 95 °C for 5 min before 2 µL of each sample was loaded on a Mini-PROTEAN TGX Precast Gel. The gel was run for 35 min at 200 V.

4.3 Characterization

4.3.1 Buffer Optimization

How the melting temperature of PerR depended on buffer conditions was examined in a thermal shift assay using the commercial screen JBScreen Solubility HTS. There seem to be two unfolding events occuring, represented as dips in the melt curve derivative as the temperature is increased. The buffers with the highest melting temperature were buffers containing N-(2-Acetamido)iminodiacetic acid (ADA). These buffers seemed to lack the first dip in the melt curve derivative, see figure 20.

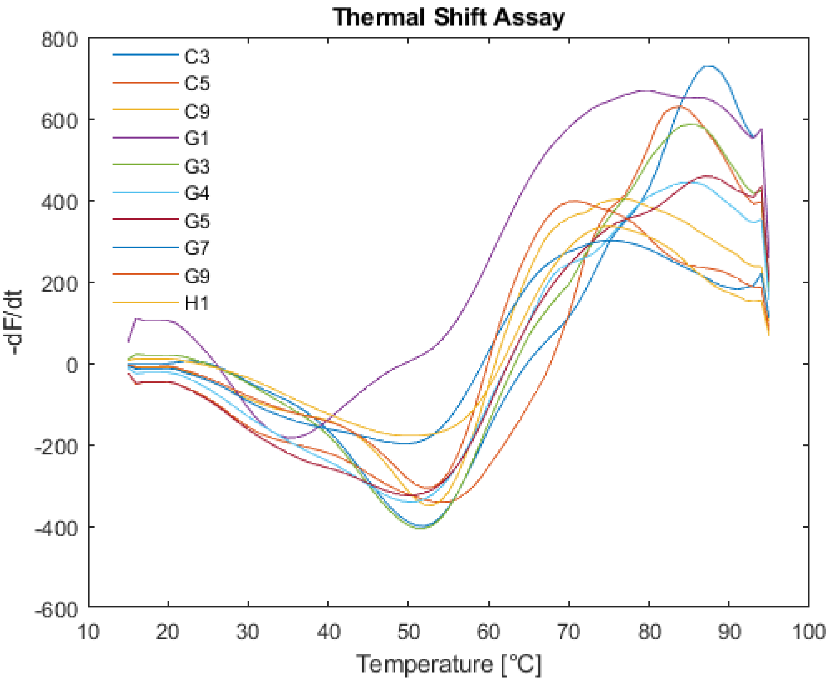

Figure 20: Thermal shift assay using the commercial screen JBScreen Solubility HTS with a CFX ConnectReal-Time PCR Detection System (BioRad). The figure shows only the reaction conditions where Tm > 50 °C. C3: 100 mM ADA, pH 6.0, 150 mM NaCl. C5: 100 mM Bis-Tris, pH 6.5, 150 mM NaCl. C9: 100 mM ADA, pH 7.0 150 mM NaCl. G1: 100 mM Potassium Acetate, pH 5.0, 150 mM NaCl 5% (v/v) glycerol. G3: 100 mM ADA, pH 6.0, 150 mM NaCl 5% (v/v) glycerol. G4: 100 mM MES, pH 6.5, 150 mM NaCl 5% (v/v) glycerol. G5: 100 mM Bis-Tris, pH 6.5, 150 mM NaCl 5% (v/v) glycerol. G7: 100 mM PIPES, pH 7.0, 150 mM NaCl 5% (v/v) glycerol. G9: 100 mM ADA, pH 7.0, 150 mM NaCl 5% (v/v) glycerol. H1: 100 mM DIPSO, pH 7.5, 150 mM NaCl 5% (v/v) glycerol.

Using the best buffer candidate from the solubility screen, 100 mM MES, pH 6.5, 150 mM NaCl 5% (v/v) glycerol, another thermal shift assay was conducted with additives. This revealed a preference for denaturing agents and glycerol, see figure 21. The reason for not choosing ADA, which had the highest melting temperatures, was because of its chelating properties that would have interfered with PerR's biological function.

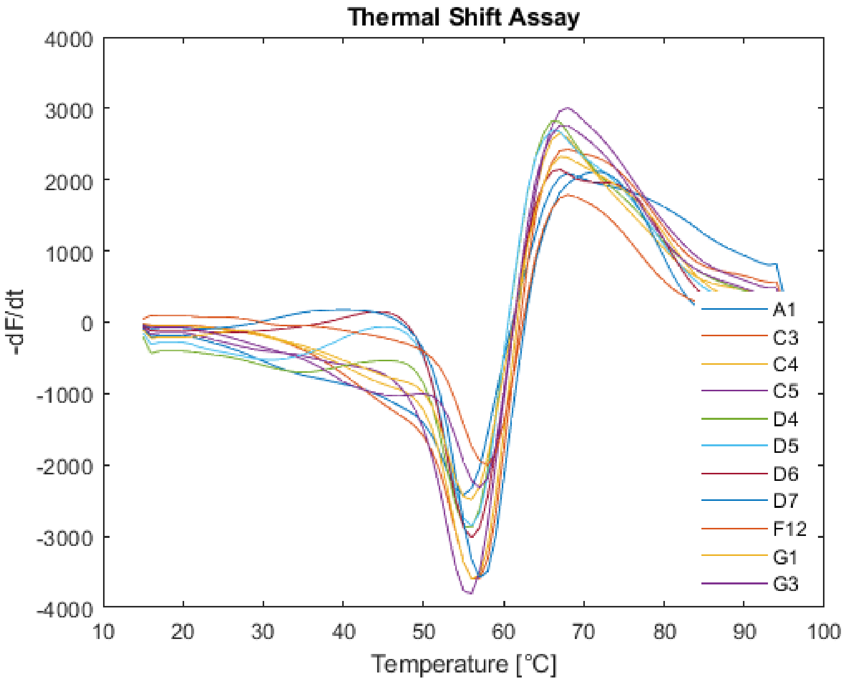

Figure 21: Thermal shift assay using the commercial screen rubic Additive screen with a CFX ConnectReal-Time PCR Detection System (BioRad). The figure shows only the reaction conditions where Tm > 50 ºC. A1: Ultrapure water. C3: 100 mM NaI. C4: 100 mM KI. C5: NaBr. D4: 0.5 M Urea. D5: 1 M Urea. D6: 2 M Urea. D7: 4 M Urea. F12: 20% (v/v) glycerol. G1: 5% (v/v) PEG 400. G3: 5% w/v PEG 3350.

Using the knowledge gained from the previous experiments a protein stability analysis was carried out in a Tycho NT.6 (NanoTemper) to determine the optimum salt and glycerol levels since both seemed to have a impact on the stability of PerR. In terms of stability PerR shows a preference for high levels of glycerol and salt, see figure 22.

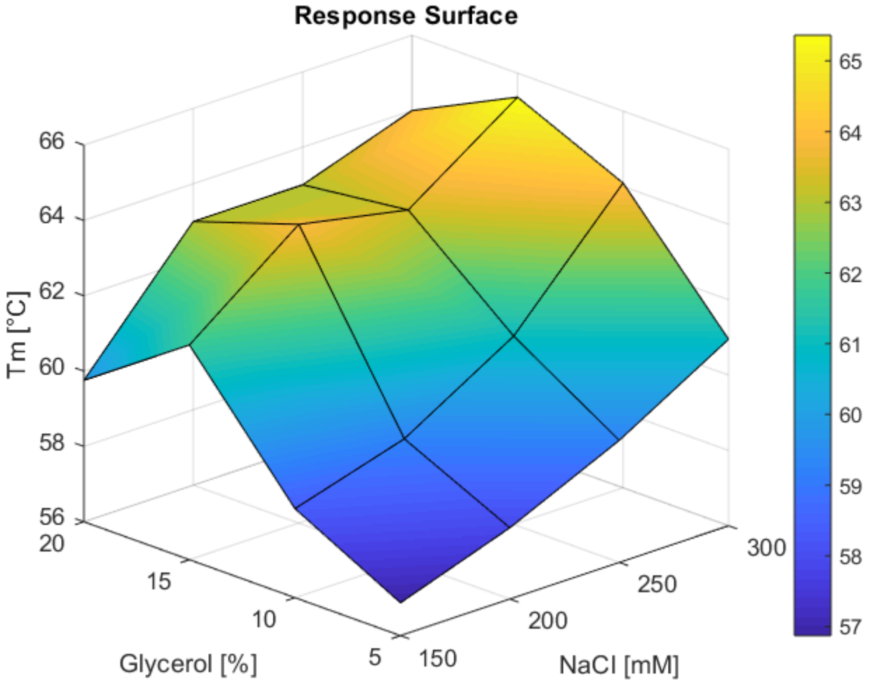

Figure 22: Response surface plot of the buffert optimization experiment showing the impact of different salt and glycerol concentrations on the melting temperature. The buffer used in the experiment was 100 mM MES, pH 6.5 and the experiment was carried out in a Tycho NT.6 (NanoTemper). The melting points were extracted from the raw data using Matlab.

