Investigating the role of the phage protein paratox in streptococcal natural competence and phage biology

By:

Nicole Rutbeek

A thesis submitted to the Faculty of Graduate Studies of the University of Manitoba

in partial fulfillment of the requirements for the degree of:

Master of Science

Department of Microbiology

University of Manitoba

Winnipeg, MB, Canada

ABSTRACT

Streptococcus pyogenes is responsible for a wide variety of human diseases ranging from minor skin infections and pharyngitis to more serious complications such as rheumatic fever and flesh-eating disease. The varying degrees of pathogenicity observed between different strains of *S. pyogenes* are largely due to the expression of prophage encoded toxins and virulence factors. Furthermore, these toxin and virulence genes are actively spread between different strains and species of *Streptococcus* by horizontal gene transfer. One important mechanism of horizontal gene transfer in *Streptococcus* is natural transformation, which involves the uptake of exogenous DNA. This processes is regulated by either the ComRS or ComCDE quorum sensing pathways. *S. pyogenes* contains the ComRS quorum sensing system, which is negatively regulated by a phage protein paratox (Prx). Prx is a small conserved protein that is encoded at the 3' end of the prophage, directly adjacent to a toxin or virulence gene. However, little was known about the global role of Prx in the biology of the phage or its detailed mechanism of natural competence inhibition. An X-ray crystal structure reveals that Prx binds directly to the DNA binding domain of the transcription factor ComR. Additional biochemical assays also show that Prx functions by directly preventing DNA binding without affecting signal peptide (XIP) recognition. Thus, Prx is able to interact with ComR in either the apo (inactive) conformation and the XIP bound (active) conformation. This mechanism of quorum sensing inhibition, while structurally unique, is analogous to the quorum sensing repression observed by a *Pseudomonas aeruginosa* phage protein, Aqs1. While the inhibition of natural competence expression could be beneficial to prophage, it is also hypothesized that Prx may play an additional biological role. To address this hypothesis, pull-down assays with *S. pyogenes* lysates were performed and revealed a number of additional putative Prx binding partners worth further investigation.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor Dr. Gerd Prehna. I am beyond grateful for all of his valuable advice and genuine support over the last few years. I have gained an incredible amount knowledge, and my academic achievements can be directly attributed to his mentorship. I simply cannot express how fortunate I have been with the opportunity to work on an intriguing project in such a wonderful research environment.

I am also greatly appreciative of Dr. Deborah Court and Dr. Mazdak Khajehpour for their time spent as members on my advisory committee and their thoughtful advice and support. Also, I am especially thankful for the opportunity to collaborate with Dr. Khajehpour, whose mentorship and insight has been incredibly valuable.

There are a number of laboratories within the departments of Microbiology and Chemistry I would like to thank for their assistance with, and use of equipment, which was essential to this project. This includes (but is not limited to) the laboratories of Dr. Brian Mark, Dr. Karen Brassinga, Dr. Sean McKenna, and Dr. Joe O'Neil. I would also like to thank Dr. Trushar Patel at the University of Lethbridge, AB, for the coordination of SEC-SAXS data collection at the Diamond light source (B21 beamline) and for helping review the processed data. Also, several bacterial strains and expression constructs used in this study were generously gifted by the Federle lab from the University of Illinois at Chicago, IL.

Finally, I would like to thank all of my lab-mates (past and present), as well as the staff and students within the department of Microbiology, who have all contributed to an exceptionally fun and collaborative research environment.

TABLE OF CONTENTS

PERMISSIONS

Many of the figures, tables, and data within this thesis have been included and adapted from the following peer-reviewed article:

> Rutbeek, N.R., Rezasoltani, H., Patel, T.R., Khajehpour, M., & Prehna, G. Molecular mechanism of quorum sensing inhibition in Streptococcus by the phage protein paratox. *J. Biol. Chem.* **279**, e100992 (2021).

Only the data for experiments I performed are included in this thesis.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACN	Acetonitrile
β-ME	2-mercaptoethanol
CD	Circular dichroism
CDM	Chemically defined media
DBD	DNA binding domain
DMAB	Dimethylamine-borane
DMSO	Dimethyl sulfoxide
DOL	Degree of labelling
DTT	Dithiothreitol
EMSA	Electromobility shift assay
FAM	Fluorescein amidite
FITC	Fluorescein-5-isothiocyanate
FPLC	Fast protein liquid chromatography
FRET	Fluorescence resonance energy transfer
GAS	Group A *Streptococcus*
HGT	Horizontal gene transfer
IAA	Iodoacetamide
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LB	Lysogeny broth
MS	Mass spectrometry
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethylene glycol
RMSD	Root-mean-square deviation
SAD	Single wavelength anomalous diffraction
SAXS	Small angle X-ray scattering
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
TFA	Trifluoroacetic acid
THY	Todd Hewitt broth (supplemented with 0.2% yeast extract)
TMT	Tandem-mass-tag
TPR	Tetratricopeptide repeat
TSA	Trypticase soy agar
WT	Wild-type
XIP	SigX inducing peptide

1 INTRODUCTION

1.1 *Streptococcus pyogenes*

1.1.1 Characteristic features and pathogenicity

Streptococcus pyogenes is a Gram-positive, lactic acid bacterium, belonging to the Streptococcaceae family. *S. pyogenes* is commonly known by its Lancefield classification[1] as Group A Streptococcus (GAS), for which its major antigenic determinant is the presence of polyrhamnose chains with β-linked N-acetylglucosamine side chains extending from the cell wall[2]. Another characteristic feature of *S. pyogenes* (along with a few other pathogenic species of *Streptococcus*) is its ability to perform beta-hemolysis, the complete breakdown of red blood cells. *S. pyogenes* is a common human pathogen with a broad range of pathogenicity. It is most commonly known to cause minor illnesses such as pharyngitis (Strep throat) and impetigo[3]. However, infection with *S. pyogenes* can sometimes lead to more serious complications such as scarlet fever and less commonly, rheumatic fever[4]. Although rare, *S. pyogenes* can also be responsible for life-threatening diseases such as necrotizing fasciitis, sepsis, and streptococcal toxic shock syndrome[4]. Most strains of *S. pyogenes* carry one or more virulence factors, including the anti-phagocytic M protein[5] as well as a variety of phage encoded virulence genes[6], all of which contribute to its success, as well as diversity as a pathogen.

1.1.2 Global relevance & economic burden

While serious complications due to *S. pyogenes* are rare in North America, it has a substantial burden on the healthcare system. *S. pyogenes* is one of the major cause of pharyngitis[4], which is among most common reasons for primary health care visit[7]. In the United States, pharyngitis due to GAS alone was predicted to have an economic burden of $224 to $539 million annually[8]. The global impact of *S. pyogenes* is considerably more significant. In 2005 the World

Health Organization (WHO) estimated there were over 18 million existing cases of serious GAS related illnesses, and that *S. pyogenes* was responsible for over 500,000 deaths per year[3]. Furthermore, the increased as well as inappropriate use of antimicrobials has led to the growing concern of antimicrobial resistance. Currently, the Centers for Disease Control and Prevention (CDC) lists Erythromycin resistant Group A Streptococcus as a "concerning threat" with an estimated 5400 invasive infections and 450 deaths caused by Erythromycin resistant GAS in 2017; over a 3 fold increase compared to 2013[9].

1.1.3 Horizontal gene transfer in *Streptococcus*

Horizontal gene transfer (HGT) plays a major role in spreading virulence traits from one strain of *S. pyogenes* to another, resulting in the emergence of new, and pathogenic strains[10,11]. Not only does HGT occur between different strains of *S. pyogenes*, it can also occur between different species of *Streptococcus*[12,13]. The two primary methods of horizontal gene transfer in *Streptococcus* include phage transduction[13] and natural transformation[14]. In *S. pyogenes* phage transduction has been shown to occur through generalized transduction, involving infection with a lytic bacteriophage[15]. However more importantly, *S. pyogenes* contains a number of lysogenic prophages within its genome with the ability to excise and infect a new host, allowing for the direct transfer of phage encoded toxins from one strain to another[16]. Another important mechanism of HGT in *Streptococcus* is natural transformation, which can occur once the bacterium becomes naturally competent allowing for the uptake of DNA from its environment[17]. It is important to note that although *S. pyogenes* contains all of the genes required for natural competence expression, natural transformation is difficult to observe[18]. This suggests natural transformation is tightly regulated in *S. pyogenes*[19].

1.2 Phage transduction in *S. pyogenes*

1.2.1 Generalized transduction

Generalized transduction is the transfer of non-specific genetic material from one bacterium to another through infection with a bacteriophage[20]. This occurs when a bacteriophage incorrectly packages bacterial genetic material in its viral particle rather than its own. Most bacteriophage capable of mediating generalized transduction package their DNA through the recognition of *pac* sites, followed by headful packaging[21]. In the event of generalized transduction, the phage terminase recognizes and cleaves pseudo-*pac* sites in the bacterial chromosome (or a plasmid) resulting in the packaging of bacterial genetic material[21]. The phage then goes on to infect a new host, and the bacterial DNA it carries has the potential to undergo homologous recombination and be incorporated in the genome of the new host cell. Although generalized transduction is commonly mediated by temperate bacteriophages, it can be facilitated by virulent bacteriophages as well. In *S. pyogenes* generalized transduction has been shown to occur with a number of bacteriophages, including the lysogenic T12 phage[22], and lytic A25 phage[15].

1.2.2 Common bacteriophages of *S. pyogenes*

Several lytic bacteriophages are known to infect *S. pyogenes*, with the best described example being the A25 phage[23]. The A25 phage belongs to the Siphoviridae family, of which identifying characteristics include linear dsDNA genomes and long non-contractile tails[13]. Recent sequencing of this bacteriophage genome revealed that it shares many similarities with the streptococcal prophages and is likely to have evolved from a temperate bacteriophage, having lost the ability to integrate into the bacterial genome[23].

Temperate bacteriophages can be found in every known strain of *S. pyogenes*[24]. The temperate phages known to infect *S. pyogenes* include T12, SF370, and similar phages, all

belonging to the Siphoviridae family[13]. These bacteriophages are found within the genomes of *S. pyogenes* as stable genetic elements, or prophages. However, given the right conditions they have the ability to excise from the genome, package their DNA in viral particles, and lyse from the cell to go on to infect new bacteria[16]. Stable phage-like elements that have lost the functional and structural genes required for replication can also be found in the Streptococcal genome[25]. These phage and phage-like elements are highly prevalent in *S. pyogenes*, with each strain carrying anywhere from 1-8 of these prophages within its genome[13]. Bacteriophage DNA integrates into the host genome at site-specific *attB* sequences in the bacterial chromosome through homologous recombination with the bacteriophage *attP* sites[26]. The *attB* sites can be found in a variety of locations in the genome, however are most often within coding regions[13]. Prophages have been found integrated in a variety of genes, with tRNA genes being the most common which account for nearly a third of targeted genes[13]. The genetic organization of these prophages is highly similar to the well-studied lambda phage[27]. These prophages consist of approximately 40 or more genes, including, from 5' to 3', genes for: a site-specific integrase, lysogeny regulation, DNA replication and modification, structural and capsid proteins, lysis, and finally (if present) toxins and virulence[13]. Despite the similarities to other well characterized temperate phages, many streptococcal prophage genes remain unannotated and uncharacterized.

1.2.3 Phage encoded toxin and virulence genes

Most temperate phages in *S. pyogenes* carry one or more toxin or virulence genes located at the 5' end of prophage[13]. These genes have been shown to aid in GAS virulence and can include superantigens/exotoxins, phospholipases, and DNases[28,29]. The exotoxins expressed by prophages (SSA, SpeA, SpeC, SpeK, and a variety of others) are secreted by the bacterium resulting in over-activation of the immune system[28,30]. The superantigens bind to the major histocompatibility

complex class II and T cell receptors simultaneously, resulting in the release of cytokines[31]. The prophages can also encode phospholipases (Sla), which aid in virulence[32] and DNases (Sdn, Spd), which allow for the escape of neutrophil traps[29]. A few temperate phage have also been shown to carry antimicrobial resistance genes, including *mefA* (erythromycin resistance) and *tetO* (tetracycline resistance)[33].

Although a variety of virulence genes are encoded by streptococcal prophages, the expression and regulation of these genes is not well understood[34]. While prophage genes play a significant role in streptococcal pathogenicity, it is possible that they are downregulated in a complex evolutionary balance between streptococcal pathogenicity and persistence in the human host.

1.3 Natural Competence

1.3.1 Natural Competence regulation in *Streptococcus*

Natural transformation is a mechanism of HGT where bacterial cells take up exogenous DNA, which is then incorporated into the genome through homologous recombination[14]. In order for this process to take place, the cells must first become naturally competent. This involves the production of the required machinery for the uptake of dsDNA, processing dsDNA to ssDNA, and shielding the product from host degradation[14,35]. In *Streptococcus,* the expression of these required genes is regulated by either the ComRS[36] or ComCDE[37] quorum sensing systems. The ComCDE pathway utilized by *Streptococcus pneumoniae,* involves the activation of the ComD transmembrane receptor by the ComC signal peptide (CSP), which results in the phosphorylation of the transcription factor ComE[38]. The ComRS pathway, found in *S. pyogenes*, differs in that the ComS signal peptide (XIP) enters the cell and directly activates ComR, which is both a signal transducer and transcription factor[39].

1.3.2 ComRS quorum sensing pathway

The ComRS quorum sensing pathway regulates natural competence gene expression for a number species of *Streptococcus*, including *S. pyogenes, S. agalactiae, S. thermophilus, S. suis, S. bovis,* and *S. mutans*[12]. The ComRS pathway, summarized in Figure 1, begins with the basal expression of the small peptide ComS which is secreted from the cell. Once outside of the cell, ComS is cleaved into the active 7-8 amino acid active pheromone XIP, which is then able to re-enter the cell[40]. After a threshold concentration of XIP has been reached within the cell, which occurs when the cells are in a high-density population, the XIP binds to ComR, the key transcription factor of the pathway[36]. Binding to XIP causes ComR to undergo a conformational change, allowing ComR to dimerize with another activated monomer and bind to DNA[41]. ComR recognizes the promoters for *comS*, creating a positive feedback loop for the expression of more signal peptide, and *sigX*, resulting in the transcription of the alternative sigma factor X (SigX)[39]. SigX when bound to RNA polymerase allows for the recognition of CIN-box promoters that contain the conserved sequence: TACGAATA[39,42]. Each of the late genes required for natural competence contains a CIN-box in its promoter region and is thus subsequently transcriptionally regulated by SigX[39].

1.3.3 ComR and related proteins

ComR is both the signal transducer and the key transcription factor of the ComRS quorum sensing pathway. It is an approximately 35 kDa protein consisting of two distinct domains that are joined by a flexible linker[12]. The large C-terminal tetratricopeptide repeat (TPR) domain contains a binding pocket responsible for recognizing the signal peptide (XIP)[12]. The TPR binding pocket has both conserved and variable surfaces to help discriminate between different secreted XIP signals[12]. The smaller N-terminal DNA binding domain (DBD) consists of a helix-turn-helix

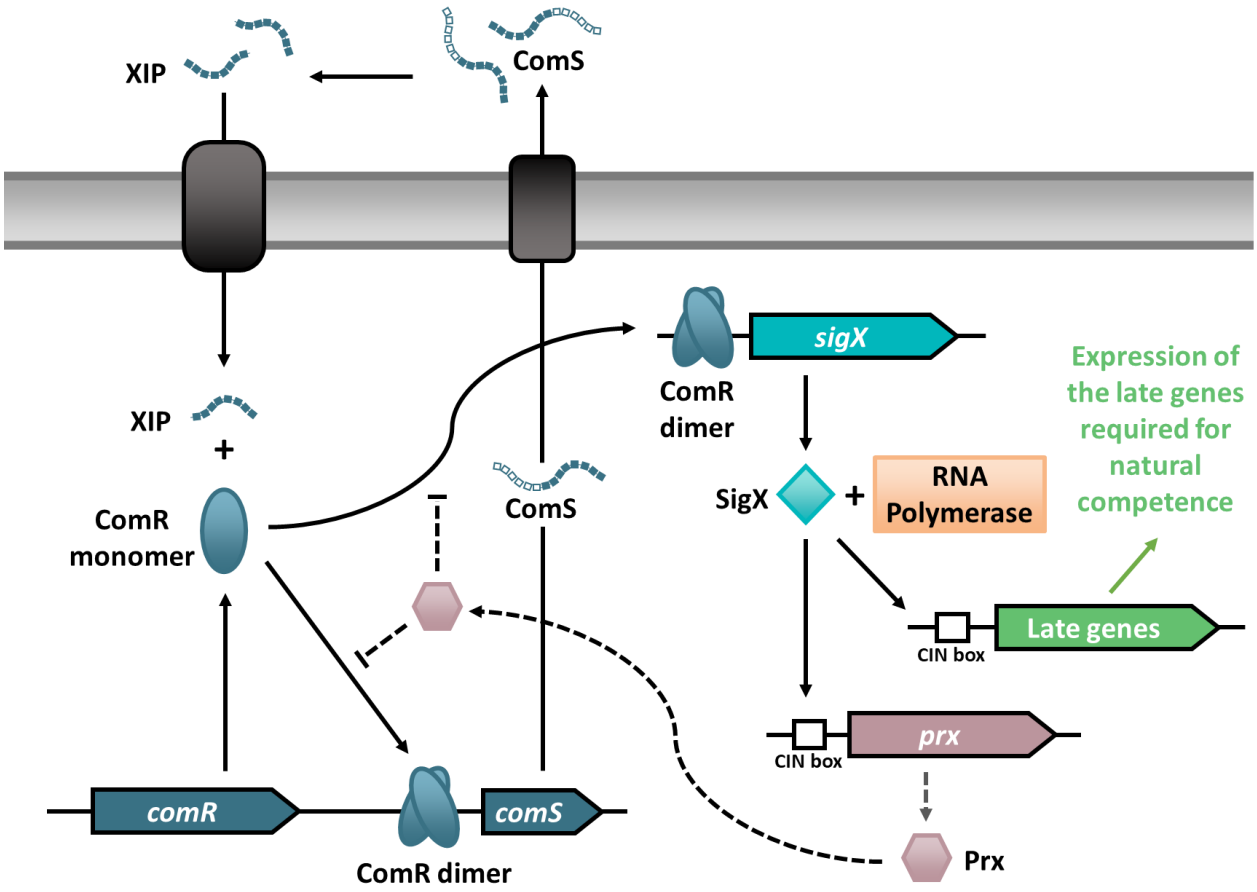

Figure 1. **Schematic diagram of the ComRS quorum sensing pathway.** A signal peptide (ComS) is secreted and cleaved once it leaves the cell[40]. Upon re-entry into the cell, XIP is able to bind and activate the transcription factor ComR[36]. This results in the expression of more ComS (XIP) and expression of the alternative sigma factor SigX[39]. SigX recognizes CIN-box promoter sequences and when bound to RNA polymerase allows for the expression of the late genes required for natural competence[39]. Expression of the phage protein Prx is also under the control of a CIN-box promoter[19]. Prx acts as a negative regulator of the pathway by preventing ComR from binding the *comS* and *sigX* promoters[19].

domain, with conserved arginine residues shown to be essential for recognizing DNA[41]. ComR belongs to the RRNPP protein family, and more specifically the Rgg sub-group of transcription factors, of which directly regulate virulence in *Streptococcus*[43]. Similar to ComR, Rgg proteins contain DBD and TPR domains, and DNA binding is regulated by a signal peptide (SHP)[43,44]. However, Rgg proteins are dimers in solution with or without the signal peptide bound whereas apo-ComR is a monomer in solution. Once the XIP binds ComR, ComR undergoes a large conformational change that results in the release of the DBD from the TPR domain, allowing ComR to dimerize and bind DNA[41].

The ComRS pathway is used by multiple different species of *Streptococcus*, with each species producing its own signal peptide. As a result there is a significant amount of variation between the different ComR proteins as well as their ability to recognize other/non-self XIP molecules[12]. ComR proteins are grouped into three distinct types based on the sequence of the XIP they recognize[12]. Type I ComR proteins, which can be found in *S. thermophilus*, recognize a hydrophobic XIP lacking tryptophan residues, while type II proteins, utilized by *S. pyogenes* and *S. mutans*, bind XIPs with a double tryptophan (WW) motif. Finally, type III proteins, found in *S. suis*, recognize XIP with a separated tryptophan (WXXW) motif[12]. ComR proteins, both within and between the different types only share approximately 30-50% amino acid sequence identity, with the largest amount of variability in the TPR domain.

1.3.4 Non-competence related genes expressed by SigX

SigX is an alternative sigma factor, which once bound to RNA polymerase allows for the expression of the late genes required for natural competence by specifically recognizing promoters with CIN-box sequences (TACGAATA)[39]. Previous microarray assays have shown that the ComRS pathway upregulates the expression of a number of genes, many of which were previously

thought to be unrelated to natural competence[39]. Of particular interest is a bacteriophage gene, *prx*, that contains a CIN-box in its promoter region[19].

1.4 Paratox

1.4.1 Paratox acts as a negative regulator of natural competence

Paratox (Prx) is an 8 kDa, highly conserved protein encoded by the prophages of many pathogenic species of *Streptococcus*[19]. Not only is Prx expression induced by the ComRS quorum sensing system, it also inhibits the pathway. Prx was shown to bind the key transcription factor ComR *in vitro* with nanomolar affinity, which prevents ComR from binding the *comS* and *sigX* promoters[19]. Prx was also shown to reduce natural competence late gene expression *in vivo*. Mutants lacking *prx* showed increased SSB (single stranded binding protein) transcription, and *prx* introduced (on a vector) into to a species of *Streptococcus* not containing any *prx* genes (*S. mutans*), leads to a decrease in SSB expression[19]. This shows that Prx acts as a negative regulator of the ComRS pathway and could provide an explanation for why natural transformation is difficult to observe in *S. pyogenes*, despite expressing the genes required for natural competence[18,19,39].

1.4.2 Linkage of *prx* to phage encoded toxin and virulence genes

While little is known about the role of Prx, besides its relation to the ComRS pathway, one of its most interesting features is the location of its gene within the prophage. Prx is always encoded at the 3' end of the prophage, and adjacent to a phage encoded toxin or virulence gene[45]. Not only is Prx found adjacent to a toxin gene, it is also genetically linked to that toxin. When aligning *prx* sequences, it will group based on the type of toxin or virulence gene next to it[45]. A simple gene tree of the *prx* sequences from the prophages within two different strains of *S. pyogenes* is shown in Figure 2, highlighting that the neighboring toxin is a better indicator of *prx* homology than the

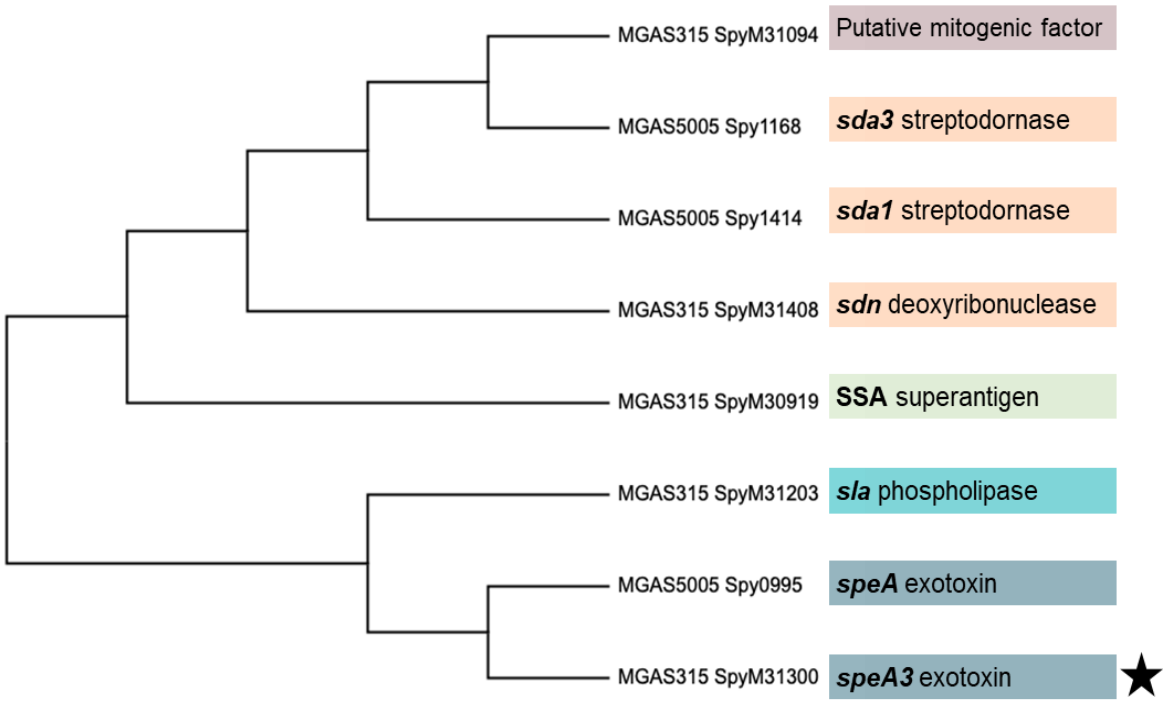

Figure 2. Simple neighbour-joining gene tree including all the Prx amino acid sequences from *S. pyogenes* MGAS315 and *S. pyogenes* MGAS5005. Each sequence is annotated by strain, followed by gene identifier. The toxin and virulence gene directly adjacent to each *prx* gene is displayed on the right. The version of Prx used in this study (SpyM3_1300, from MGAS315) is starred.

bacterial strain. The linkage of these two genes indicates they are likely to remain together in the event of homologous recombination[45]. Furthermore, Prx is highly conserved, sharing over 60% and often 80% amino acid sequence identity between different species of *Streptococcus*[19]. The genes directly 5' of toxin/virulence genes (hyaluronidases, holins, and lysins) are also highly conserved. Therefore, it has been hypothesized that these genes along with *prx,* could serve as frequent sites for homologous recombination and allow for the spread of the toxin genes between different phages[45].

1.4.3 Homology and similarities of Prx to other phage proteins

Prx is a highly conserved protein; however no known homologues have been found in any other genus. Even *Staphylococcus aureus,* which shares a few highly similar exotoxins with *S. pyogenes,* has no obvious Prx homologs. Prx also has a novel fold[19], with no known significant structural homologs. A few distant structural homologs can be found, all with insignificant Z-scores (all are below 3) as indicated by the Dali server[46]. These hits mainly consist of significantly larger proteins for which Prx partially aligns with a small portion. One potential distant homolog is a small phage protein GpW, which shares a 20% amino acid sequence identity, and some structural similarities[19]. GpW is a structural lambda phage protein responsible for joining the head tail regions of the viral particle[47].

1.5 Objectives

1.5.1 OBJECTIVE 1: Characterize the mechanism of natural competence inhibition by paratox

Since Prx has a novel protein fold[19] and ComR undergoes significant conformational changes upon activation[12,41], it is hypothesized that Prx employs a novel mechanism of protein-protein binding and inhibition. Fully characterizing this mechanism of natural competence inhibition will allow for the understanding of a potentially novel mechanism of protein-protein

binding and inhibition, as well as provide additional insight into the evolutionary role of the bacteriophage protein paratox.

1.5.2 OBJECTIVE 2: Search for additional roles of paratox in the biology of the phage

There are numerous examples of bacteriophage proteins that have adapted to perform multiple functions, ultimately economizing their restricted genomes[48,49]. In addition, because Prx shares some structural homology to the lambda phage GpW[19,47] and is genetically linked to a toxin gene within the prophage[19,45], we hypothesize that Prx plays an additional role for the survival of the bacteriophage.

2 MATERIALS & METHODS

2.1 Bacterial strains & growth

All bacterial strains used are displayed in Table A1. For protein expression and cloning *Escherichia coli* BL21-Gold (DE3) cells were used. Overnight cultures of *E. coli* were grown directly from freezer stocks stored at -80 °C (20% glycerol) in Lysogeny Broth (LB) with 100 µg/mL ampicillin at 37 °C with shaking, unless stated other-wise.

Streptococcus pyogenes strains MGAS315, MGAS5005, and MGAS5005Δ*prx3* were grown from the freezer stocks on Trypticase soy agar II (TSA) plates supplemented with 5% sheep's blood (ThermoFisher) to ensure no contaminates were present. Isolated colonies were then used to inoculate overnight cultures in either Todd Hewitt broth supplemented with 0.2% yeast extract (THY), or chemically defined media (CDM)[39,50]. *Streptococcus mutans* UA159 was grown directly from a freezer stock (20% glycerol) in THY broth. All *Streptococcus* strains were grown in 7.5 mL of media in 8 mL sealed screw cap tubes, at 37 °C.

Because MGAS5005 and MGAS5005Δ*prx3* were to be used for proteomics, primers were designed as an extra step of verification. This was done to ensure they were the correct strains. The primers (listed in Table A2) were designed for PCR either around the *prx3* gene, which would result in an approximately 800 base pair product in the WT strain and 1500 bp product in the Δ*prx3* strain (due to the presence of kan^R in its place), or including the *prx3* gene, which would result in a 500 bp product in the WT and no product in the Δ*prx3* strain. These PCRs were performed using Taq polymerase.

2.2 Plasmids and generation of expression constructs

2.2.1 Expression constructs

The protein expression constructs used in this study include Prx MGAS315[19], ComR *mutans*[19], and ComR *suis*[12], as well as ComR MGAS5005, ComR MGAS315, and Rgg3 *S. pyogenes*, which were generously gifted by the Federle lab at the University of Illinois at Chicago. The expression constructs for ComR *thermophilus* in pET21a and Prx MGAS315 in pGEX6P1, were ordered codon optimized, from Genscript. All protein constructs, vectors, along with all relevant information are listed in Table A3.

2.2.2 Generation of new protein variants and constructs

Each Prx variant and the ComR DBD construct were generated by Q5 mutagenesis (New England Biolabs), using either the Prx or ComR *mutans* construct as a template. The ComR *mutans* DBD construct was generated through Q5 mutagenesis by the substitution to a stop codon after residue D66. For the ComR TPR construct, an insert containing the TPR domain of ComR *mutans* (residues S74 to T304), was cloned into the NdeI and BamHI sites of an empty pET15b vector. The primers for each new expression construct are listed in Table A4. All new protein constructs were verified by induction trials (described in section 2.3.1) as well as Sanger sequencing (Sickkids - The Centre for Applied Genomics) of the multiple cloning site region of the plasmid.