Total reflection X-ray fluorescence (TXRF) was used to determine the metal conent of PerR. This experiment was carried out by *Dirk Maurer* from Julia Griese´s lab and showed only trace amounts of iron, manganese, copper, and nickel. The TXRF experiment also showed that PerR contained about 11% of zinc, see table 14 and figure 23.

Table 14: Metal concentrations found in the TXRF experiment.

Metal	Ratio PerR:Metal	Concentration
Mn	99400 : 1	0.8 µM
Fe	8200 : 1	6.6 µM
Ni	3110 : 1	2.5 µM
Cu	755 : 1	60.8 µM
Zn	10.6 : 1	85.6 µM

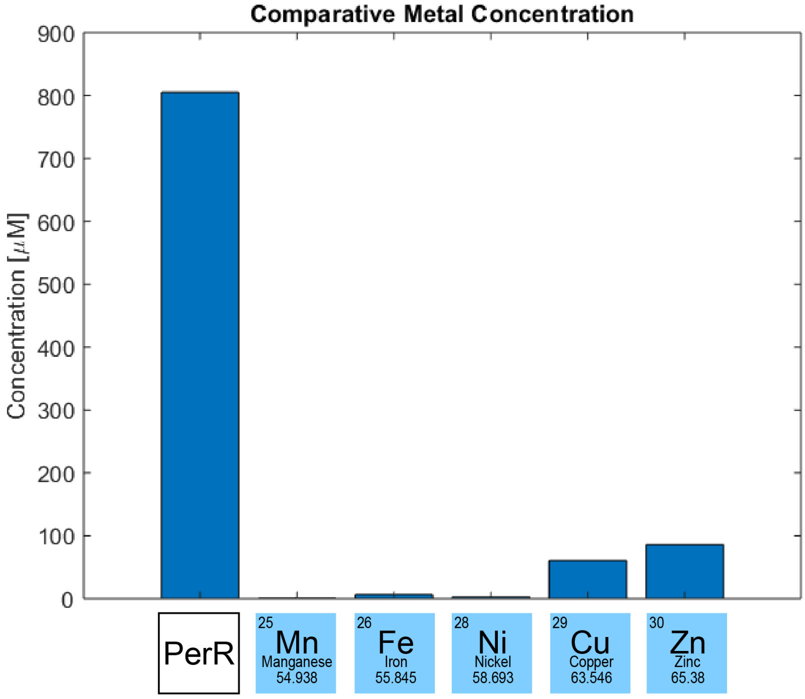

Figure 23: Metal concentrations found in the TXRF experiment compared to the concentration of PerR.

4.3.2 Small-angle X-ray Scattering

The data from the small-angle X-ray scattering experiment showed that the signal to noise changed in response to the addition of various metal ions and DNA. The dimer peak at approximately image number 400, is the smallest relative to the sharp monomer peak at image number 430 for the sample treated with EDTA, higher for the sample with zinc added and highest when zinc, manganese, and DNA are added. PerR seems to aggregate or oligomerize which can be seen on the slope of the calculated radius of gyration, see figure 24.

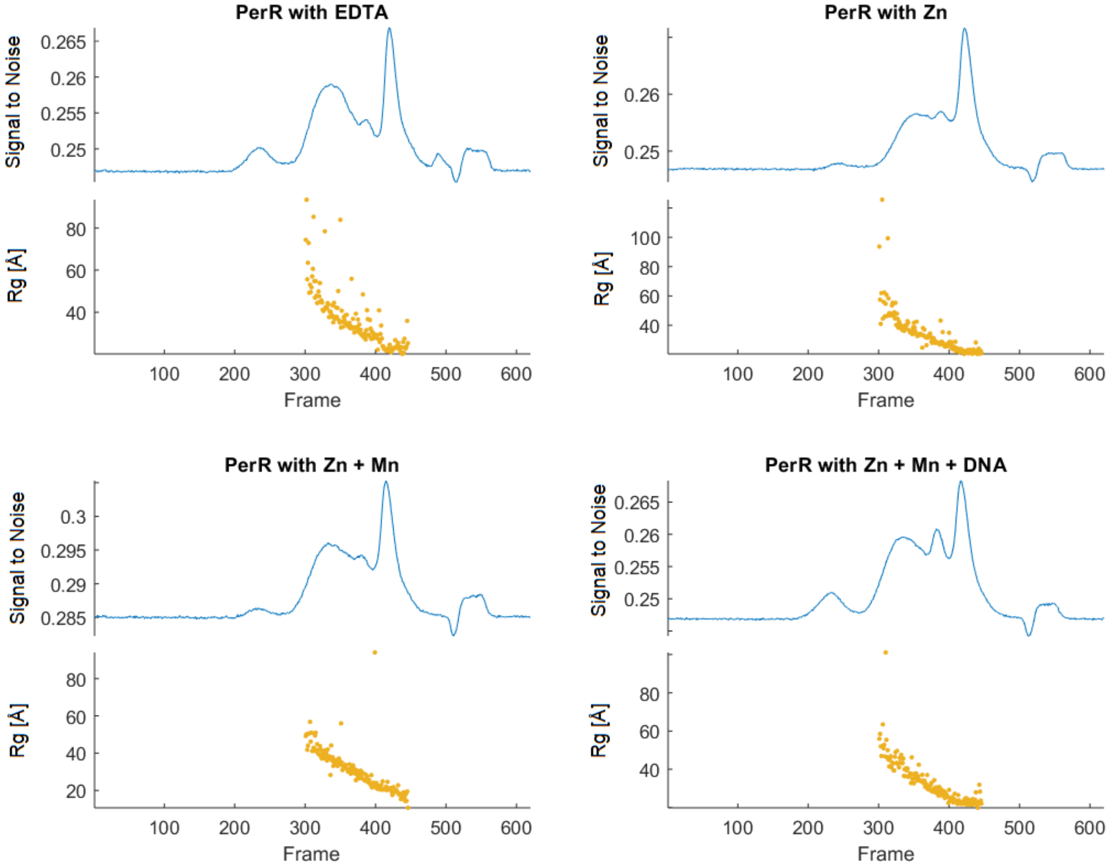

Figure 24: SEC-SAXS Signal Plot. Each point represents the integrated area of the ratio of the sample SAXS curve to the estimated background. The yellow scatter plot is the estimated radius of gyration. The relative dimer (frame 400) to monomer peak (frame 430) changes in response to the additives. The slope of the scatterplot indicates the oligomerization of the protein.

The Guinier analysis on the SAXS data of the dimeric peak with additives showed no problems with intermolecular effects such as aggregations or repulsion. Furthermore, the R_g and the maximum intramolecular distance of the particles, D_{max}, varied between the samples, see table 15. This indicates that the shape of the PerR changes its compactness in response to the addition of metal ions and ligand. The Kratky Plot shows that PerR differs from a perfect globular protein and is slightly elongated. It also seems that PerR is not perfectly folded and the spread for large **q**-values suggests that the sample may be polydisperse. From the scattering pattern and the fitted curve, the protein seems to have a dumbbell shape or a rod, see the figures 25-28.

Table 15: Calculated values from the SAXS data of the dimeric peak.