2.3 Protein expression and purification

2.3.1 Induction trials

For each new protein expression construct, small-scale induction trails were performed prior to large scale protein expression. *E. coli* BL21-Gold (DE3) cells containing the protein expression plasmid of interest were grown overnight in LB + 100 µg/mL ampicillin at 37 °C. The next morning 3 mL of cells were grown at a 1 in 100 dilution in LB + 100 µg/mL ampicillin at 37

°C to an optical density of approximately 0.6 at 600 nm. At this point half of the cells (1.5 mL) were transferred to a new test tube and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Both induced and non-induced cells were allowed to grow at 20 °C overnight. The following morning the cells were collected (centrifugation at 4000 rpm for 10 min, 4°C), and re-suspended in 300 µL of a lysis buffer anticipated to be appropriate for the protein of interest (usually 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 25 mM imidazole) and kept on ice from this step forward. The cells were then lysed by sonication (5 seconds each), followed by centrifugation at 18000 rpm, 4°C, for 10 minutes. The lysates were transferred to new 1.5 mL centrifuge tubes and the pellets were re-suspended in the same volume (300 µL) of lysis buffer. Finally, the same volume of each non-induced pellet, non-induced lysate, induced pellet, and induced lysate (for each protein expression construct of interest) were run on a 15% SDS-PAGE gel. The presence of a distinct band correlating to the expected protein size in the induced samples, more-so than the non-induced indicates the expression plasmid is functional and the protein is inducible. The presence of a more distinct band in the induced lysate rather than the induced pellet, also indicates the protein is soluble in the given conditions.

2.3.2 Large-scale protein expression

Each protein in this study was expressed and purified as previously described[19]. To summarize, *E. coli* BL21-Gold (DE3) cells were grown in LB + 100 µg/mL ampicillin at 37 °C to an optical density of approximately 0.6 to 0.8 at 600 nm. At this point the temperature was reduced to 20 °C and IPTG was added at a concentration of 1 mM to induce protein expression. The cells were allowed to grow overnight for approximately 20 hours, and were collected the following morning by centrifugation at 4200 rpm, at 4°C. The cell pellets were then re-suspended in the appropriate wash/lysis buffer (typically 50 mM Tris pH 7.5, 500 mM NaCl, 25 mM imidazole),

and stored at -80 °C until lysis. On average, 3 L of cells would be re-suspended in 100 to 150 mL of buffer.

2.3.3 Se-MET labelled protein expression

In order to purify ComR *mutans* with Se-Met replacing Met, for the purpose of collecting an anomalous diffraction dataset, Bl21-Gold (DE3) cells containing the expression plasmid were grown in a defined media with Se-Met added. First a normal overnight culture containing the ComR *mutans* expression plasmid was grown in LB + 100 µg/mL ampicillin. The following morning the cells were centrifuged at 2000 rpm and 4 °C for 30 minutes. The cells were then re-suspended in M9 media (11.28 g/L of M9 salts), spun down again, and finally re-suspended in their original volume with M9 media. While the cells were being washed, the following supplements were sterile filtered and added (per 1 L) to autoclaved M9 medium: 1 mL 1 M $MgSO_4$, 0.1 mL 1 M $CaCl_2$, 1 mL 0.01 M $FeCl_3$, 1 mL 1 mg/mL thiamine, 25 mL 40% glucose, and 1 mL 100 mg/mL ampicillin. Once washed, the cells were added at a 1:100 dilution (10 mL per 1 L media), and allowed to grow at 37 °C until an OD_{600nm} of approximately 0.8 was reached. At this point the following amino acids (dissolved in water) were added (per 1 L): 0.05 g Se-Met, 0.1 g Lys, 0.1 g Thr, 0.1 g Phe, 0.05 g Leu, 0.05 g Val and 0.05 g Pro. Each of the amino acids were dissolved together in water, with the exception of Se-Met which was dissolved and added separately. The temperature was reduced to 20 °C, and the cells were allowed to grow for an additional 30 minutes, after which IPTG was added to a concentration of 0.6 mM. The cells were collected after 16 hours.

2.3.4 Cell lysis

For each of the following steps, cells and or lysates were kept on ice or at 4 °C when possible. Prior to lysis, cell pellets were thawed in icy water, and 1 mM phenylmethylsulfonyl

fluoride (PMSF), 10 mM $MgCl_2$, and a small amount of DNase I (approximately 1-5 mg) was added. The cells were then lysed using an EmulsiFlex-C3 (Avestin). The soluble lysate, containing the protein of interest was separated from the insoluble material by centrifugation at 16000 rpm, 4 °C, for 30 minutes.

2.3.5 Purification of His-tagged proteins

The expression constructs for Prx, and each ComR protein all contained 6-His tags, and were purified by Nickel-affinity chromatography. For most protein purifications, a gravity column was loaded with 8 mL of Nickel NTA agarose resin (GoldBio). First the resin (stored in 20 % EtOH) was washed with water, charged (~10 mL of 200 mM $NiSO_4$), washed again with water, and equilibrated into the lysis/wash buffer of the protein (usually 50 mM Tris pH 7.5, 500 mM NaCl, 25 mM imidazole). Next the lysate was run over the column, and then washed with 200 mL of the same wash buffer. Finally, the protein was eluted with 50 mL elution buffer (50 mM Tris pH 7.5, 500 mM NaCl, 500 mM imidazole). The column was then washed with a small volume of water, ~10 mL of 200 mM EDTA pH 8.0, 50 mL of water, and equilibrated into 20% EtOH for storage. Samples were collected from each step (pellet, lysate, flow-through, wash, and elution) and run on a 15% SDS-PAGE gel.

Each of the ComR constructs contained cleavable His-tags, either with a thrombin protease site (ComR *mutans, suis,* MGAS5005, MGAS315) or a HRV-3C protease site (ComR *thermophilus*). When removing the His-tag, the protein was dialyzed in a minimum of 1L of digestion buffer (50 mM Tris pH 7.5, 500 mM NaCl), with either thrombin or HRV-3C added at an estimated ratio of 1:100 mg enzyme:protein, overnight at 4°C. Following digestion, the protein was centrifuged at 4200 rpm, 4 °C, to remove any precipitation, and re-run over the Nickel-NTA column with the flow-through containing the digested protein and the elution contain non-digested

protein (if remaining). Samples were always run on an SDS-PAGE gel to ensure complete digestion.

Each protein was purified in the listed buffers, with the exception of ComR MGAS315 and ComR MGAS5005, which had 10% glycerol added to each buffer.

2.3.6 Purification of GST-tagged protein

GST-Prx was purified using a gravity column with glutathione-sepharose resin (GoldBio). However, first the lysate was run over a Q-sepharose column (about 10 mL of resin in a 50 mL gravity column) as an initial cleaning step of the lysate to help preserve the glutathione resin. The column was first washed with water, then equilibrated with lysis buffer (50 mM Tris-HCl pH 7.5, 200 mM NaCl, 2 mM β-Me). The lysate was then run over the column and the flow-through was collected. The Q-sepharose resin was washed with a high salt/high pH buffer, regeneration buffer 1 (50 mM Tris-HCl pH 8.8, 1M NaCl), followed by washing with water, and equilibration into 20% EtOH for storage. The glutathione-sepharose column (8 mL resin added to a 50 mL gravity column), also stored in 20% EtOH, was washed with water and equilibrated with lysis buffer. Next the flow-through from the Q-sepharose column (containing the protein of interest) was run over the column, followed by 200 mL of wash buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 2 mM β-Me). The GST-tagged protein was then eluted from the column with 50 mL of elution buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM glutathione). Glutathione was added to the elution buffer immediately prior to use, at which point the pH adjusted to 7.5. Following elution, the glutathione-sepharose column was washed with a minimum of 50 mL of each regeneration buffer 1, regeneration buffer 2 (20 mM sodium acetate pH 4.5, 1 M NaCl), water, and then equilibrated into 20% EtOH. After approximately every 4 uses, the glutathione-sepharose column was washed

with 6 M guanidine hydrochloride after the second regeneration buffer, followed by multiple washes with water.

Following elution, the GST-tagged protein was dialyzed overnight, at 4 °C, in 2 L of digestion buffer (50 mM Tris-HCl pH 7.5, 200 mM NaCl, 2 mM β-Me) along with HRV-3C protease at an estimated ratio of 1:100 mg enzyme:protein. The following morning, the protein was centrifuged at 4200 rpm, 4°C, for 15 minutes, to remove precipitation. The digested protein was re-run over the glutathione-sepharose column with the flow-through (ideally) containing the protein of interest, and the elution containing GST, along with any undigested protein. While this step helped separate GST from the protein of interest, a significant amount of GST was always present in the flow-through as well, and could only be completely separated from the protein of interest by size exclusion chromatography.

2.3.7 Purification of SUMO-tagged protein

The Rgg3 expression construct contained an N terminal 6-His-SUMO tag. Rgg3 was purified identically to the other His-tagged proteins by nickel affinity chromatography. Following elution, the protein was digested by dialysis with an approximate ratio of 80:1 mg protein:SUMO protease in a low-salt digestion buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM β-Me, 10 % glycerol) overnight at 4°C.

2.3.8 Size exclusion chromatography

Following affinity chromatography and/or affinity tag removal, each protein was further purified by size exclusion chromatography (SEC) with complete buffer exchange into a gel filtration buffer (usually 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM β-Me). GST-Prx (after digestion) had a gel filtration buffer lacking β-Me as this allowed for excess GST to dimerize, making it easier to separate from Prx (as they are relatively close in size). Different buffers were

required (for the purposes of circular dichroism, primary amine labelling, etc.) and are listed under the relevant sections.

Size exclusion chromatography was performed at 4°C using an AKTA pure (GE Healthcare) and a HiLoad 16/600 superdex 75 gel filtration column (GE Healthcare). Before each use, the column (stored in 20% EtOH) was equilibrated with 2 column volumes (240 mL) of filtered RO water, followed by 2 column volumes of the filtered gel filtration buffer. Prior to running the protein over the AKTA, it was concentrated to a volume of approximately 2 mL using a centrifugal filter (Amicon) with a 3 kDa molecular weight cut off for Prx proteins, and a 10 kDa cut off for all other proteins. After running the protein over the size exclusion column, each fraction was run on an SDS-PAGE gel to ensure sample purity, and all pure fractions were collected and concentrated.

2.4 Protein-Protein binding assays

2.4.1 Size exclusion chromatography

For the purpose of protein-protein binding experiments by SEC, proteins were incubated for a minimum of 15 min on ice, at a molar ratio of 1:1.5 ComR:Prx, TPR:Prx, or Rgg3:Prx. For the DBD:Prx a 1.5:1 molar ratio was used as the DBD is slightly smaller than Prx. The complexes were run either over a HiLoad 16/600 superdex 75 column, or the smaller Superdex75 increase 10/300 column, using an AKTA pure (GE Healthcare). For each gel filtration binding assay, the standard gel filtration buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM β-Me) was used.

2.4.2 Pull-down experiments using purified protein

Pull down experiments were performed using Prx-6His, and various ComR orthologs with the His-tag removed. Each step was performed at 4 °C where possible. First Ni-NTA resin (GoldBio) was prepared by washing (incubating/rotating at for 5 minutes, followed by

centrifugation at 1500 *g* for 5 minutes, and removal of supernatant) with water 2 times, and then 2 times with buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM β-Me). Buffer was then added to bring the volume to a 50 % slurry of resin with buffer. A final volume of 20 µL of resin (40 µL of slurry) was then added to individual centrifuge tubes, and 0.36 mg of Prx-6His was added to each tube. Prx was allowed to incubate with the resin for 30 minutes, followed by two wash steps with buffer. Next 0.4 mg of each ComR ortholog was added to Prx saturated resin. The protein mixtures were allowed to incubate (rotating) for 1 hour, followed by two wash steps with Buffer. For each incubation and wash step the final volume was adjusted to 700 µL with buffer. Finally, all of the remaining buffer was carefully removed and 15 µL of 4× Laemmli sample buffer was added to each centrifuge tube. The samples were boiled at 93 °C for 5 min, followed by centrifugation at 21000 *g* for 3 minutes. Each sample (5 µL) was then run on a 15 % SDS-PAGE gel. No-Prx (ComR background) controls were run identically to and in parallel with the Prx pull down samples, with the only difference being the addition of buffer rather than Prx. Additionally, ComR-only controls containing the same amounts of ComR added to the pull-down samples were run on a SDS-PAGE gel to verify sample purity and provide a reference for the amount of protein added, versus pulled down by Prx.

2.4.3 Isothermal titration calorimetry

Prior to isothermal titration calorimetry (ITC) experiments, each protein was dialyzed overnight at 4 °C in the same buffer (gel filtration buffer). For each experiment, 300 µM Prx (or Prx D32A) was injected into 20 µM ComR *mutans* or DBD at a constant temperature of 25 °C. Controls of each Prx injected into buffer, as well as buffer injected into ComR were also run. The ITC was performed using a MicroCal iTC200 (GE Healthcare), and the data analysis was completed using the Origin software (Malvern) with all the results based on a one site model.

2.5 Protein crystallization, data collection, and refinement

2.5.1 Crystallization screening

ComR *mutans* was screened for crystallization in a variety of commercially available buffer conditions (Qiagen, Molecular dimesons, Jena Bioscience), by sitting drop vapor diffusion. Crystallization trays were set up using a Crystal Gryphon Robot (Art Robbins Instruments). All of these conditions were screened at a variety of protein concentrations ranging from approximately 5 to 25 mg/ml, both with and without the His-tag. A few protein crystals formed in the following buffer conditions: PACT G4 (0.2 M KSCN, 0.1 M Bis-Tris propane, 20% PEG 3350), JCSG++ E4 (0.1 MES pH 6.5, 1.6 M $MgSO_4$), and PACT G3 (0.2 M NaI, 0.1 M Bis-Tris propane pH 7.5, 20% PEG 3350), with 1:1 gel filtration buffer to crystallization buffer. All of these conditions were further optimized (same buffer condition with slight variations in salt, buffer, and/or precipitant concentration, as well as varying protein concentrations) in larger volumes, sometimes seeding with crystals from the same condition. ComR *mutans* (without the His-tag) crystallized readily in multiple conditions after optimization of the PACT G3 buffer condition.

The ComR *mutans*:Prx complex was screened for crystallization in all of the same conditions as ComR *mutans*. After months, no crystals formed so many additional conditions were screened including incubating crystal screens at 20 °C rather than 4 °C, *in situ* proteolysis (discussed in section 2.5.2), reductive methylation (section 2.5.3), as well as the addition of various ligands (Hampton) to buffer conditions which appeared to result in the formation of spherulites.

The DBD:Prx complex readily crystallized in a variety of conditions, so only a few trays were set up and crystals were picked directly from the screening trays. A few of the conditions used for downstream data collection include PEGs E9 (0.2 M NH_4Cl, 20 % PEG 3350), and PEGs G5 (0.2 M Potassium acetate, 20% PEG 3350).

2.5.2 *In situ* proteolysis

Since the ComR:Prx complex did not crystallize, the Proti-Ace kit (Hampton) was used in an attempt to remove a possible flexible/surface exposed region of the ComR:Prx complex that could be interfering with crystallization. First, 10 mg/mL ComR:Prx complex was incubated with 0.1 mg/mL of each of the enzymes from the kit (α-chymotrypsin, trypsin, elastase, papain, subtilisin, and endoproteinase Glu-C) for 45 min at 37 °C. Each sample was run on a 15 % SDS-PAGE gel in order to determine if any of the enzymes were appropriate to use. A condition where the protein of interest appears to be broken down into only a few slightly smaller yet clean bands, and not completely digested (multiple low molecular weight bands) was considered ideal. Both α-chymotrypsin and papain appeared to be appropriate enzymes, resulting in the formation of a few bands slightly smaller than ComR, without resulting in any precipitation. A variety of buffer conditions were screened for crystallization with each condition including ComR:Prx undigested, ComR:Prx with α-chymotrypsin, and ComR:Prx with papain. For each sample ComR:Prx was at a concentration of 10 mg/mL and the digested samples were incubated with 50 µg/mL enzyme for 1 hour at 4 °C prior to setting up the crystallization trays. After several months at 4 °C, crystals were observed in 1:1 gel filtration buffer and 0.1 M HEPES pH 7.0, 18 % PEG 12000.

2.5.3 Reductive methylation

Reductive methylation of ComR:Prx was also attempted as an additional effort to crystallize the complex, based on the rationale that the methylation of lysine residues could potentially change the protein surface enough to allow for new crystallization contacts. First the ComR:Prx complex was dialyzed at 4 °C overnight, in a buffer lacking free amines (20 mM HEPES pH 7.5, 100 mM NaCl). The complex, at a concentration of 5 mg/mL, was allowed to incubate for 2 hours, rotating, at 4 °C, protected from light along with 20 µL (per mL protein) of

1 M dimethylamine-borane (DMAB) and 40 µL (per mL protein) of 1 M formaldehyde. After 2 hours, the same amounts of each DMAB and formaldehyde were added again, and the protein was allowed to incubate for another 2 hours. Next, an additional 10 µL (per mL protein) of 1 M DMAB was added and the protein was allowed to incubate overnight (18 hours). The next morning the reaction was quenched with the addition of 500 µL of 1 M Tris-HCl (pH 7.5). The ComR:Prx complex was equilibrated back into the original buffer (gel filtration buffer) by SEC, followed by concentration, and screening for crystallization in a variety of available buffer conditions.

2.5.4 Data collection

Protein crystals were cryo-protected in crystallization buffer supplemented with PEG (PEG 3350 or 12000) to a concentration of over 35 %, flash frozen in liquid nitrogen and were tested for diffraction using the in house MicroMax-007 HF X-ray source and R-axis 4++ detector (Rigaku) or sent to the Canadian light source (CLS) or Advanced light source (ALS) for remote data collection.

For ComR *mutans* multiple datasets were collected, however the most successful dataset was of Se-Met ComR in 1:1 gel filtration buffer and PACT G3 crystallization buffer. This dataset was collected remotely at the CLS beamline CMCF-ID (081D-1) at a wavelength tuned to the K absorption edge for selenium (0.979590 nm).

For the DBD:Prx complex, a dataset resulting from in-situ proteolysis of the full ComR:Prx complex was collected at the CLS beamline CMCF-ID (081D-1). Data sets of the DBD:Prx complex were also collected using the in-house X-ray source as well as remotely at the ALS beamline 8.3.1.

2.5.5 Data processing and refinement

For each dataset, initial processing was performed using XDS[51], followed by the conversion of intensities into amplitudes and truncating the data in CCP4[52]. The phases for each dataset were solved by molecular replacement using Prx (PDBid: 6CKA)[19] and the DBD of ComR *suis* (PDBid: 5FD4)[12] for the DBD:Prx structure, and a combination of ComR *suis* (PDBid: 5FD4) and ComR *thermophilus* (PDBid: 5JUF)[41] for the ComR *mutans* structure. Although the Se-Met ComR *mutans* dataset was collected at a wavelength tuned to the absorption edge of selenium, there was not enough anomalous signal for SAD determination of phases and molecular replacement was used instead. All molecular replacement was performed using Phenix[53]. The models were further built using Coot[54] and refined with both Phenix[53] and Refmac5[55]. The final structural models for DBD:Prx were deposited in the Protein Data Bank (PDB) with PDB ids 7N10 and 7N1N.

2.6 Circular dichroism

Circular dichroism (CD) spectras were collected for each Prx variant, along with the wild-type protein in order to ensure that the introduced amino acid substitutions had no effect on the protein's secondary structure. Prior to CD, each of the concentrated Prx variants were dialyzed in 1.5 L of filtered CD Buffer (20 mM potassium phosphate pH 7.5, 100 mM NaF) overnight at 4 °C. The next morning, each protein was diluted to a measurable (by absorbance at 280 nm) stock concentration of exactly 2 mg/ml using the CD dialysis buffer. The CD spectra were collected using a Jasco J-810. Each protein was further diluted with CD buffer to a concentration of 0.2 mg/mL and loaded into a 0.05 cm cell. Three accumulations of each sample were collected and averaged from 260 nm to 190 nm. The output traces were converted from ellipticity units (mdeg) to $\Delta\varepsilon$ ($M^{-1}\cdot cm^{-1}$) using the following formula:

$$\Delta \varepsilon = \frac{\theta \cdot MWR}{C \cdot I \cdot 2398}$$

where θ = ellipticity (mdeg), *MWR* = the mean residue weight (molecular mass (Da) / number of amino acids -1), *C* = concentration (mg/mL), and *I* = pathlength (cm).

2.7 Small angle X-ray scattering

2.7.1 Data collection

Protein samples of ComR *mutans*, ComR *mutans*:Prx, and Prx were sent to the Diamond light source, B21 beamline, for SEC-SAXS data collection. Each protein sample was at a concentration of 9 mg/mL and run over a Shodex KW402.5-4F column equilibrated in gel filtration buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM β-Me). The samples were run at a flow rate of 0.160 mL/min with an exposure time of 3 seconds for each frame, as previously described[56].

2.7.2 Data analysis and processing

Data processing was predominantly performed using the ATSAS software package[57]. First CHROMIXS[58] was used for the sample peak selection and buffer subtraction, followed by PRIMUS[59], which was used for Guinier analysis and Kratky analysis. Next GNOM[60] was used to estimate the D_{max} and calculate the P(r) distributions.

The buffer subtracted datasets were also processed identically (Guinier and Kratky analysis) in ScÅtter (bioisis.net). By manually fitting a linear curve to the Porod region of the Porod-debye plot, the Porod exponent (P_X) could be determined. This provided a semi-quantifiable unit for comparing protein flexibility, and also allowed for the estimation of the Porod volume (V_P)[61].

Finally, the GNOM processed datasets were used in DENSS[62], for the estimation of low resolution electron density maps. For each protein sample, 20 unique density maps were

calculated, averaged, and refined back into the original datasets. The PDB structures of ComR and Prx, including ComR *S. suis* (PDBid: 5FD4), and the DBD:Prx complex, were manually aligned into the density maps using Chimera[63].

2.8 Fluorescence resonance energy transfer

2.8.1 Double labelled ComR

ComR *S. mutans* (no His tag) was double labeled with two separate fluorophores. First, ComR was singly labelled on the N terminus with an AlexaFuor 555 succinimidyl ester, after which ComR was labelled again with a C2 malemide (AlexaFluor 647). ComR has three cysteines, however only one of the cysteines (C199) is surface exposed, potentially allowing for single residue labelling.

For N terminal labelling of ComR with the succinimidyl ester, the recommended protocol (ThermoFisher) was loosely followed, with adjustments made to allow for less efficient labelling, therefore targeting only the N terminus and not the primary amines found on lysine residues. Labelling was attempted at two different conditions: A (pH 6.5) and B (pH 7.0). First the purified ComR *mutans* was dialyzed into a labelling buffer lacking primary amines (20 mM sodium phosphate pH 6.8, 100 mM NaCl). The following morning ComR was concentrated to a concentration 7.1 mg/mL (or 200 µM), allowing for buffer exchange multiple times into the same buffer, but with the appropriate pH (either 6.5 or 7.0). Once ready to label the protein, the fluorophore/label (1 mg) was dissolved in 100 µL DMSO resulting in a concentration of 8 mM. Next 80 µL protein (0.016 mmol) was incubated with 10 µL label (0.08 mmol) for 2 h, protected from light, shaking, and at room temperature. The reaction was stopped with the addition of 10 µL of 1.0 M Tris-HCl pH 7.5, and excess dye was removed using a Zeba dye removal spin column (ThermoFisher: A44296) and following the accompanying protocol. Finally, the degrees of

labelling (DOL) were calculated and recorded with a DOL of 0.76 for A and 1.18 for B, as recommended in the labelling protocol (ThermoFisher).

Following the N terminus labelling and removal of excess dye, much of ComR was lost, resulting in final concentrations of 1 mg/mL (or 0.03 mM). Both samples A and B from the initial labelling step were labelled the same way with the second fluorophore. The C2 Maleimide was dissolved in 100 µL DMSO, and 8 µL (0.064 mmol) was added to 100 µL of protein (0.003 mmol) to be in a 20-fold molar excess. The protein and dye were incubated for 3 hours at room temperature, shaking, and protected from light, after which the reaction was quenched with the addition of 5 µL of 20 mM β-ME (final concentration of 1 mM). Next 2 µL of 5 M NaCl was added (to reduce protein binding to the spin column) and the samples were cleaned up with another dye removal spin column (ThermoFisher A44296). After the second labeling the DOL was calculated again for both samples and for both the succinimidyl ester (555) and C2 maleimide (647). The DOLs for sample A were 0.528 (555) and 1.56 (647) and for sample B were 0.61 (555) and 1.25 (647). It is important to note however, that calculating the DOL of a singly labelled protein included the correlation factor (CF) at 280 nm for the dye used. When calculating the DOL of the double labelled protein, there was no realistic way to determined correlation factors at the specific wavelengths of the other dye, therefore the backgrounds of the dyes relative to one another were not taken into account, making the DOL less accurate after the second labelling.

The final buffer condition of the proteins after each labelling step was approximately: 14 mM sodium phosphate (pH 6.5 or 7.0), 85 mM Tris-HCl (pH 7.5), 160 mM NaCl, 0.9 mM β-ME, and 16% DMSO. The double labelled ComR was flash frozen and stored at -80 °C, for a number of months. Directly before use the protein was thawed and the DOLs were recorded again: with DOLs of 1.4 (555) and 3.3 (647) for sample A, and 0.55 (555) and 0.86 (647) for sample B. The

protein concentration of sample A was significantly lower however, possibly accounting for inaccurate DOL values.

2.8.2 Experimental parameters

The FRET experiments were performed using a Jobin-Yvon Fluorolog-3 fluorescence spectrophotometer (Horiba), at an excitation wavelength of 545 nm, scanned from 500 to 700 nm, and with the slit widths set to 5, 5 nm bandpass. For each experiment, double labelled ComR from sample B was used, and samples were diluted with gel filtration buffer, unless stated otherwise. First two controls were run: labelled ComR on its own, followed by labelled ComR in 7 M urea (completely denatured ComR). Next a binding assay was performed with 0.05 µM labelled ComR and increasing concentrations of Prx, ranging from 0.1 to 4 µM.

2.9 Electro-mobility shift assays

First, the fluorescently labelled *comS* promoter region was generated with primers used in a previous study[19]: UA159_Pcoms_F and FAM_UA159_PcomS_R. *Streptococcus mutans* UA159 was grown in 14 mL THY media at 37 °C overnight, in a sealed 15 mL conical tube. The following morning cells were collected by centrifugation at 4200 rpm, 4 °C, and a genome extraction was performed using the GeneJET Genomic DNA purification kit and corresponding protocol for Gram positive bacteria (ThermoFisher). The approximately 300 base-pair region around the *comS* promoter was then amplified using the UA159_Pcoms_F and FAM_UA159_PcomS_R primers[19] and Q5 polymerase (New England Biolabs), run on a 0.8 % agarose gel to verify sample purity, and cleaned up using either the GeneJET PCR cleanup kit (ThermoFisher) or the Monarch PCR & DNA cleanup kit (New England Biolabs). The FAM-labelled primer, as well as the resulting FAM-labelled DNA, were always kept in the dark and protected from light.

The electro-mobility shift assays (EMSAs) were performed as previously described[19,64]. For each 20 µL sample, 5 µL of 4× sample buffer (80 mM HEPES pH 7.8, 80 mM KCl, 20 mM $MgCl_2$, 0.8 mM EDTA) and 5 µL 50 % glycerol was added, and the entire sample was run on a 5% native-PAGE gel, in a potassium phosphate running buffer (50 mM potassium phosphate pH 7.5). The gels were made using 40% Acrylamide/Bis-Acrylamide 29:1, and potassium phosphate pH 7.5 buffer, as described previously[64]. The first experiment was performed identically as previously described[45], with the exception that the His-tag of ComR *mutans* had been removed. First 4 µM ComR *mutans*, 8 µM XIP, and increasing concentrations of Prx were incubated together for 30 min, followed by the addition of 100 ng of FAM-labelled P*comS* and incubation for an additional 15 min. In addition, the same experiment was repeated with the pre-formation of the ComR:Prx complex, as well as the pre-formation of the ComR:XIP:DNA complex. For the pre-formed ComR:Prx complex, ComR and Prx were incubated together for 30 min, followed by the addition of XIP and labelled DNA for 15 min. For the other experiment, ComR, XIP and labelled DNA were incubated together for 30 min, after which Prx was added and the samples were allowed to incubate for an additional 15 min. Control experiments were also performed using PrxD32A which is unable to bind to ComR. For the XIP competition experiment, 4 µM ComR, 8 µM Prx, and 100 ng of FAM-labelled DNA were incubated together for 30 min. Next increasing concentrations of XIP were added, and the samples were allowed to incubate for an additional 15 min before loading onto the gel.

For each experiment samples were incubated at room temperature, and protected from light after the addition of the labelled DNA. Stock concentrations/dilutions of all proteins were made such that consistent volumes would be added at each step. Samples were run on 5 % native PAGE

gels at 100 V for 50 min, with pre-chilled running buffer, at 4 °C, and protected from light. The gels were imaged using a Fluorochem Q imager (Protein Simple).

2.10 Fluorescence competition assays

2.10.1 Experimental conditions and parameters

For all of the fluorescence assays a dansyl labelled XIP (GLDWWSL), synthesized by Abclonal, was used. The peptides were stored at -80 °C, and re-suspended in the experimental buffer immediately prior to use. FITC-XIP was initially used; however non-specific interactions between Prx and FITC were observed at high concentrations. Each of the fluorescence assays was performed using a Jobin-Yvon Fluorolog-3 fluorescence spectrophotometer (Horiba) with an excitation wavelength of 335 nm and scanned from 400 to 600 nm and excitation and emission slit widths at 5 nm bandpass. The maxima of the fluorescence intensity (535 nm) was used plot the binding curves. For the fluorescence anisotropy experiment, the slit widths were set to 10 nm bandpass, and samples were scanned from 510 to 570 nm. All samples were prepared in triplicate, and measured in a 1.4 mL quartz cuvette (fireflysci).

2.10.2 Sample preparation

For each experiment, ComR *mutans* and Prx were equilibrated into an assay buffer of 20 mM HEPES pH 7.5, 100 mM NaCl, and 1 mM β-ME by SEC, as the regular (Tris) gel filtration buffer had significant background fluorescence in the 400-450 nm range. Before running each of the experiments, the fluorescence intensity of dansyl-XIP (0.2 µM) with increasing concentrations Prx (up to 10 µM) was measured to ensure that Prx was not interacting with the labelled XIP.

First a binding curve of dansyl-XIP with increasing concentrations of ComR was performed in order to determine an appropriate concentration of ComR to use for competition assays. Dansyl-XIP at a concentration of 0.2 µM, was incubated with increasing concentrations of

ComR *mutans*, ranging from 0.125 to 15 µM, for 1 hour at room temperature (protected from light).

Competition assays were then performed with 0.2 µM dansyl-XIP, 0.5 µM ComR, and increasing concentrations (0.125 to 4 µM) of either non-labelled XIP, or Prx. A control competition assay with non-labelled XIP was first performed to ensure that the dansyl-XIP interaction was specific. 10x stock concentrations and dilutions of each protein were prepared the night before, and XIP (both labeled and unlabeled) was dissolved in the assay buffer immediately prior to the experiment. The proteins and dansyl-XIP were added at the same time (Prx or XIP dilution, buffer, dansyl-XIP, and ComR last) and incubated together for 1 hour, room temperature, protected from light. For fluorescence anisotropy, samples were prepared identically to the Prx competition samples, with 0.2 µM dansyl-XIP, 0.5 µM ComR, and Prx ranging from 0.25 to 8 µM.