Experiment	R_g	D_{max}
PerR with EDTA	34.0	78.42
PerR with Zn	26.0	81.28
PerR with Zn + Mn	27.7	79.88
PerR with Zn + Mn + DNA	26.9	80.56

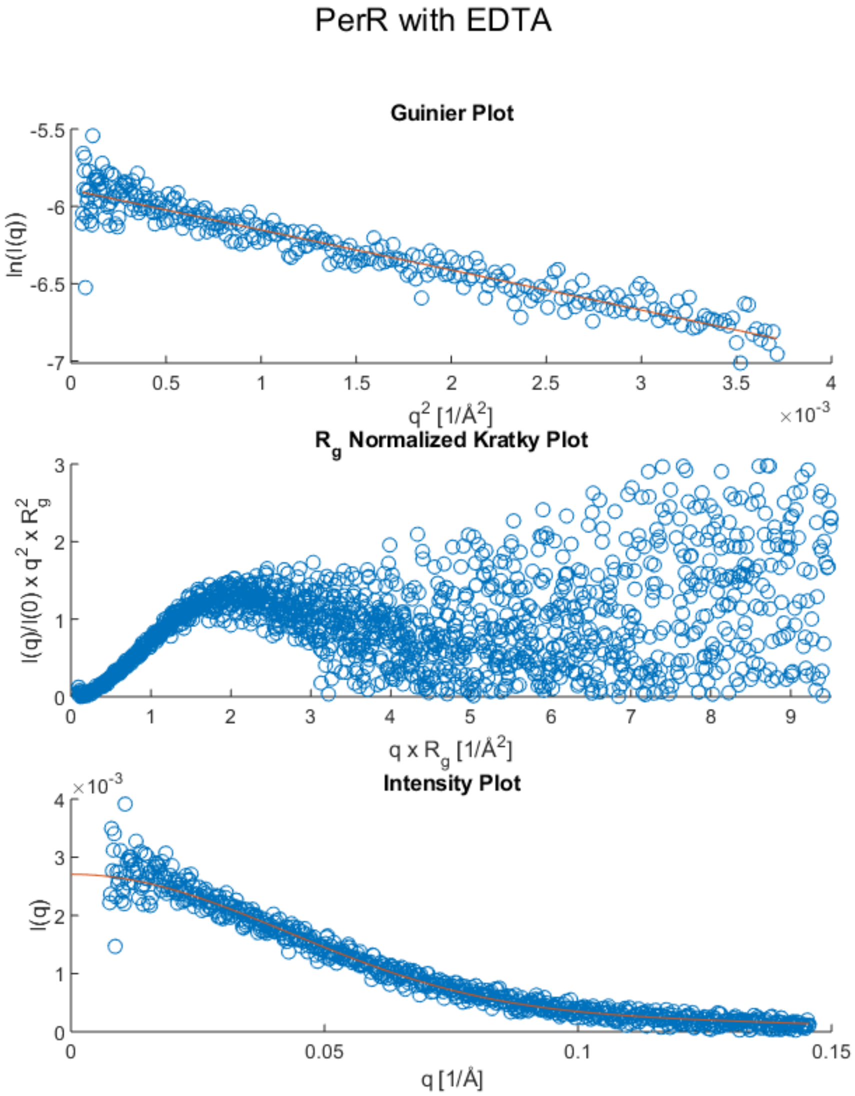

Figure 25: Guinier plot, a dimensionless Kratky plot, and the corresponding intensity plot for PerR with EDTA added in a ratio of 1:4. The Guinier plot shows the linear relationship between the natural logarithm of the intensity and low q^2-values. The normalized Kratky plot is based on the values obtained from the Guinier fitting and shows the folded state of the protein. The intensity plot shows the fitted curve used to calculate the $p(r)$ function.

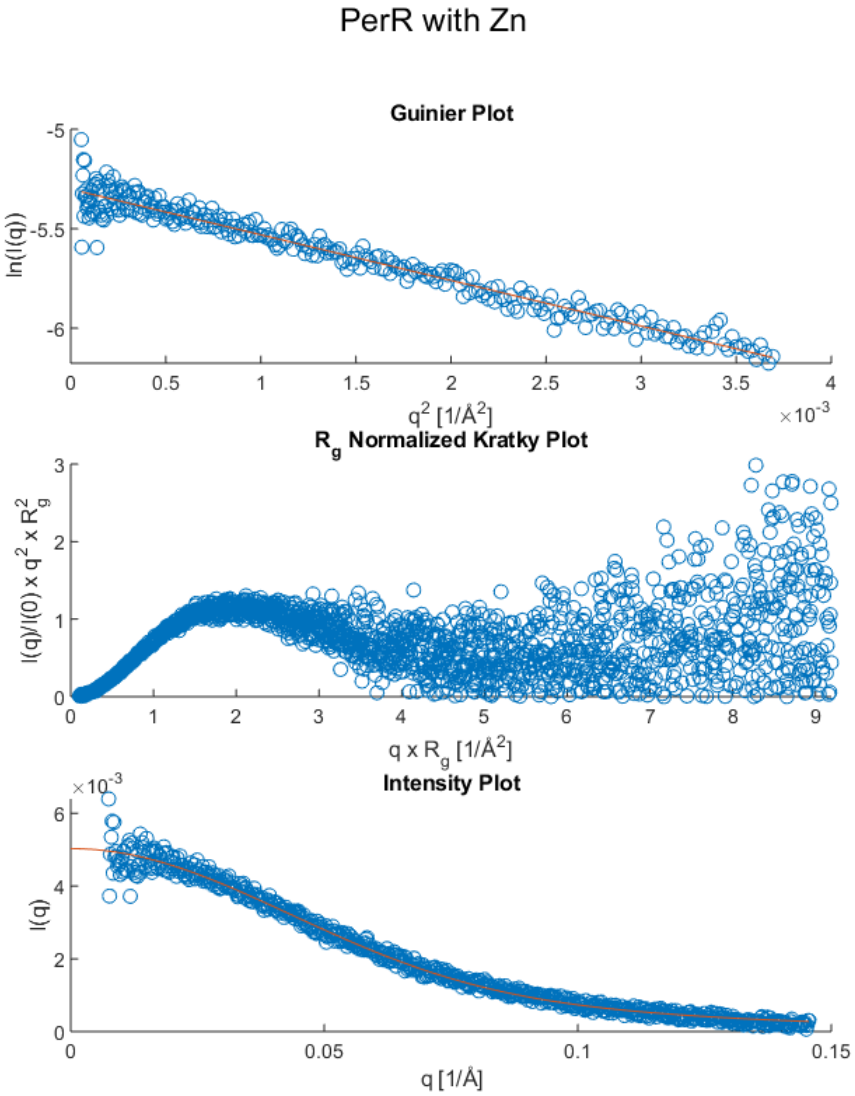

Figure 26: Guinier plot, a dimensionless Kratky plot, and the corresponding intensity plot for PerR with Zn added in a ratio of 1:2. The Guinier plot shows the linear relationship between the natural logarithm of the intensity and low q^2-values. The normalized Kratky plot is based on the values obtained from the Guinier fitting and shows the folded state of the protein. The intensity plot shows the fitted curve used to calculate the $p(r)$ function.

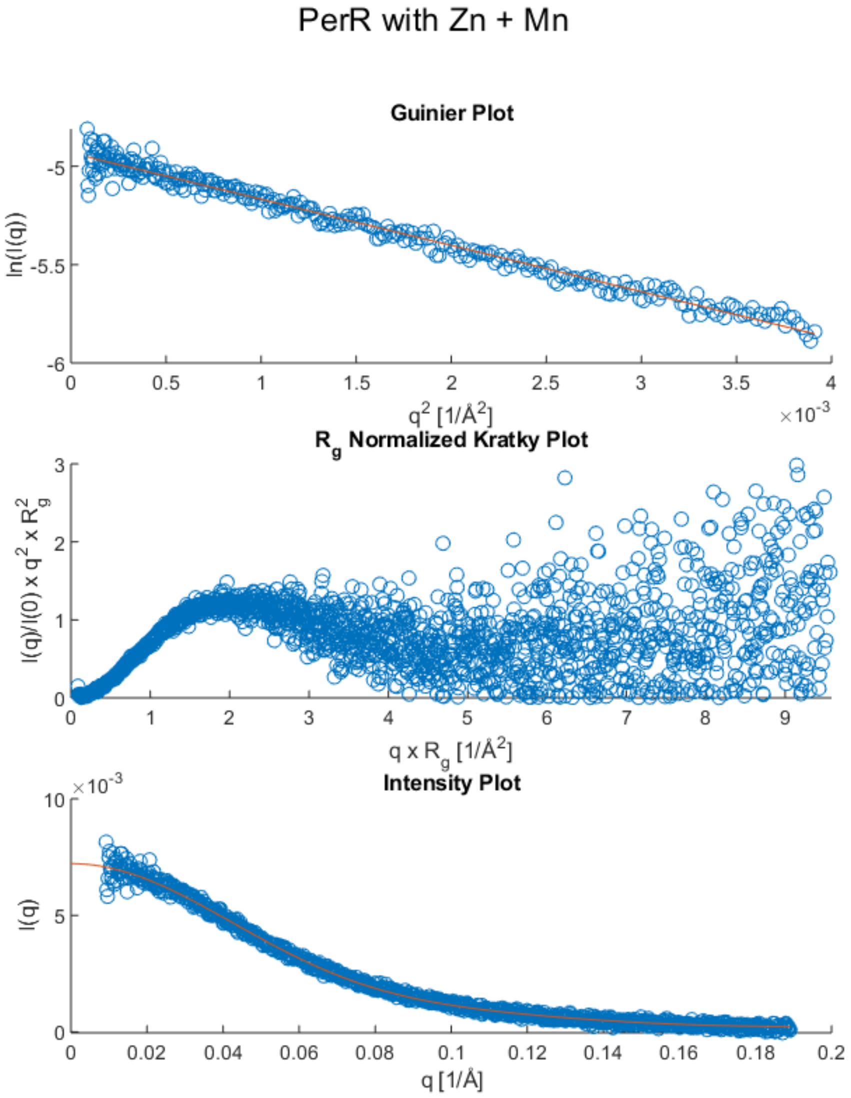

Figure 27: Guinier plot, a dimensionless Kratky plot, and the corresponding intensity plot for PerR with Zn and Mn added in a ratio of 1:2:3. The Guinier plot shows the linear relationship between the natural logarithm of the intensity and low q^2-values. The normalized Kratky plot is based on the values obtained from the Guinier fitting and shows the folded state of the protein. The intensity plot shows the fitted curve used to calculate the $p(r)$ function.

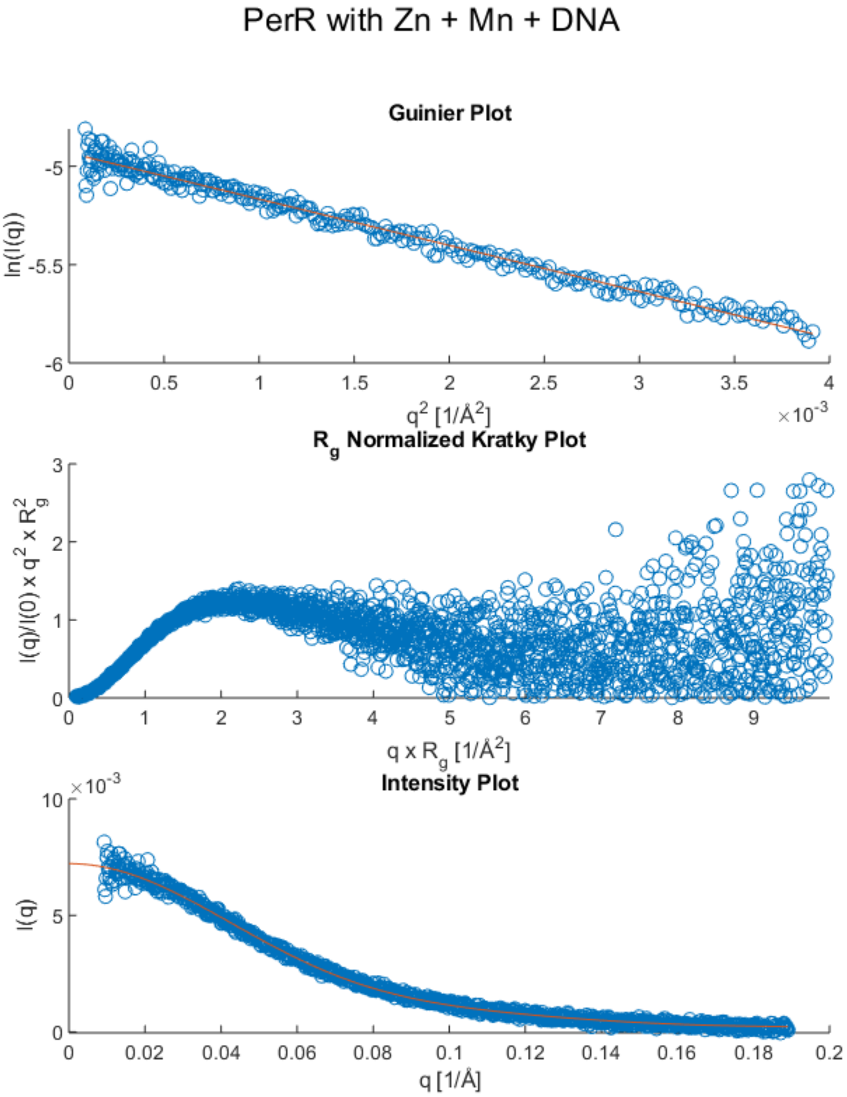

Figure 28: Guinier plot, a dimensionless Kratky plot, and the corresponding intensity plot for PerR with Zn, Mn, and DNA added in a ratio of 1:2:3:1/25. The Guinier plot shows the linear relationship between the natural logarithm of the intensity and low q^2-values. The normalized Kratky plot is based on the values obtained from the Guinier fitting and shows the folded state of the protein. The intensity plot shows the fitted curve used to calculate the $p(r)$ function.

The pair distance distribution function from the reverse transformation of the fitted data reveals that PerR becomes more elongated during treatment with EDTA, more compact in the addition of metal ions and most compact in the addition of both metal ions and

DNA. Here, too, one can see that PerR has an elongated, dumbbell shape. It also seems that the elongated region is more pronounced in the sample with metal ions and DNA, indicating that there might be a protruding portion of the protein, see figure 29.

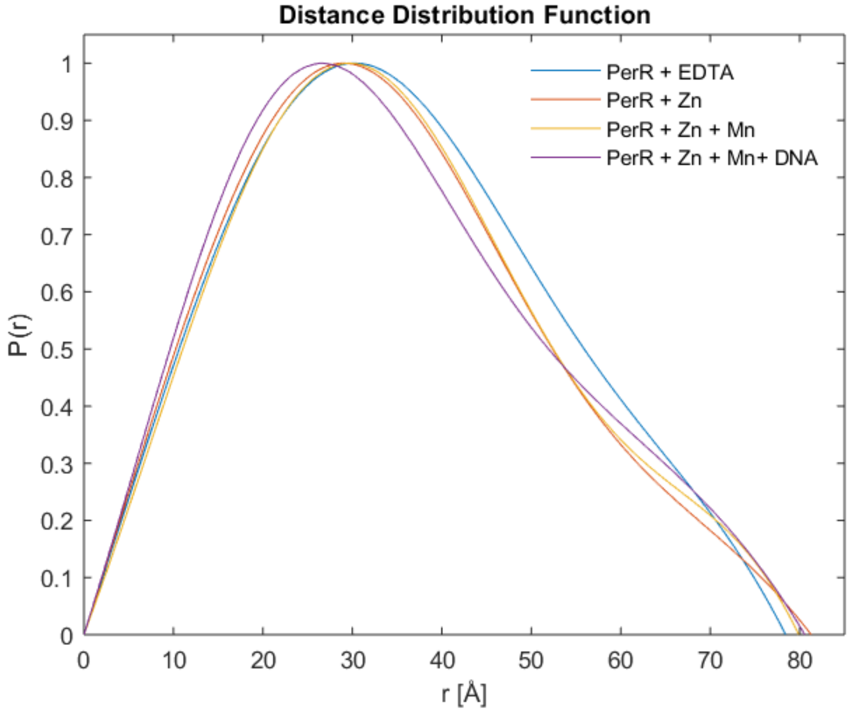

Figure 29: Distance distribution function .The plot shows the distribution of the summed pairwise intraparticle distances. The curves were normalized between [0 1] and then superimposed. The y-axis is in arbitrary units and is not to be taken as the true probability.

Analyzing the scattering data with the DAMMIF tool from the ATSAS package, several interesting features could be seen. The dumbbell shape was confirmed by the models, and a portion is indeed protruding out of the sample with PerR and the DNA. Furthermore, the models roughly fit with the homologous PerR from *B. subtilis*, see figures 30 and 31. The monomer peak for the sample treated with EDTA was also analyzed with DAMMIF and resulted in a shape that was consistent with the size of the monomer from *B. subtilis* but did not match completely. For the monomer treated with EDTA there is an extra protruding part compared to the monomer with zinc added, see figure 30.