2.11 Pull down experiments with Streptococcal lysates and media

2.11.1 Growth and induction of *S. pyogenes* MGAS315

S. pyogenes MGAS315 was streaked directly from the freezer stock (in 20% glycerol) on trypticase soy agar (TSA II) plates supplemented with 5% sheep's blood (ThermoFisher) and allowed to grow for a minimum of 24 hours at 37 °C. Because MGAS315 contains no selective marker, cells were first streaked on the blood agar to ensure complete/beta-hemolysis and that no contaminants were present. The next day, isolated colonies were picked and inoculated in 7.5 mL of chemically defined media (CDM), which was pre-warmed to 37 °C, in an 8 mL, sealed screw-capped tube and allowed to grow overnight at 37 °C. The following morning, overnight cultures were diluted into fresh CDM at a 1:20 ratio (375 µL overnight in a final volume of 7.5 mL media) in triplicates, and allowed to grow until an optical density at 600 nm of approximately 0.2. At this point cells were either induced with 50 nM XIP (MGAS315) to induce the ComRS pathway[40] or

200 ng/mL mitomycin C, which is a DNA damaging agent and has been shown to induce lysogenic bacteriophages in *S. pyogenes*[24]. The XIP induced cells were allowed to grow for an additional 2 hours, reaching an OD_{600nm} of 0.8. The mitomycin C induced samples initially continued to grow after induction; however they decreased in optical density as bacteriophage induced lysis occurred. These cells were allowed to grow for an additional 4 hours after induction and the OD_{600nm} values ranged from 0.23 to 0.84. The three cultures of XIP induced cells, and mitomycin C induced cells were then combined and spun down at 2000 *g*, 4 °C, for 20 min, and re-suspended in 600 µL lysis buffer consisting of 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% glycerol, and divided into 3 centrifuge tubes (approximately 250 µL each). The media was also collected for each sample, filtered with a 0.2 µm syringe filter, and then concentrated using a 3 kDa centrifugal filter (Amicon) to a volume of approximately 500 µL. The cells and media were then flash-frozen and stored at -80 °C, until use.

2.11.2 Pull-down experiments

Samples were thawed in icy water and 1 mM PMSF was added to each. Each 250 µL sample containing cells was then lysed by sonication. For each XIP and mitomycin C induced cells, one sample was lysed for 1 round of sonication (1 s on 2 s off, × 10), one sample lysed for 2 rounds, and one for 3 rounds (with 30 second breaks in-between) on ice. Since the purpose of this experiment was to search for a potential binding partner (rather than quantification), this approach was chosen to increase the range of possible soluble proteins present. Sonication for a shorter amount of time, resulted in less efficient lysis, however more sonication, while increasing lysis efficiency, also resulted in a loss of soluble protein. The media was subjected to one round of sonication, to lyse any potential phage particles. The combined XIP induced cell lysate, combined

mitomycin C induced cell lysates, and the media for each, were then centrifuged at 21000 *g* for 10 min, at 4 °C, to remove insoluble material.

The supernatant for each sample was divided in 2 centrifuge tubes in equal volumes. To one, (the Prx pull-down sample) 120 µg of purified Prx (with His tag) was added, and to the other (background control) lysis buffer was added instead. Each of the volumes were increased to 700 µL with lysis buffer to allow for more efficient sample mixing and reduce the amount of resin lost. The samples were allowed to incubate for 1.5 h, at 4 °C rotating. During this incubation step the Ni-NTA resin (Goldbio) was prepared, as described in section 2.4.2. After washing and equilibrating into the lysis buffer, 40 µL of a 50% slurry (20 µL resin), was added to each sample and allowed to incubate for an additional hour, rotating, and at 4 °C. The samples were then washed 3 times with 600 µL wash buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 20 mM imidazole, 10% glycerol) mixing for 10 min at 4 °C, followed by centrifugation at 1200 *g* for 5 min, and aspiration of the remaining buffer. Finally, the samples were eluted with 40 µL of elution buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 500 mM imidazole, 10% glycerol). Each sample was pipette mixed, and centrifuged at 1200 *g* for 10 min. The supernatant of each sample (40 µL), was then carefully transferred to a new centrifuge tube, flash frozen, and stored at -80 °C.

2.11.3 Trypsin digestion and sample clean-up for mass spectrometry

Each pull-down sample (40 µL) was adjusted to final concentration of 100 mM ammonium bicarbonate, in a volume of 100 µL. Next the samples were reduced, adding 10 µL of fresh 100 mM DTT (dissolved in 100 mM ammonium bicarbonate). Samples were gently vortexed to mix, and allowed to incubate for 35 min at 58 °C. After cooling to room temperature, 10 µL 500 mM iodoacetamide (IAA), also dissolved in 100 mM ammonium bicarbonate, was added to alkylate the proteins. Samples were allowed to incubate for 45 min at room temperature, protected from

light. Next, 16 µL of 500 mM IAA (in 100 mM ammonium bicarbonate), and allowed to react for an additional 10 min, in the dark, to quench the excess IAA. Finally, sequencing grade trypsin (Promega) was added to each sample (2 µg to pull-down samples with Prx, and 1 µg to background samples without Prx). The samples were then incubated overnight for 18 h, at 37 °C.

The following morning the digested proteins were acidified and cleaned up for the downstream mass spectrometry using Peirce C18 spin columns (ThermoFisher: 89870) and following the associated protocol. First the samples were acidified by adding 50 µL 4× sample buffer (2% TFA, 20% ACN), to 150 µL sample. The spin columns were then prepared by washing with activation solution (50% ACN) 2 times, followed by two washes with equilibration solution (0.5% TFA, 5% ACN). Samples were then loaded onto the columns, centrifuged at 1500 *g* for 1 min, and the flow-through was re-run over the column. The column was then washed 3 times with 200 µL equilibration solution. Finally, the peptides were eluted with 20 µL of elution buffer (0.1% TFA, 70% ACN). The peptide samples were then stored at -20 °C before being sent away for mass spectrometry. Mass spectrometry was performed at the Manitoba Centre for Proteomics and Systems Biology (pay-per-service).

2.11.4 Preliminary data analysis

Because the pull-downs revealed numerous hits for both the background and experimental samples, the lists for Prx pull-down samples (for both XIP and mitomycin C induction) were individually filtered. This was done by selecting the pull-down sample list, to a cut-off of values below Log(E) = -10.0 (larger negative values are more probable hits) and subtracting the list by all the samples that appeared in the background control (no Prx) list. The cut-off was chosen visually, based on where the frequency of the same E values begins to significantly decrease[65]. The filtered pull-down lists for XIP and mitomycin C induction were also filtered to make sure

none of the remaining hits were significantly present in the background samples of the other condition. Three separate lists remained including protein hits that were: 1) significant in both XIP and mitomycin C induced samples, 2) significant in just XIP induced samples, and 3) significant in just mitomycin C induced samples. Each of these potential hits were further looked into, using a variety of bioinformatic servers (BLAST[66], Uniprot, Phyre2[67], etc.), and proteins considered essential and/or constitutively expressed (for example RNA polymerase processing proteins) or that are known to bind metals (high potential for interacting with the Ni-NTA resin) were removed.

3 RESULTS

3.1 Prx interacts with each ComR protein type

3.1.1 Prx pull-downs with various ComR proteins

ComR is a highly variable protein (ranging from 50 to 30% identical between different species), and is grouped into three different types based on the XIP it recognizes[12]. It was therefore of interest to determine if the ComR:Prx interaction was specific to only similar ComR proteins, or broad enough for Prx to interact with ComR proteins from different species and of different types. In order to show the specificity of Prx to various ComR orthologs, a pull-down assay was performed. Purified Prx MGAS315 with a 6-His tag was used as bait, and its ability to pull down a number of ComR proteins of different types and from different species was assessed. The purified ComR proteins (with their His-tags removed) included ComR *mutans* (positive control), ComR MGAS5005, ComR *suis*, and ComR *thermophilus.* Pull-downs with Rgg3, a related transcription factor to ComR were also attempted. However, Rgg3 interacted with the Ni-NTA resin and was present in the background control. Pull-downs with GST-Prx were also attempted; however the GST tag prevented Prx from pulling down any ComR proteins, including the ComR *mutans* positive control.

The pull-down assay, shown in Figure 3, shows that Prx was able to pull down each ComR protein, including examples from each ComR type I, II and III. However, Prx displayed a clear preference for the type II ComR proteins, including ComR *mutans* and ComR MGAS5005. This is expected as the Prx used is from *S. pyogenes* MGAS315, which also expresses a type II ComR. Together this indicates that the interaction is both broad and specific, possibly with Prx having evolved to effectively target the ComR type II while interacting with conserved region that is similar among all three ComR types.

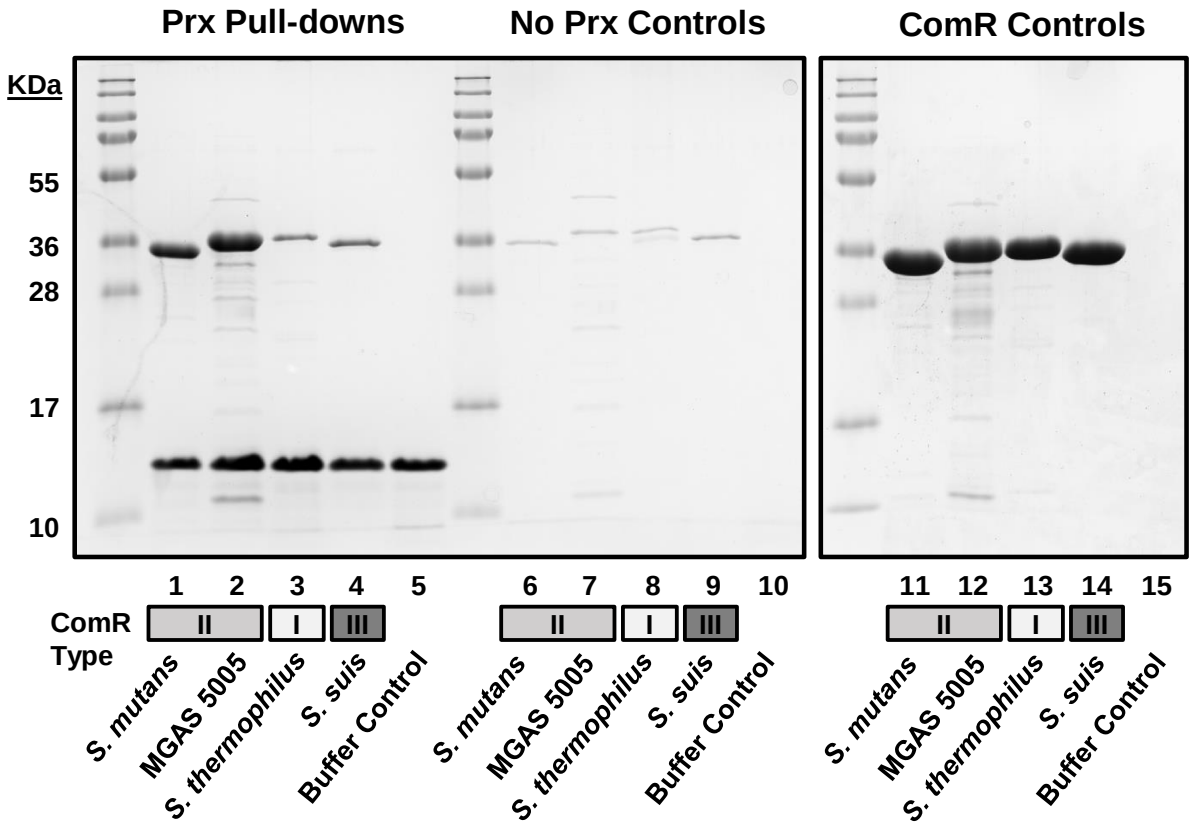

Figure 3. Prx pull down assay with various ComR orthologs. In lanes 1 to 5 Prx (with a His tag) was used as bait and incubated with various ComR Proteins (ComR mutans, MGAS5005, suis, thermophilus). Following numerous wash steps and elution of Prx from the resin, the samples were run on 15% SDS-PAGE gels. Prx (near 12kDa on SDS-PAGE) interacted with each ComR protein, and as a result ComR (~36 kDa) was present in the elution as well. Lanes 6-10 are of the same experiment, however lacking Prx, showing the ComR background levels. In lanes 11-14 the ComR proteins (the same amounts as added to the pull-down experiments) were run on an SDS-PAGE gel. The shaded boxes below the lane numbers indicate the type of ComR protein used (type I, II, or III).

3.1.2 Size exclusion chromatography binding assays

The interaction of Prx with each ComR protein was also shown by size exclusion chromatography (SEC) in Figure 4. Prx was incubated with each ComR ortholog at a 1.5 molar excess, after which the complex was run over a SEC column. Since SEC separates proteins based on hydrodynamic radius (size), with larger proteins eluting first and smaller proteins eluting later, we would expect to see a size-shift in the event that Prx interacts with ComR. Complex formation was observed with all four ComR orthologs based on the leftward shift of the complex (Prx and ComR incubated together) vs ComR on its own (upper panels of Figure 4). This is further shown by the SDS-PAGE gels of each fraction (lower panels of Figure 4), where the fractions corresponding to the complex or ComR (approximately 10 or 11 mL), contain bands for both ComR and Prx, while Prx on its own elutes at a much later volume (~14 mL).

3.2 Prx does not interact with Rgg3

Rgg3 could not be used in the pull-down assay (due to background binding to the resin), however its ability to interact with Prx was assessed by SEC (Figure 5). Rgg3 *S. pyogenes* is a transcription factor that is closely related to ComR, belonging to the same Rgg protein sub-family[43]. Rgg3 *S. pyogenes* also shares approximately 15-20% amino acid sequence identity with various ComR proteins, however despite its similarities and the fact that Prx is able to interact with ComR proteins sharing only 30% sequence identity, no interaction was observed by SEC. Since both proteins are expressed by *S. pyogenes*, this suggests that Prx specifically evolved to inhibit ComR and not Rgg.