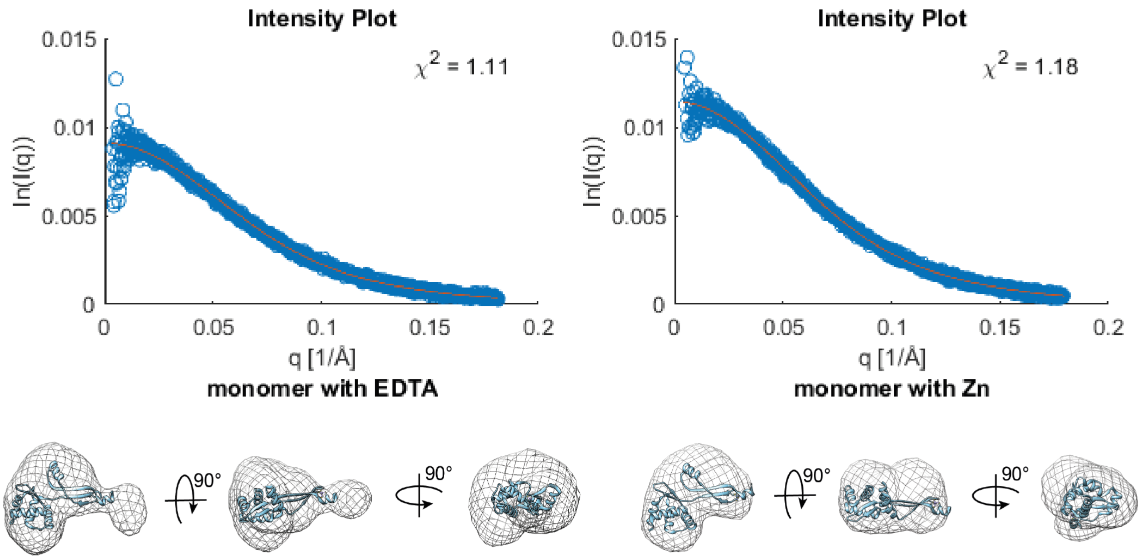

Figure 30: The averaged DAMMIF model result out of 20 runs with the experimental data from the monomeric peak for PerR treated with EDTA and zinc. The mesh represents the averaged model from the DAMMIF run with one PerR monomer from B. subtilis superimposed, shown in blue. No symmetry was applied. The final DAMMIF models had a χ^2-value of 1.057. The intensity plot above is the scattering from the experimental data with the calculated scattering curve from the superimposed monomer. PDB ID: 2FE3.

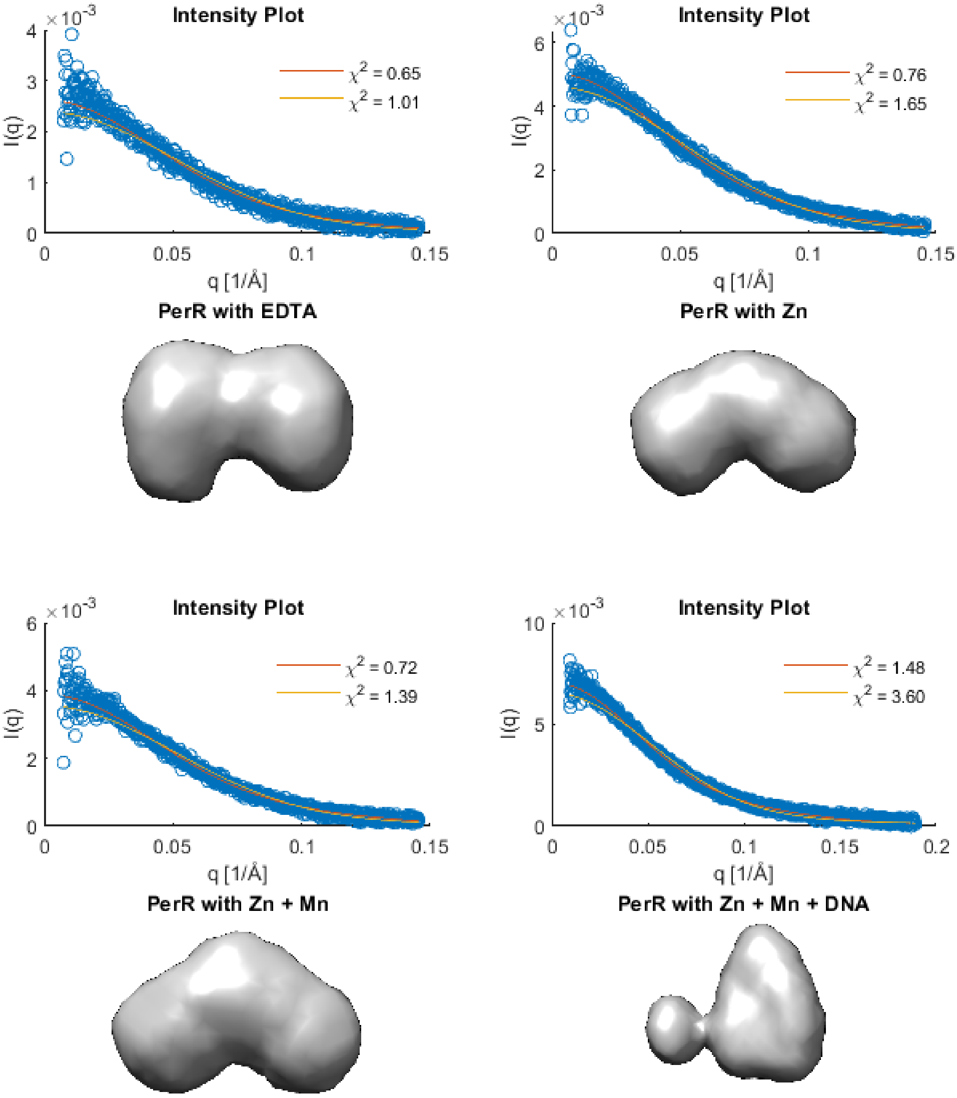

Figure 31: The averaged DAMMIF model result for 20 different runs on the experimental data from the dimeric peak for each sample. The final DAMMIF models had a χ^2-value of 0.61 for PerR with EDTA, 0.89 for PerR with Zn, 0.67 for PerR with Zn + Mn, and 1.22 for PerR with Zn + Mn + DNA. No symmetry was applied. Above the model is the scattering plot of the experimental data and calculated scattering curves for PerR in its apo and holo form shown in red and yellow respectively. PDB ID: 2FE3 (apo), 3F8N (holo). Visualized with UCSF Chimera.

The analysis of the EOM results suggested that PerR exists in multiple conformations in response to the addition of metals, see figure 32. When zinc is added it is still in an elongated but flexible conformation, and upon addition of manganese, it exists predominantly in the elongated conformation with a small fraction in a more compact conformation resembling the holo-form of PerR, see table 16.

Table 16: Calculated values from the EOM analysis.

Experiment	**Fraction**	R_g	D_{max}
PerR with EDTA	1.000	28.45	100
PerR with Zn	0.091	27.92	97.94
	0.738	27.24	94.24
	0.182	26.46	88.90
PerR with Zn + Mn	0.929	28.60	101.20
	0.081	23.81	80.75

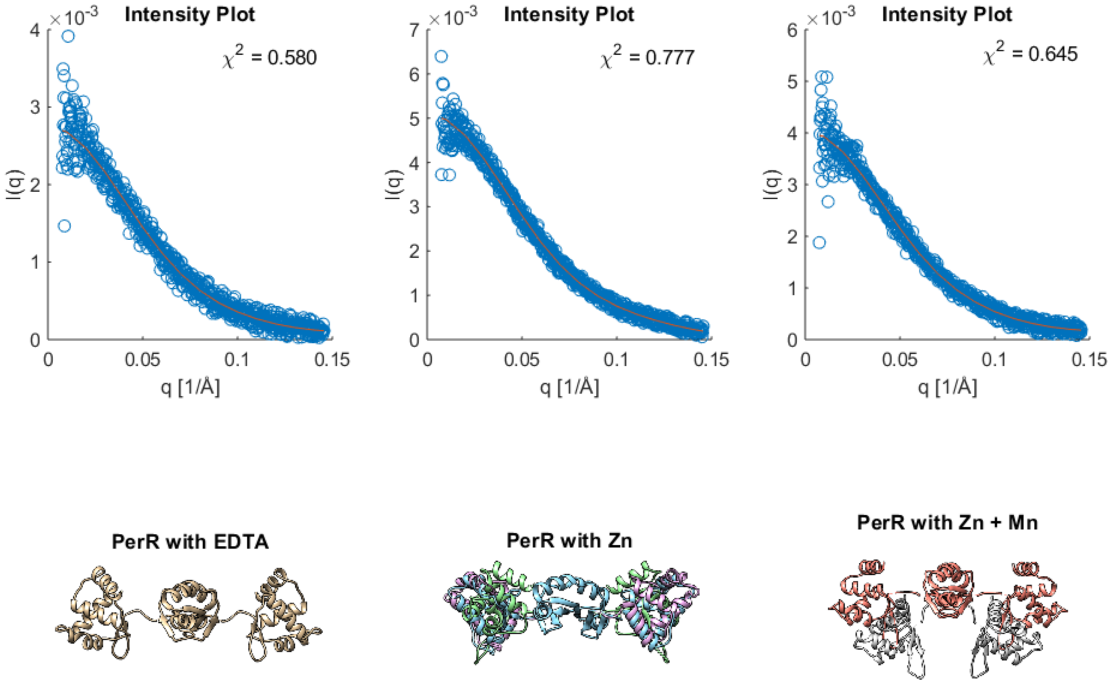

Figure 32: EOM results. The apo structure of PerR from text B. subtilis was used as a starting model, PDB ID: 2FE3. The three domains were treated as rigid bodies connected with random coil linkers in between. In the top row, the scattering data from the dimeric peaks of the different samples and the final fit for the model with the corresponding χ^2-value and in the bottom row, the different states that the algorithm found superimposed over each other.

4.4 Crystallization

The crystallization screens showed a lot of phase-separations in many different conditions. In conditions containing citrate as buffer crystals rapidly formed only to dissolve after a while. These crystals were fluorescent under UV light. One crystal seemed to grow in condition A7.1 (2.0 M ammonium sulfate, 0.1 M sodium HEPES, pH 7.5, 2% (v/v) PEG 400) in one of the SG1 screens containing PerR in 50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 10% (v/v) glycerol, 1 mM $ZnSO_4$, but it was too small to “fish”. This crystal was fluorescent under UV light and showed weak birefringence when viewed with a light microscope using a rotatable polarizing filter.

Upon the addition of 1 mM $ZnCl_2$, the protein samples formed a light grey and cloudy precipitate of low density. The precipitate could be redissolved, at least by visual inspection, by pipetting.

5 Discussion

5.1 Cloning

The cloning succeeded after some unsuccessful attempts. This was confirmed both by the plasmid fingerprinting experiment, see figure 15, as well as DNA sequencing. The reason for the difficulties was probably because the construct was slightly toxic to the cells and thus had a strong selective pressure to lose the plasmid. This toxicity was likely due to the DNA binding properties of PerR. Since it was from another organism it would, therefore, have DNA off-targets effects in the *E. coli* genome. Furthermore, by overexpressing a metal binding protein it would rapidly deplete the intracellular metal concentrations stressing the cells even more.

The critical step to success was probably the dephosphorylation of the ends of the plasmid that helped to lower the background of recirculated plasmids taken up in the transformation.

5.2 Expression & Purification

By growing the bacteria to a high density before inducing the PerR production and by inducing with high concentrations of IPTG and arabinose, a large amount of protein was obtained, see figure 17. Because PerR was moderately toxic, a high density was needed before the production could be turned on and once it was turned on, it was maximally induced to squeeze out as much protein as possible before the cells either died or inactivated the production of PerR, and overgrowing the culture. In the end, the efforts with the auto-induced media were abandoned even though it showed the highest expression levels of PerR in the small expression tests, see figure 16. This was due to problems when scaling up the operation, mainly because of poor aeration of the cultures. Foaming was a constant problem with the baffled Erlenmeyer flasks, effectively reducing the surface area for the gas exchange. When switching to bigger flasks without baffles and with a greater media to air volume the cultures grew very slowly, presumably because of the depletion of glucose before reaching a sufficiently high OD.

The purification was never fully optimized because the plasmid used in the first half of the project (pET28-6xHis-TEV-PerR) was replaced by a new one (pET28-NStrep-TEV-PerR) which required new purification steps. The protocol presented in the method section, however, is a good and reliable method that guarantees a good amount of protein. Some steps can be optimized, such as adding an IMAC step after the Tactin-affinity chromatography or using TEV with a Strep-tag. The wash fractions that shows a lot of PerR should also be saved and run over the column again in the Tactin-affinity chromatography step to make sure as much PerR is kept as possible. The yield was never properly calculated partly because I forgot and partly because of the technical difficulties faced when trying to optimize the protocol. Sometimes protein was lost due to sheer clumsiness and sometimes because of the need to repeat unoptimized steps which would have caused a very biased estimate of the yield. In the end, I got at least 23.4 mg protein from 4.8 L of cultivated bacteria, probably more like 30 mg in total as estimated from the pooled eluted fractions from the Tactin-affinity chromatography and the lost fractions in the washing steps, see figure 17.

The theoretical value of the molecular weight of PerR was calculated to 16.9 kDa using the ProtParam tool from the ExPASy web service (Gasteiger *et al.* 2005), but on the SDS-PAGE, PerR migrated as a 21 kDa protein. Since the SDS-PAGE is based on the principle that SDS binds and denatures protein to the same extent, linearly dependant on the size, differences in the calculated and observed molecular weight on an SDS page may occur when this relationship does not hold. It has been found that peptides with many acidic residues migrate more slowly in SDS-PAGE and that some sequences of amino acids can provide a significant size shift. PerR from the *S. erythraea* has a theoretical pI value of 5.24, calculated using the ProtParam tool, and contains many acidic residues that may explain the observation. An experimental ratio of the content of acidic amino acids and the expected difference in size has been determined as

$$y = 276.5x - 31.33$$

where x is the content of acidic amino acids between $[11.4\%, 51.1\%]$ and y represents the average ΔMW per amino acid (Guan *et al.* 2015). PerR contains 15.5% acidic amino acids which used as input in the formula gives an expected size shift of 4.3 kDa, which is consistent with the SDS-PAGE result within 0.2 kDa.

5.3 Characterization

The thermal shift assay on the solubility experiment shows that metal chelators stabilize PerR in solution. Presumably because it stabilizes PerR in its monomeric form by chelating the structural zinc in the dimerization domain. The first observed dip in the thermal shift assay is probably the dissociation event of the dimer which was mostly lacking for the buffers containing the chelator ADA, see figure 20.

Even though buffers containing ADA showed a high melting temperature, it was not advised to use this buffer with metalloproteins because of its strong chelating properties and would have probably interfered with downstream experiments. According to Ferreira *et al.* (2015) MES is a good buffer for metalloproteins without disturbing downstream experiments and its biological activity. Incidentally, the low pH of the buffer and MES itself is very good to keep the iron from oxidizing. Therefore the buffer with 100 mM MES, pH 6.5, 150 mM NaCl, 5% (v/v) glycerol was used for further investigation.

The same experiment with additives to the buffer mentioned above showed that the addition of denaturing agents strongly influenced PerR's stability. The melting curves for increasing amounts of urea seems to support the idea that the small dip before the melting point is when the dimers break apart, seen in figure 21. glycerol also seemed to have a stabilizing effect, maybe due to the stabilizing molecular crowding effect of glycerol (Perham *et al.* 2007).

To understand the effect of glycerol on the stability of PerR, the melt temperature of PerR was investigated with different concentrations of glycerol and salt. This showed a nonlinear relationship between salt and glycerol content and the melting temperature. The highest melting temperature was measured for the buffer with 15% (v/v) glycerol and 300 mM NaCl, see figure 22. Therefore, these concentrations were used in the final buffer which was 100 mM MES, pH 6.5, 300 mM NaCl 15% (v/v) glycerol.

From the TXRF experiment, see figure 23, we can see that very low levels of zinc are present in the protein sample. This explains why we see PerR mainly in its monomeric state which may not be surprising since the monomeric fraction from the SEC was used in the experiments.

5.3.1 Small-angle X-ray Scattering

The sloping line in the SEC-SAXS signal plot for the calculated R_g-value, figure 24, might indicate that some PerR is aggregated or oligomerized. This slope can be explained by the dissociation of monomers from this state as the equilibrium shifts on the column, and therefore a sloping trend is obtained in the calculated R_g-value of the particles. If this is true then the oligomeric state of PerR could be concentration dependent as this would explain the reversible association and dissociation seen in the elution profile. As the different fractions are separated on the column the equilibrium shifts in response to the lower concentrations of the monomeric PerR and thus the oligomers dissolve. The oligomer peak is also not a part off the void peak that occurs earlier in the elution profile and thus seems to be more ordered than a simple aggregation of proteins.

Furthermore, it can be seen that the relative size of the dimer peak compared to the monomer peak changes in response to the addition of ligands or metal chelator. With the addition of EDTA, the volume of the dimer peak is the smallest compared to the monomer peak, which is reasonable if EDTA destabilizes the dimer by chelating the zinc. When zinc is added, the dimer peak becomes a little bit bigger, by stabilizing the dimerization of PerR and in the addition of zinc and manganese, the dimer peak is unchanged since the manganese does not participate in the stabilization of the dimer. When both zinc, manganese, and DNA are added, the dimer peak has the biggest volume. What causes this is uncertain, but one thing is sure; the addition of DNA causes a visible change in the elution profile and the DAMMIF models, see image 24 and 31. Perhaps a monomer binds the DNA and therefore, has roughly the same size as a dimer causing both peaks to overlap. A previous electrophoretic mobility shift assay carried out by *Francisco Javier Marcos Torres* in Julia Griese's lab did not show any binding activity but this possibility should not be completely ruled out since the EMSA was carried out in a completely different buffer and temperature.