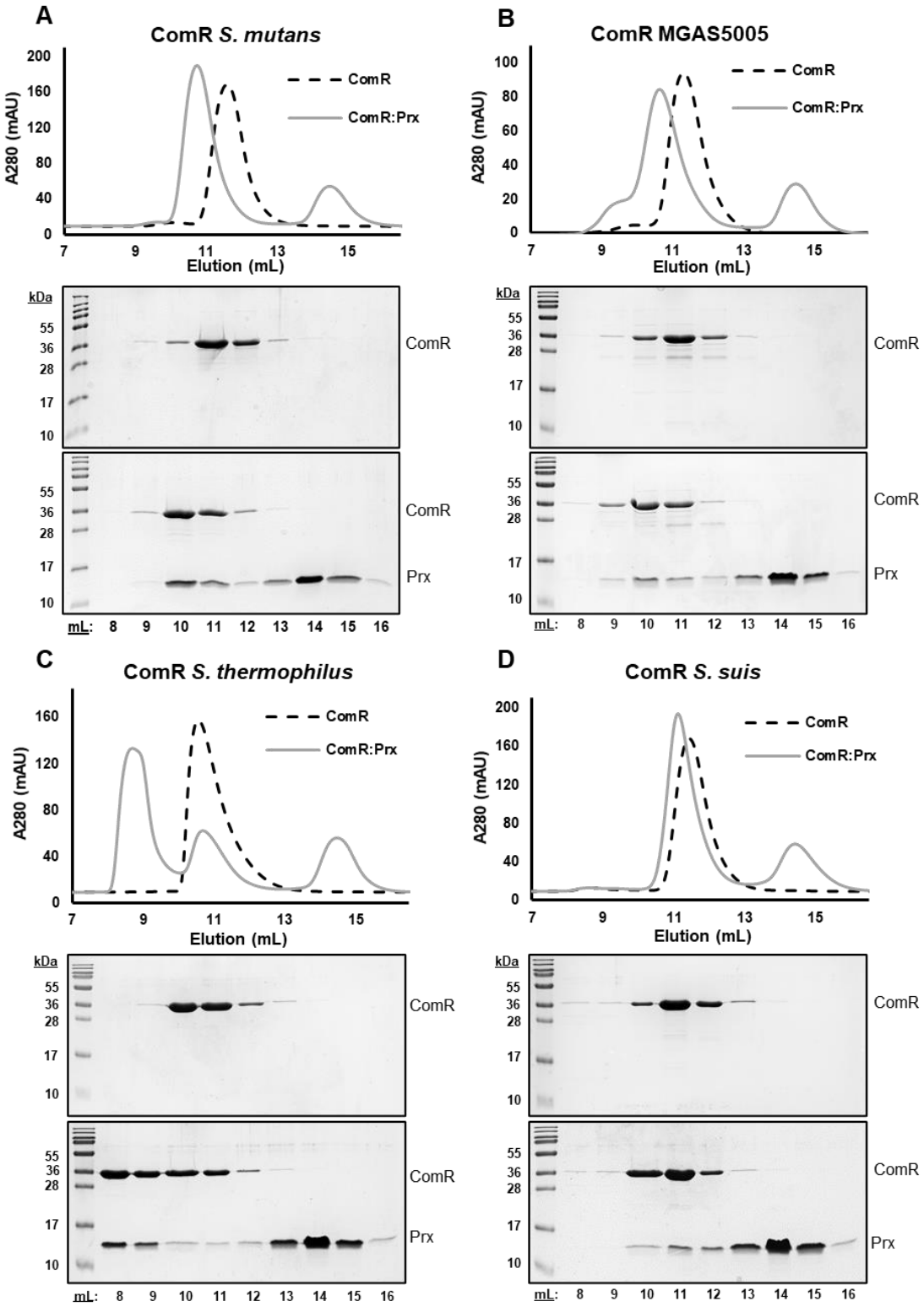
A
ComR S. mutans
ComR
ComR:Prx
A280 (mAU)
Elution (mL)
kDa
ComR
Prx
mL:
B
ComR MGAS5005
C
ComR S. thermophilus
D
ComR S. suis

Figure 4 (Previous page). Size exclusion chromatography binding assays of Prx with each ComR type. Prx was incubated with each ComR at a 1.5 molar ratio, followed by running the complex over size exclusion column. For each complex, the upper panel shows the SEC elution profile, and the lower panel shows the corresponding protein fractions run on an SDS-PAGE gel. **A)** shows Prx in complex with ComR *mutans*, a type II ComR, **B)** shows Prx in complex with ComR MGAS5005, also a type II, **C)** shows Prx and ComR *thermophilus*, a type I, and **D)** shows Prx with ComR *suis*, a type III protein.

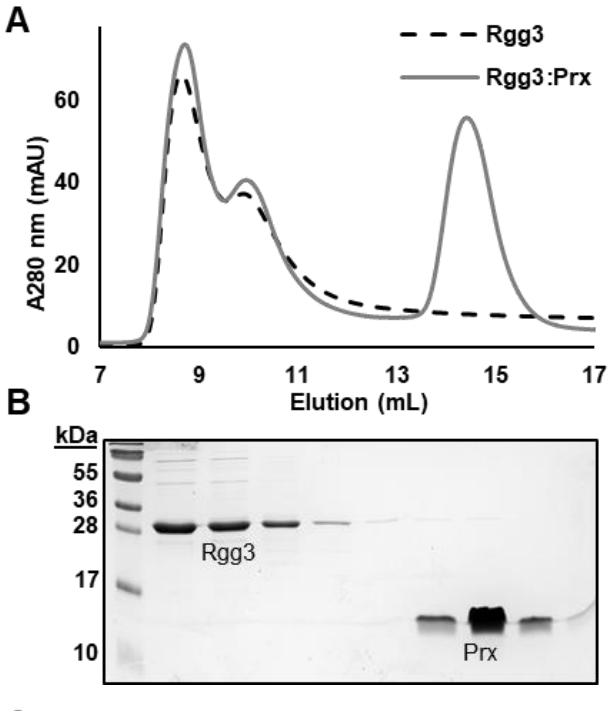

Figure 5. Size exclusion chromatography binding assay of Prx with Rgg3. **A)** Size exclusion chromatography trace of Prx and Rgg3 incubated together, and Rgg3 run on its own. **B)** SDS-PAGE gel with wells corresponding the above fractions from the SEC trace, highlighting that no Prx is co-eluting with Rgg3.

3.3 Prx interacts with the ComR DNA binding domain

3.3.1 Prx interacts with the ComR DBD and not the TPR

One of the primary goals of this project was to characterize the interaction of ComR:Prx by X-ray crystallography. However, the full ComR:Prx complex did not crystallize so partial constructs of ComR *mutans* were made including only the DNA binding domain (DBD) and the tetratricopeptide repeat (TPR) domain. This was done in order to both narrow down where the interaction takes place, as well as to generate new constructs for crystallization. Both the DBD and TPR domains expressed and purified well, and were stable in the same buffer conditions as ComR.

In order to determine if either one of the partial domains could interact with Prx, SEC binding assays were performed. Based on the elution profiles and SDS-PAGE gels in Figure 6, Prx appears to be able to bind the DBD and not the TPR domain. Although Prx and the DBD are similar in size, making it difficult to visualize complex formation by the SDS-PAGE gel, a noticeable shift is observed when the complex is run over the SEC column. For the TPR domain, no shift is observed and Prx does not co-elute with the TPR on the SDS-PAGE gel.

3.3.2 Prx has a higher affinity for the DBD than the full ComR protein

Since the DBD:Prx interaction was difficult to observe by SEC, isothermal titration calorimetry (ITC) was performed, shown in Figure 7. Prx at a concentration of 300 µM was titrated into both 20 µM ComR *mutans* and DBD. The ComR:Prx interaction had a K_d of 390 +/- 80 nM, similar to previous results[19]. The thermodynamic properties of the interaction were also similar as previously described[19], with a ΔH of -11.8 +/- 0.3 kcal/mol and a ΔS of -10 +/- 1 cal/mol/K. Prx also bound to the DBD, with a significantly higher affinity of K_d = 50 +/- 20 nM, and had a ΔH of -20.8 +/- 0.3 kcal/mol and ΔS of -37 +/- 1 cal/mol/K. For both ComR:Prx and DBD:Prx the interactions were enthalpy driven, with favorable ΔH values and unfavorable ΔS values. This

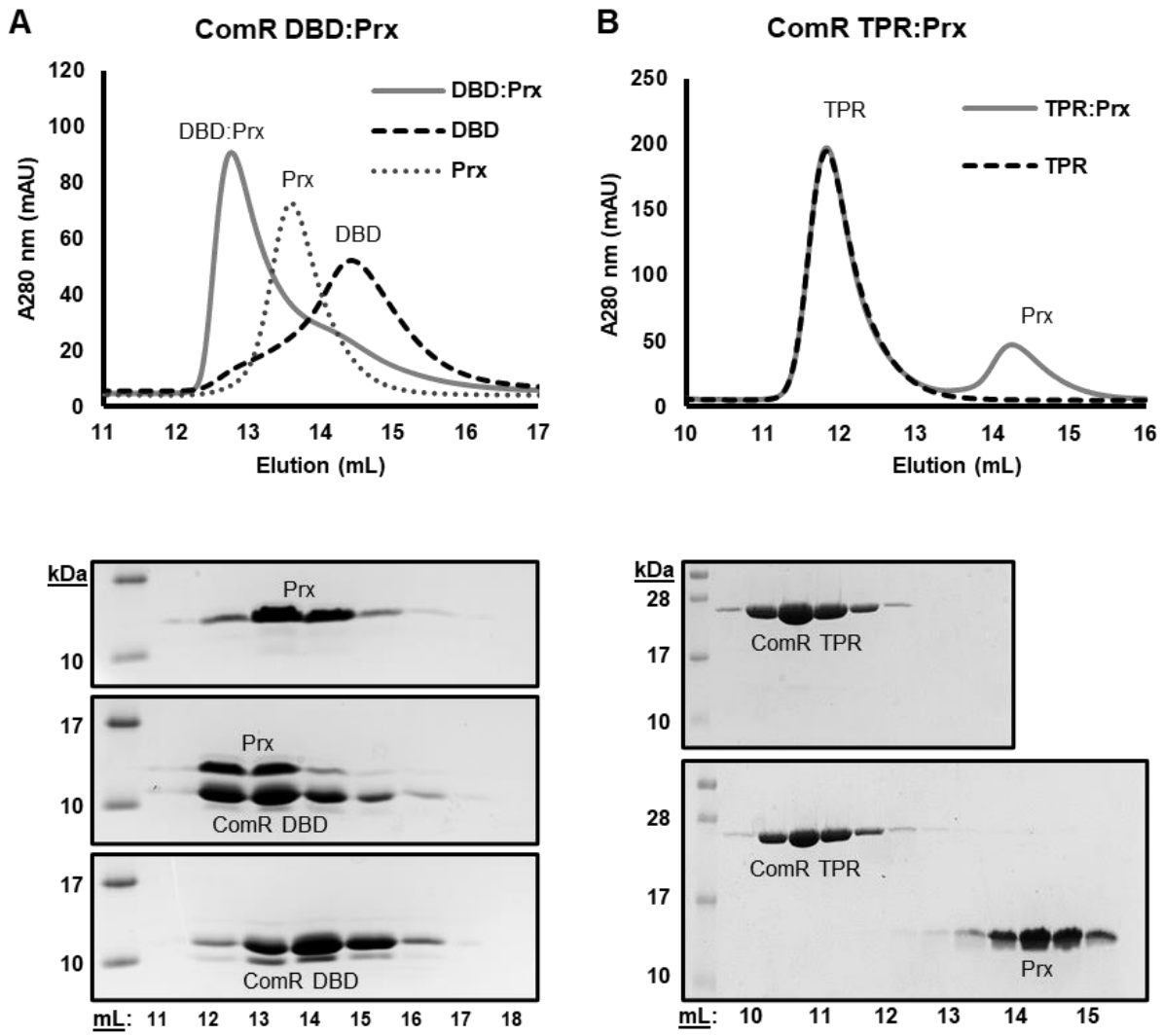

Figure 6. Size exclusion chromatography binding assays of Prx with the ComR DBD and TPR domains. A) SEC binding assay of Prx with the ComR DBD. **B)** SEC binding assay of Prx with the ComR TPR domain. For both **A)** and **B)** the upper panels show the SEC elution traces, and the lower panels show SDS-PAGE gels with the wells corresponding to the fractions of the elution traces above.

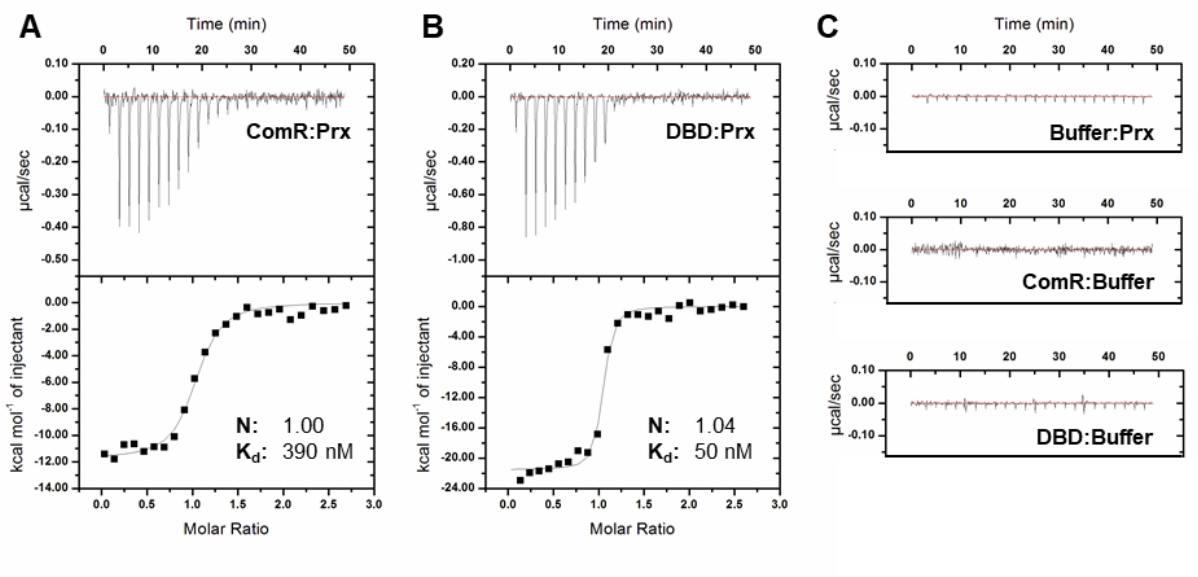

Figure 7. Isothermal titration calorimetry of Prx with ComR and the ComR DBD. ITC of **A)** Prx injected into ComR *mutans*, **B)** Prx injected into the ComR DBD, and **C)** various background controls including buffer injected into Prx, or ComR injected into buffer, and DBD injected into buffer.

indicates that the interaction is more likely to be stabilized by electrostatic, rather than hydrophobic interactions. Relative to ComR:Prx, the DBD:Prx complex had a much more favorable enthalpy and much less favorable entropy. A possible reason for this could be that Prx causes a conformational change in ComR when the TPR domain is present, resulting in an increase in entropy.

3.4 X-ray crystal structure of Prx in complex with the ComR DBD

3.4.1 Prx interacts directly with residues responsible for recognizing DNA

Although the full-length ComR:Prx complex did not crystallize, an X-ray crystal structure of the DBD of ComR in complex with Prx was solved (PDBid:7N10) to a resolution of 1.65 Å, with an R_{free} of 20.4. All of the data collection and refinement statistics are displayed in Table 1.

The DBD:Prx structure reveals that Prx interacts specifically with the DNA binding region of the DBD (Figure 8A) . The interaction surfaces of both proteins are highly conserved (Figure 8B and Figure 9), as well as electrostatic (Figure 8B). The Prx interaction surface is electronegative, with many acidic residues interacting with the electropositive/basic DNA binding surface of the DBD. Prx forms salt bridges and hydrogen bonds with key residues known to recognize DNA including R33, R37 and K49[41], as well as the residue Q34. Q34 is responsible for holding the DBD directly adjacent to the TPR domain of ComR, rather than in the extended/active conformation[12]. Residue D32 of Prx, which is highly conserved, interacts with three of these important and conserved residues of ComR (R33, Q34, R37) and was shown to be essential for the interaction by SEC binding assays (Figure 11A), DNA binding inhibition in EMSAs, as well as ITC. Prx also forms several other hydrogen bonds throughout the interaction surface, which when disrupted, interfere with complex formation. This includes E6 of Prx which interacts with ComR T46, and Prx E9 which interacts with ComR T44 and the ComR peptide backbone.