In the Guinier plot in the initial SAXS data analysis, one can see that all samples show a linear relationship for small **q**-values and have no intermolecular effects such as aggregation or repulsion that affects the data quality. In the Kratky plot, one can further see that all samples deviate slightly from a perfect globular protein and are somewhat elongated. The protein sample is not unfolded but still not completely folded. There is also a significant spread at high **q**-values which might indicate that the sample is not homogeneous, see figures 25-28. In table 15, we see that the calculated data for the fitted curve from the figures referenced above shows that PerR treated with EDTA has the largest R_g-value, which is reasonable as it should linearize PerR to its apo form. Since the data lacked standard deviations, this data should be interpreted with some caution.

From the distance distribution function, PerR generally seems to have an slightly elongated shape with two spheres at the ends. PerR treated with EDTA appears to have a more elongated structure with a more distinct dumbbell shape. Treated with Zn, PerR becomes slightly more compact, and with manganese added, the structure does not change significantly from the previous sample. With DNA added, the structure becomes more compact with a protruding part, see figure 29. The results from the DAMMIF modeling partly confirmed these observations. PerR treated with EDTA has a more elongated structure with a clear dumbbell shape. Treated with zinc or with zinc and manganese, PerR has a slightly curved structure. With both metal ions and the DNA, a triangular-shaped structure with a protruding part is obtained. This protruding part may be an artifact from the modeling or a part of the DNA helix to which a monomer is bound but this is purely speculative, see figure 31. A more reasonable explanation for this structure is that it is actually the monomer or a mixture of a monomer and a dimer being shown. Since the images selected for the analysis were from the second half of the dimer peak, it may have contained monomers because of the overlap between the peaks. This could explain the similar structure between this structure, and the structure found for the monomer.

Interestingly when PerR is treated with EDTA, PerR adopts a linear conformation as expected in the EOM analysis. When only zinc is added, PerR still has a linear structure, but with some flexibility, and when manganese is added, two conformations of PerR, an elongated and a bent conformation is seen. The bent conformation is probably holo PerR with manganese bound to it, see figure 32. Only 8% of PerR is in this conformation which may indicate that the remaining PerR is already oxidized and can no longer bind manganese as the oxidation of PerR irreversibly inactivates the metal binding site by the oxidation of the His residues, see table 13.

The apparent miss-match between the model found by DAMMIF for the monomeric PerR and the monomer from the crystal structure can be because of the dimerization of the monomer in the crystal. Since the monomer in the crystal is packed as a dimer, merely taking the monomer from this structure may not reflect its true conformation in solution. There is also a difference in the shape of the monomer for the PerR that was treated with EDTA and PerR with zinc added. The zinc atom is stabilizing two parts of the monomer, and if stripped away, the monomer should adopt a more relaxed structure, perhaps even with a long tail sticking out, which may explain the differences in the structures. The zinc is located just at the region where the density bulges out. But once again this may just be an artifact from the modeling due to a polydisperse sample, see figure 30.

The results discussed above are from the interpretations of models and should be treated with caution. To be more confident in the results, the experiments have to be repeated with the same conditions. Ultimately a crystal structure would yield more information and confirm these observations.

5.4 Crystallization

In the crystallization trials, a trend could be observed. In the wells containing citrate as a buffer, crystals were obtained almost immediately. These may have been protein crystals. The reason why they were suspected of being protein crystals, besides being fluorescent under UV light, was that they redissolved almost overnight. Salt crystals have strong ion-ion bonds and should not dissolve so easily once formed. The chelating properties of the citrate may have helped in the initial crystallization process by bringing PerR closer together. Because of its chelating properties it should be avoided when crystallizing the holo structure.

Another thought that emerged during the project was that to succeed in the crystallization of PerR, the conformations may need to be “frozen” in their respective positions. Either with the addition of manganese to get the holo structure or oxidized to get the apo structure and thus preventing the dimer from “flapping”.

6 Conclusions

It is evident that more work is needed to characterize PerR. The main focus should be on trying to prevent the oxidation of PerR when expressed by replacing the iron with manganese as soon as possible by chelating the bound iron with EDTA and then adding manganese or by growing the bacteria in media containing manganese to displace the iron in PerR.

During the purification of PerR, an IMAC step should be performed to get rid of the TEV protease. Since previous studies done by *Julia Griese* indicates that the Strep-tag blocks the DNA binding of PerR the tag has to be cleaved off; therefore the best solution would be to construct a TEV with a Strep-tag and get rid of it the same way as the uncleaved product in a reverse Tactin-affinity chromatography step.

There is a lot of biological data still to be analyzed from the crystallization experiments. Given more time, one could have tried to discern trends and extract information about how PerR reacts to different precipitants something that might say something about the nature of the protein.

The same can be said about the SAXS data that still contains lot of data that can be extracted. Especially the data concerning the monomeric form of PerR. Only two monomer peak were thoroughly analyzed, but it is reasonable to assume that the monomer also reacts to the other conditions, which could be of importance in later crystallization attempts.

To stabilize the dimer, enough zinc needs to be present when expressing PerR. One way to make sure that enough zinc is present is to add some to the growth media, ensuring that the cells need for zinc is satisfied. Adding zinc at a later stage is not recommended since it may cause precipitation of the protein as observed during the setup of a crystallization screen.

7 Acknowledgements

I would like to thank,

My supervisor, professor *Julia Griese*, who supported me through the project, gave expert advice and took me under her wings to do this project.

Researchers *Dirk Maurer* and *Francisco Javier Marcos Torres*, who, with their expertise in prokaryotic microbiology, structural biology, and biochemistry, helped me to carry out my experiments and supervise me in my daily work.

Scientist *Robert Gustafsson* who assisted me in the lab and shared his knowledge and expertise in structural biology and protein purification.

Everyone at the structural biology laboratory, Uppsala 2019, who made me feel at home and a part of the group.

8 References

Ahn BE, Baker TA. 2016. Oxidization without substrate unfolding triggers proteolysis of the peroxide-sensor, PerR. Proceedings of the National Academy of Sciences 113: E23–E31.

Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Meier-Kolthoff JP, Klenk HP, Clément C, Ouhdouch Y, van Wezel GP. 2016. Taxonomy, Physiology, and Natural Products of Actinobacteria. Microbiology and molecular biology reviews: MMBR 80: 1–43.

Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. 2007. Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. Journal of the American Chemical Society 129: 5656–5664.

Bernhard R. 2009. Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology. New York, 1 edition edition.

Bresgen N, Eckl PM. 2015. Oxidative Stress and the Homeodynamics of Iron Metabolism. Biomolecules 5: 808–847.

Chandrangsu P, Rensing C, Helmann JD. 2017. Metal homeostasis and resistance in bacteria. Nature Reviews Microbiology 15: 338–350.

Changela A, Chen K, Xue Y, Holschen J, Outten CE, O'Halloran TV, Mondragón A. 2003. Molecular basis of metal-ion selectivity and zeptomolar sensitivity by CueR. Science (New York, N.Y.) 301: 1383–1387.

Culotta VC, Daly MJ. 2013. Manganese Complexes: Diverse Metabolic Routes to Oxidative Stress Resistance in Prokaryotes and Yeast. Antioxidants & Redox Signaling 19: 933–944.

Deng Z, Wang Q, Liu Z, Zhang M, Machado ACD, Chiu TP, Feng C, Zhang Q, Yu L, Qi L, Zheng J, Wang X, Huo X, Qi X, Li X, Wu W, Rohs R, Li Y, Chen Z. 2015. Mechanistic insights into metal ion activation and operator recognition by the ferric uptake regulator. Nature Communications 6: 7642.

Dinos GP. 2017. The macrolide antibiotic renaissance: The present and future of macrolide antibiotics. British Journal of Pharmacology 174: 2967–2983.

Dudev T, Lim C. 2014. Competition among metal ions for protein binding sites: determinants of metal ion selectivity in proteins. Chemical Reviews 114: 538–556.

Faulkner MJ, Ma Z, Fuangthong M, Helmann JD. 2012. Derepression of the Bacillus subtilis PerR Peroxide Stress Response Leads to Iron Deficiency. Journal of Bacteriology 194: 1226–1235.

Ferreira CMH, Pinto ISS, Soares EV, Soares HMVM. 2015. (Un)suitability of the use of pH buffers in biological, biochemical and environmental studies and their interaction with metal ions – a review. RSC Advances 5: 30989–31003.

Fillat MF. 2014. The FUR (ferric uptake regulator) superfamily: Diversity and versatility of key transcriptional regulators. Archives of Biochemistry and Biophysics 546: 41–52.

Franke D, Svergun DI. 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of Applied Crystallography 42: 342–346.