Table 1. X-ray data collection and refinement statistics for DBD:Prx

	DBD:Prx	DBD:Prx (proteolysis)
Data Collection		
Wavelength (Å)	1.28329	0.97911
Space group	$P2_12_12_1$	$P4_12_12$
Cell dimensions		
a, *b*, *c* (Å)	38.56 41.92 90.0	36.8 36.8 189.3
α, β, γ (°)	90.0 90.0 90.0	90.0 90.0 90.0
Resolution (Å)	45 -1.65 (1.68-1.65)	36.14-1.60 (1.66-1.60)
Total reflections	111682 (5567)	215198 (7232)
Unique reflections	18234 (918)	18256 (1746)
CC(1/2)	0.998 (0.867)	0.999 (0.641)
R_{merge}	0.067 (0.727)	0.09 (1.30)
R_{pim}	0.029 (0.316)	0.03 (0.46)
$I/\sigma I$	15.2 (2.3)	15.6 (1.5)
Completeness (%)	99.9 (100.0)	100.0 (99.7)
Redundancy	6.1 (6.1)	11.7 (8.5)
Refinement		
R_{work} / R_{free} (%)	18.3 / 20.4	17.9 / 20.4
Average B-factors (Å^2)	31.7	25.7
Protein	30.3	24.1
Ligands	-	53.6
Water	39.8	32.1
No. atoms	1250	1312
Protein	1062	1091
Ligands	0	32
Water	188	206
Rms deviations		
Bond lengths (Å)	0.004	0.005
Bond angles (°)	0.660	0.690
Ramachandran plot (%)		
Total favored	100.00	99.24
Total allowed	0.00	0.76
PDB code	7N10	7N1N

Statistics for the highest-resolution shell are shown in parentheses

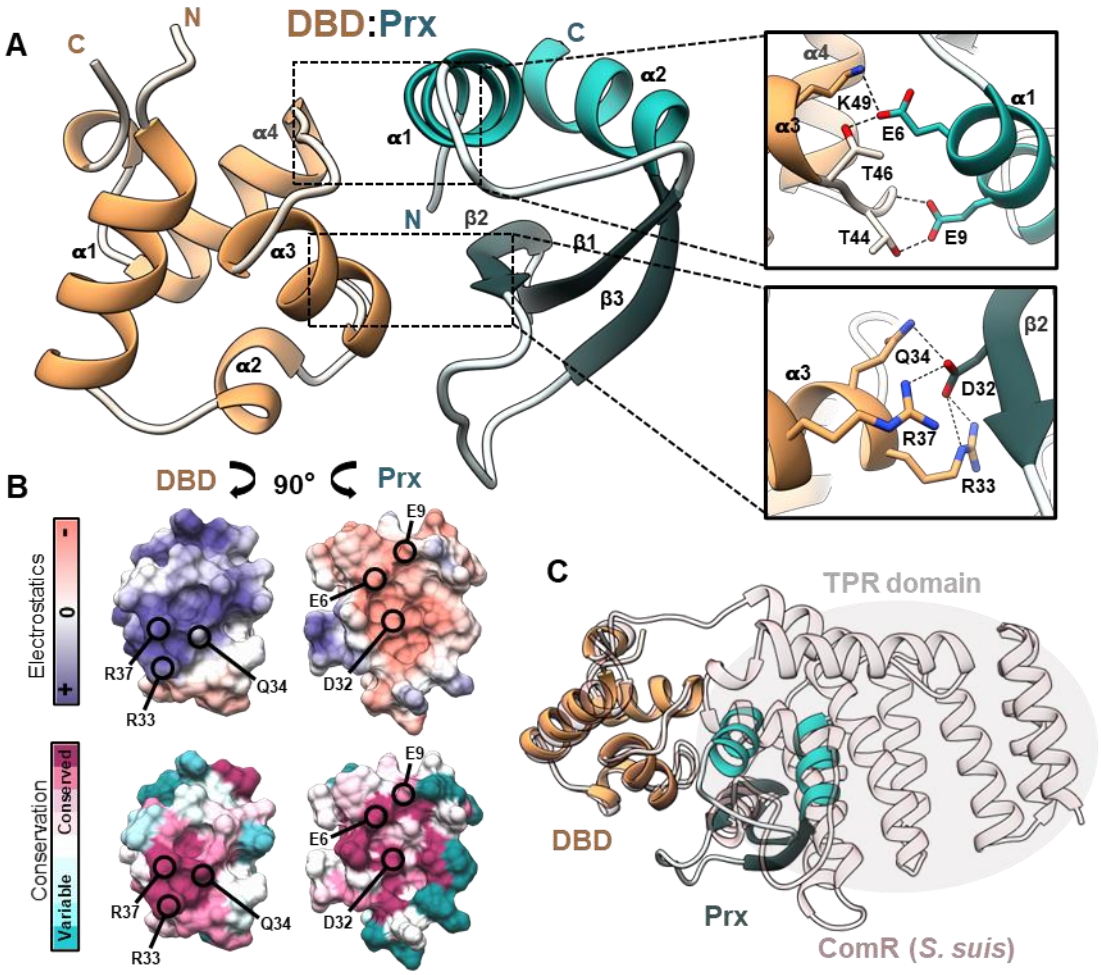

Figure 8. X-ray crystal structure of Prx in complex with the ComR DBD. A) Structure of the DBD:Prx complex with various residues highlighted within the interaction surface. **B)** Electrostatics and conservation overlayed on the interaction surfaces of both the ComR DBD and Prx with key residues highlighted. **C)** The DBD:Prx complex aligned to the DBD of ComR *suis* (PDBid: 5FD4), showing that Prx clashes with the TPR domain of ComR in its apo conformation and a conformational change must take place.

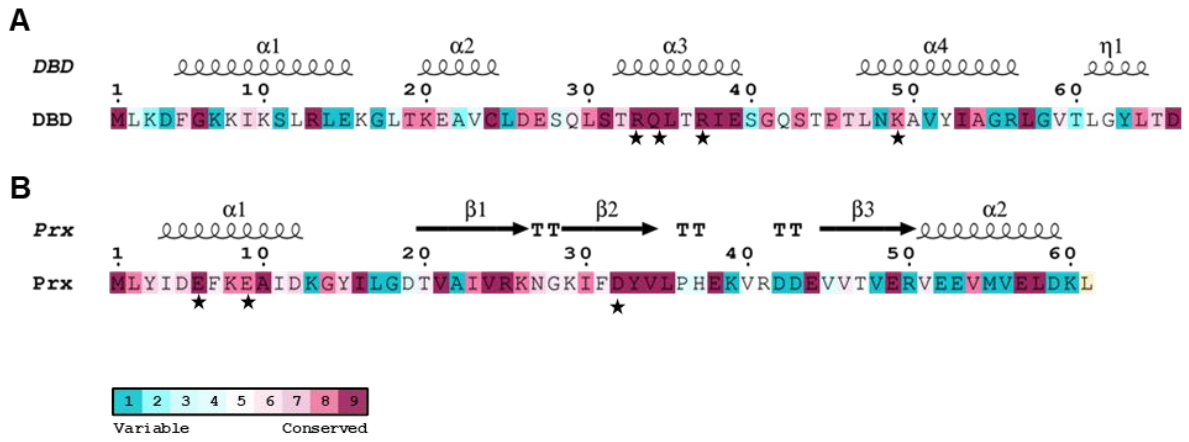

Figure 9. Amino acid sequence conservation of the ComR DBD and Prx. Sequence conservation was determined and mapped onto the protein sequences for **A)** the ComR DBD and **B)** Prx, using the ConSurf server (https://consurf.tau.ac.il/). The residues shown to form contacts in the DBD:Prx complex are starred.

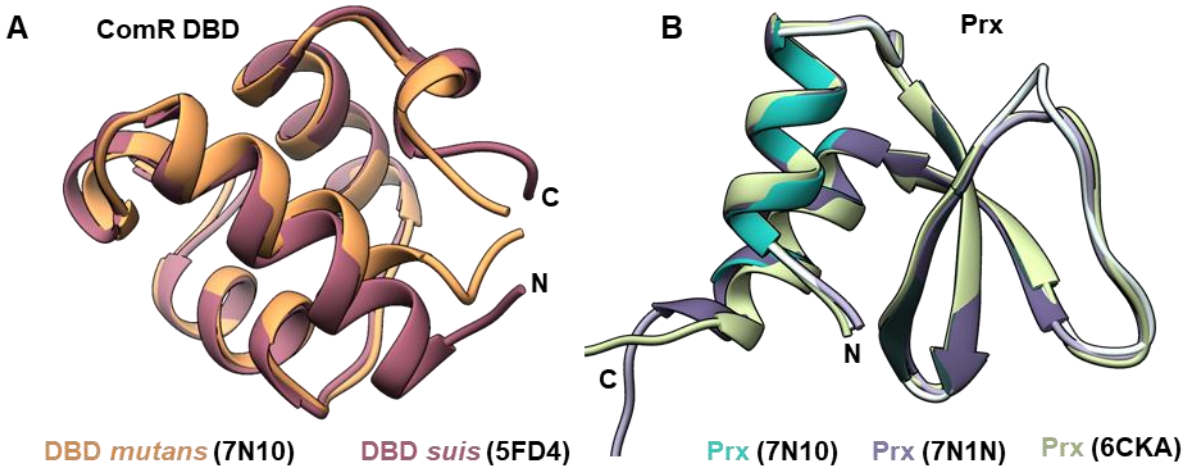

Figure 10. Structural alignments of the ComR DBD and Prx with existing structures. A) The ComR DBD from the DBD:Prx structure (7N10) was aligned to the DBD of ComR *suis* (PDBid: 5FD4) using UCSF Chimera[63], with a resulting RMSD of 0.771 Å^2. **B)** Prx from both DBD:Prx complex structures (7N10 and 7N1N) aligned with an existing structure of Prx not in complex, showing each Prx structure is nearly identical.

When the individual DBD of the solved complex structure was compared to the DBD of ComR *suis* (PDBid: 5FD4)[12] the structures were very similar with an RMSD of 0.771 Å^2 for the ComR *mutans* DBD compared to the DBD from *suis*, which share 50% sequence identity. When comparing Prx solved in complex with the DBD to an existing Prx structure (PDBid: 6CKA)[19] the structures were nearly identical. Overlays of the ComR *mutans* with the ComR *suis* DBDs along with Prx on its own and in complex with the DBD can be found in Figure 10A and B. Although crystallographic conditions are not always an accurate representation of a protein's conformation in solution, it is interesting to note that Prx had the same conformation when crystallized on its own and when in complex with the DBD.

The DBD:Prx structure was also aligned with the full ComR *suis* protein in its apo conformation (PDBid: 5FD4)[12]. With the DBDs aligned, Prx completely clashes with the TPR domain of ComR. This suggests that in order to bind, Prx must induce a conformational change in ComR to release the DBD from the from the TPR domain.

3.4.2 Prx point mutants verify the interaction surface

To confirm that the structural contacts shown in the crystal structure were biologically relevant, Prx protein variants were designed and assessed for their ability to interact with ComR. A number of residues, both thought to be important and unimportant for the interaction with ComR, were substituted to alanine. This includes Prx E6A, E9A, D12A, F31A, and D32A, (as well as a number of other protein variants that were designed before the DBD:Prx crystallized). Each of these protein variants were assessed for their ability to form a complex with ComR by SEC (Figure 11A and B). PrxD12A and F31A all resulted in a peak at the same elution volume as the ComR:PrxWT complex, meaning that removing the functional groups of these residues had no observable effect on the interaction. The PrxE6A, PrxE9A, and PrxD32A variants however,

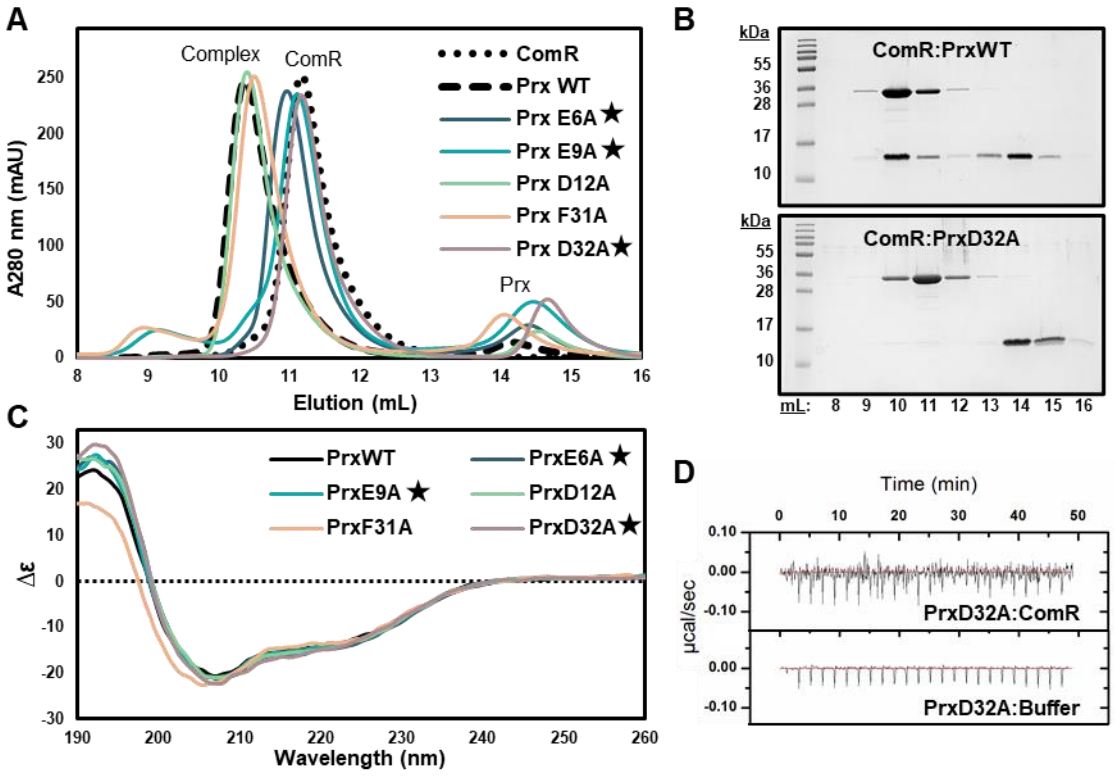

Figure 11. Binding assays with Prx variants to verify structural contacts. A) SEC binding assay of various Prx point mutants, with residues shown to contact the DBD starred. **B)** Example SDS-PAGE gels with wells corresponding to the fractions in A) of WT Prx which complexes with ComR, and PrxD32A which does not complex with ComR. **C)** Circular dichroism of the Prx variants compared to the WT protein to ensure structural integrity. **D)** PrxD32A injected into ComR, showing no observable interaction by ITC.

showed a significantly decreased ability to interact with ComR, as the resulting peak of the complex eluted with the peak for ComR on its own. These three variants all involved the substitution of a residue shown to form structural contacts in the DBD:Prx crystal structure (Figure 8A).

One important consideration when generating protein mutants is whether a change in the proteins fold or stability is introduced. In order to be sure that the observed inability to interact with ComR was due to the specific substituted residue and not a significant change in secondary structure, circular dichroism (CD) was performed with each of the Prx variants and compared to the WT (Figure 11C). For the most variants, the spectra for each Prx variant was identical to the WT, indicating the same secondary structure. One exception was PrxF31A, which showed a slightly less pronounced peak in the α-helical range. However, this protein variant was still able to form a complex with ComR.