Frawley ER, Fang FC. 2014. The Ins and Outs of Bacterial Iron Metabolism. Molecular microbiology 93: 609–616.

Gaballa A, Helmann JD. 2002. A peroxide-induced zinc uptake system plays an important role in protection against oxidative stress in Bacillus subtilis. Molecular Microbiology 45: 997–1005.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. Walker JM, editor, The Proteomics Protocols Handbook, Humana Press, Totowa, NJ, 571–607.

GE Healthcare. 2010. Gel Filtration Principles and Methods 123.

Grant TD, Luft JR, Carter LG, Matsui T, Weiss TM, Martel A, Snell EH. 2015. The accurate assessment of small-angle X-ray scattering data. Acta Crystallographica Section D: Biological Crystallography 71: 45–56.

Guan Y, Zhu Q, Huang D, Zhao S, Jan Lo L, Peng J. 2015. An equation to estimate the difference between theoretically predicted and SDS PAGE-displayed molecular weights for an acidic peptide. Scientific Reports 5: 13370.

Helmann JD. 2014. Specificity of Metal Sensing: Iron and Manganese Homeostasis in Bacillus subtilis. The Journal of Biological Chemistry 289: 28112–28120.

Hine RH, Martin EM. 2015. Electron transport chain. Hine R, Martin E, editors, A Dictionary of Biology, Oxford University Press.

Hong Enriquez RP, Do TN. 2012. Bioavailability of Metal Ions and Evolutionary Adaptation. Life : Open Access Journal 2: 274–285.

IBA Lifesciences. 2019. Protein Purification with Strep-Tactin®. `https://www.iba-lifesciences.com/strep-tactin-system-technology.html`.

Ilbert M, Bonnefoy V. 2013. Insight into the evolution of the iron oxidation pathways. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1827: 161–175.

Imlay JA. 2008. Cellular Defenses against Superoxide and Hydrogen Peroxide. Annual Review of Biochemistry 77: 755–776.

Irving H, Williams RJP. 1948. Order of Stability of Metal Complexes. Nature 162: 746.

Jacquamet L, Traoré DaK, Ferrer JL, Proux O, Testemale D, Hazemann JL, Nazarenko E, Ghazouani AE, Caux-Thang C, Duarte V, Latour JM. 2009. Structural characterization of the active form of PerR: insights into the metal-induced activation of PerR and Fur proteins for DNA binding. Molecular Microbiology 73: 20–31.

Jelić D, Antolović R. 2016. From Erythromycin to Azithromycin and New Potential Ribosome-Binding Antimicrobials. Antibiotics 5: 29.

Jiang J, He X, Cane DE. 2007. Biosynthesis of the earthy odorant geosmin by a bifunctional Streptomyces coelicolor enzyme. Nature chemical biology 3: 711–715.

Johnson DC, Dean DR, Smith AD, Johnson MK. 2005. STRUCTURE, FUNCTION, AND FORMATION OF BIOLOGICAL IRON-SULFUR CLUSTERS. Annual Review of Biochemistry 74: 247–281.

Joseph A Cotruvo J, Stubbe J. 2012. Metallation and mismetallation of iron and manganese proteins in vitro and in vivo: the class I ribonucleotide reductases as a case study. Metallomics 4: 1020–1036.

Kikhney AG, Svergun DI. 2015. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. FEBS Letters 589: 2570–2577.

Koch M, Vachette P, I Svergun D. 2003. Small-angle scattering: A view on the properties, structures and structural changes of biological macromolecules in solution. Quarterly reviews of biophysics 36: 147–227.

Konarev PV, Petoukhov MV, Svergun DI. 2001. MASSHA – a graphics system for rigid-body modelling of macromolecular complexes against solution scattering data. Journal of Applied Crystallography 34: 527–532.

Kozin MB, Svergun DI. 2001. Automated matching of high- and low-resolution structural models. Journal of Applied Crystallography 34: 33–41.

Lebeis SL, Paredes SH, Lundberg DS, Breakfield N, Gehring J, McDonald M, Malfatti S, Rio TGd, Jones CD, Tringe SG, Dangl JL. 2015. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. Science 349: 860–864.

Lee JW, Helmann JD. 2006. The PerR transcription factor senses H2o2 by metal-catalysed histidine oxidation. Nature 440: 363–367.

Ma Z, Jacobsen FE, Giedroc DP. 2009. Metal Transporters and Metal Sensors: How Coordination Chemistry Controls Bacterial Metal Homeostasis. Chemical reviews 109: 4644–4681.

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic acids research .

van der Meij A, Worsley SF, Hutchings MI, van Wezel GP. 2017. Chemical ecology of antibiotic production by actinomycetes. FEMS microbiology reviews 41: 392–416.

Nasioudis D, Witkin SS. 2015. Neutrophil gelatinase-associated lipocalin and innate immune responses to bacterial infections. Medical Microbiology and Immunology 204: 471–479.

Oliynyk M, Samborskyy M, Lester JB, Mironenko T, Scott N, Dickens S, Haydock SF, Leadlay PF. 2007. Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. Nature Biotechnology 25: 447–453.

Palmer LD, Skaar EP. 2016. Transition Metals and Virulence in Bacteria. Annual Review of Genetics 50: 67–91.

Perera L, Freudenthal BD, Beard WA, Pedersen LG, Wilson SH. 2017. Revealing the role of the product metal in DNA polymerase β catalysis. Nucleic Acids Research 45: 2736–2745.

Perham M, Stagg L, Wittung-Stafshede P. 2007. Macromolecular crowding increases structural content of folded proteins. FEBS Letters 581: 5065–5069.

Putnam CD, Hammel M, Hura GL, Tainer JA. 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Quarterly Reviews of Biophysics 40: 191–285.

Rambo R. 2019a. Bioisis: tutorials. `http://www.bioisis.net/tutorial/7`.

Rambo R. 2019b. ScÅtter. `http://www.bioisis.net/tutorial/9`.

Razin SV, Borunova VV, Maksimenko OG, Kantidze OL. 2012. Cys2his2 zinc finger protein family: classification, functions, and major members. Biochemistry. Biokhimiia 77: 217–226.

Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. 2016. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Research 44: W424–W429.

Schopf JW. 2012. The Fossil Record of Cyanobacteria. Whitton BA, editor, Ecology of Cyanobacteria II: Their Diversity in Space and Time, Springer Netherlands, Dordrecht, 15–36.

Seiple IB, Zhang Z, Jakubec P, Langlois-Mercier A, Wright PM, Hog DT, Yabu K, Allu SR, Fukuzaki T, Carlsen PN, Kitamura Y, Zhou X, Condakes ML, Szczypiński FT, Green WD, Myers AG. 2016. A platform for the discovery of new macrolide antibiotics. Nature 533: 338–345.

Stensmyr M, Dweck H, Farhan A, Ibba I, Strutz A, Mukunda L, Linz J, Grabe V, Steck K, Lavista-Llanos S, Wicher D, Sachse S, Knaden M, Becher P, Seki Y, Hansson B. 2012. A Conserved Dedicated Olfactory Circuit for Detecting Harmful Microbes in Drosophila. Cell 151: 1345–1357.

Stigall AL. 2017. Ordovician oxygen and biodiversity. Nature Geoscience 10: 887.

Szpunar J. 2005. Advances in analytical methodology for bioinorganic speciation analysis: metallomics, metalloproteomics and heteroatom-tagged proteomics and metabolomics. Analyst 130: 442–465.

Tottey S, Waldron KJ, Firbank SJ, Reale B, Bessant C, Sato K, Cheek TR, Gray J, Banfield MJ, Dennison C, Robinson NJ. 2008. Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. Nature 455: 1138–1142.

Traoré DAK, Ghazouani AE, Ilango S, Dupuy J, Jacquamet L, Ferrer JL, Caux□Thang C, Duarte V, Latour JM. 2006. Crystal structure of the apo-PerR-Zn protein from Bacillus subtilis. Molecular Microbiology 61: 1211–1219.

Waksman SA. 1919. CULTURAL STUDIES OF SPECIES OF ACTINOMYCES. Soil Science 8: 71.

Waldron KJ, Rutherford JC, Ford D, Robinson NJ. 2009. Metalloproteins and metal sensing. Nature 460: 823–830.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics (Oxford, England) 25: 1189–1191.

Zhao K, Li J, Zhang X, Chen Q, Liu M, Ao X, Gu Y, Liao D, Xu K, Ma M, Yu X, Xiang Q, Chen J, Zhang X, Penttinen P. 2018. Actinobacteria associated with Glycyrrhiza inflata Bat. are diverse and have plant growth promoting and antimicrobial activity. Scientific Reports 8: 13661.