3.5 Small angle X-ray scattering provides a model for the full ComR:Prx complex

3.5.1 The ComR:Prx complex is more flexible than apo ComR

Because there was only a structural model of the ComR DBD bound to Prx, size exclusion chromatography coupled small angle X-ray scattering (SEC-SAXS) data was collected in order to provide a model of the full complex. Datasets were collected for ComR, ComR:Prx and Prx, each at 9 mg/mL and in the same buffer by SEC coupled SAXS. Figure 12A shows the resulting intensity (I) vs scattering angle (q) of each dataset after peak selection and buffer subtraction. The transformation of this plot to the Guinier plot (ln of the intensity vs scattering angle squared) allowed for the determination of the Rg (radius of gyration), shown in Figure 12B. Each dataset was linear in the low-q region of the Guinier plot, indicating the samples were monodisperse and no significant inter-particle effects were taking place (repulsion or aggregation). The resulting Rg

values of ComR and ComR:Prx agreed with the theoretical molecular weights of 35 kDa and 43 kDa. The Rg of Prx was larger than expected however, correlating to a molecular weight near 12 kDa, rather than 8 kDa. It is unlikely that Prx is dimerizing, as Prx was previously shown to be a monomer in the current buffer by analytical ultracentrufugation[19].

Next, the overall flexibility of each sample was assessed by Kratky and Porod analysis. The dimensionless Kratky plot which is normalized to protein size (Rg), (Figure 12C) shows that each protein is globular in nature, however the ComR:Prx complex was slightly more flexible. A completely globular/folded protein and/or complex would have a distinct peak, where a completely flexible (unfolded) protein would be expected to plateau[61]. The flexibility of each sample was also assessed by Porod analysis, which does not involve normalizing the samples based on the Rg value. This provides a semi-quantifiable value for flexibility, the Porod exponent (P_X), by fitting a linear curve to the Porod region (high q region) of a Porod-Debye plot. These plots are shown in Figure 13B. The resulting Porod exponents of each sample were 3.9 for ComR, 3.7 for ComR:Prx, and 3.6 for Prx. Again, each protein would be considered folded or globular as a completely globular protein would have a P_X of 4.0 and a P_X of 2.0 is indicative of a completely flexible protein.[61]. Although mostly globular, both ComR:Prx and Prx were observed to be more flexible than ComR.

The Porod analysis also allowed for the estimation of the protein volume (using the Porod exponent). The resulting Porod volumes for ComR, and ComR:Prx agreed with the corresponding theoretical molecular weights. The Porod volume for Prx correlates to a molecular weight of 10 kDa, slightly closer to the theoretical molecular weight of 8 kDa, than based off the Guinier analysis. All of the resulting data from the Guinier and Porod analysis can be found in Table 2.

Table 2. Small-angle X-ray scattering data

	ComR	ComR:Prx	Prx
Guinier Analysis[a]			
Rg (Å)	23.59 ± 0.03	27.83 ± 0.04	16.10 ± 0.05
I (0)	0.074 ± 5.1 x10^{-5}	0.085 ± 7.0x10^{-5}	0.020 ± 3.0x10^{-5}
Porod Analysis[b]			
Porod Exponent (P_X)	3.9	3.7	3.6
Porod Volume (Å^3)	61460	76200	17450
P(r) Distribution[a]			
D_{max} (Å)	78.0	98.8	43
Rg (Å)	23.5 ± 0.07	28.3 ± 0.09	15.4 ± 0.02
P(r) quality of fit	0.823	0.772	0.891

[a]Data analysis in PRIMUS [b]Data analysis in ScÅtter IV

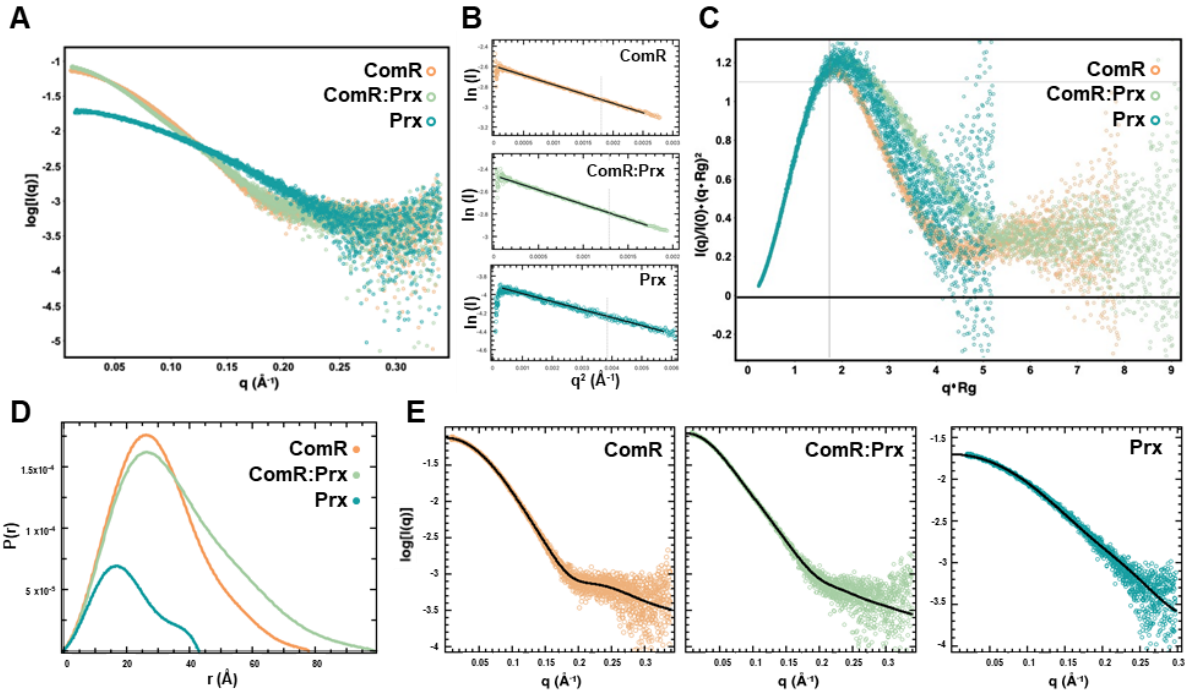

Figure 12. Small angle X-ray scattering data analysis. Scattering data is shown for each ComR (orange), ComR:Prx (light green) and Prx (teal). **A)** A scattering intensity plot of the intensity vs the scattering angle (q). **B)** Guinier fit in the low q region of the scattering intensity plot for each sample. **C)** A dimensionless Kratky plot (normalized to Rg) for each protein, highlighting relative flexibility **D)** The estimated distance distribution plots (P(r)) for each protein. **E)** The fit of the calculated distance distribution refined back to the original scattering data.

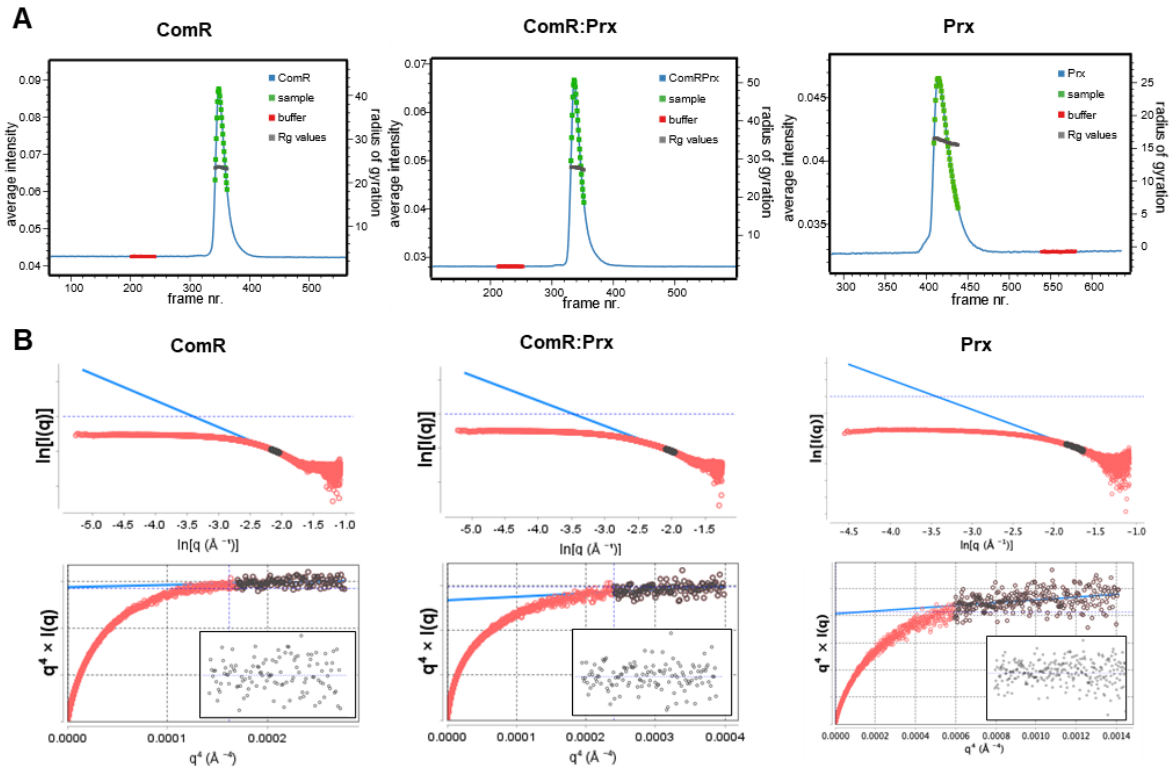

Figure 13. Additional SAXS data analysis **A)** SEC-SAXS elution profile showing the scattering intensities for each sample, along with the background buffer selection (red) and sample selection (green). The Rg values for each corresponding sample point are shown in gray. **B)** Power law plot, along with a Porod-Debye plot, showing the linear curve fit for the determination of the Porod exponent for each protein sample. The inset of each Porod-Debye plot highlights the even distribution of residuals.

3.5.2 Low resolution model of the full ComR:Prx complex

For ComR, ComR:Prx and Prx the real-space electron pair or P(r) distribution was calculated (Figure 12D). The resulting Rg values all agreed with the Guinier derived Rg values, and when refined back to the original scattering data (Figure 12E) had a good quality of fit (over 0.75). All of the relevant data is listed in Table 2. The data was then used for the estimation of low-resolution density maps using DENSS[62]. PDB structures of the DBD and Prx from the complex structure (PDBid: 7N10) and ComR *suis* (PDBid: 5FD4)[12] were aligned into the calculated density maps by hand, in UCSF Chimera[63] (Figure 14). Multiple views of the alignments are shown to allow for better visualization of the fit to the density.

ComR *suis* fits very well into the calculated density map for ComR *mutans*, with a clear larger density corresponding to the TPR domain and smaller section for the DBD. For the ComR:Prx complex, a significant increase in density can be observed. The full-length ComR *suis* protein, although significantly smaller than the overall map, did not align well into the density. No matter how ComR was placed the DBD was angled in such a way it did not fit within the map. This further supports that Prx induces a significant conformational change in ComR. Given this, the TPR domain of ComR *suis* was separately placed into the map, followed by the placement of the DBD:Prx complex.

Additionally, Prx on its own also did not fit well into the calculated density map. Correlating to larger than expected Rg value, the map also had extra density that was unaccounted for. This could be a result of the 6His tag of Prx, which may have significant effect on the folding of Prx as it accounts for approximately 13% of its size. However, it could also be possible that Prx displays an alternate conformation in the current buffer conditions relative to the structure observed in the crystallization conditions.

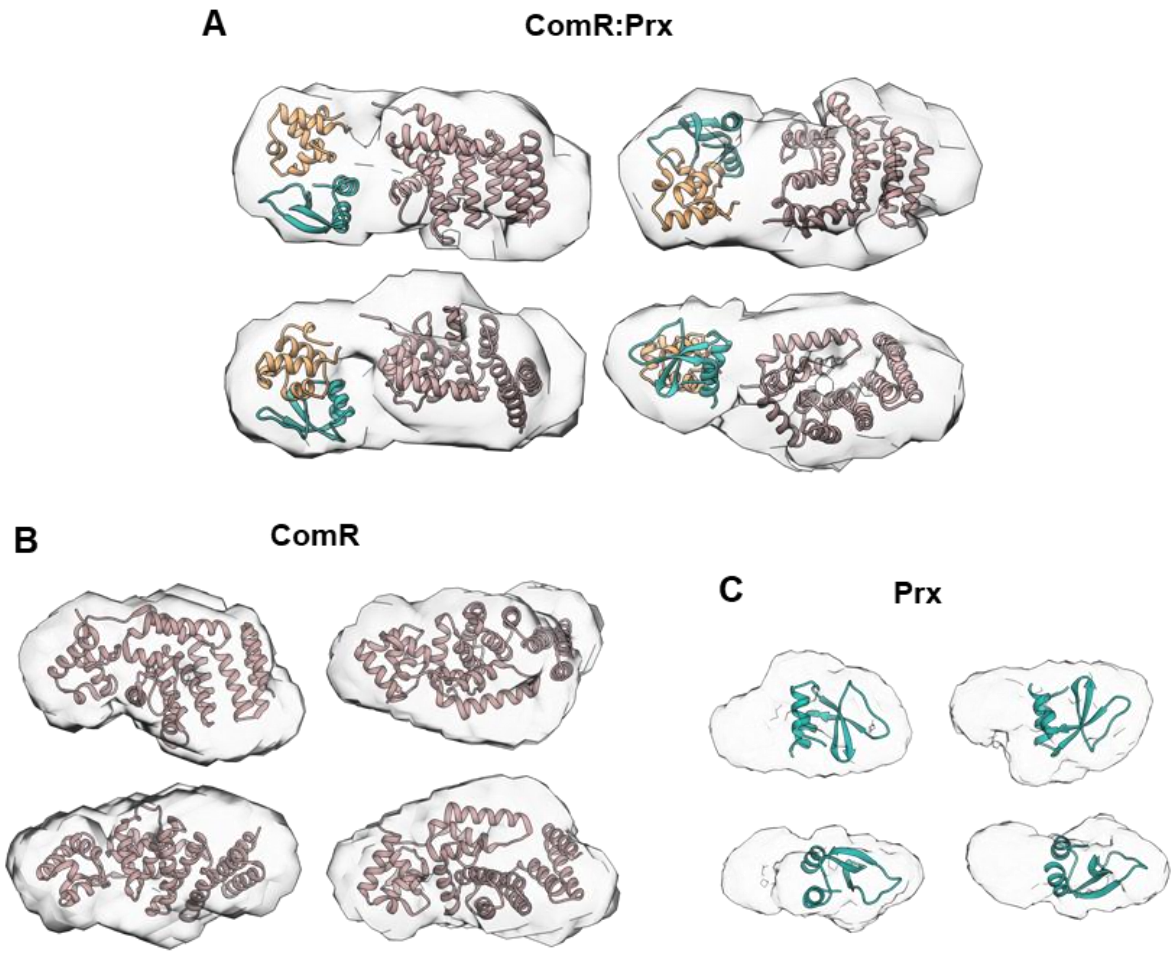

Figure 14. Electron density maps determined by small angle X-ray scattering. A) Low resolution model of the full ComR:Prx complex. The TPR domain of ComR *suis*[12] (5FD4) and the DBD:Prx complex (7N10) is aligned into the DENSS[62] calculated electron density using Chimera[63]. **B)** ComR *suis* (5FD4) aligned into the low resolution density map of ComR *mutans* and **C)** Prx (7N10) aligned into the SAXS determined density map. Multiple views of each alignment are shown (rotated roughly every 90° along the x-axis) to allow for better visualization of the fit of the structures in the density maps.

3.6 Partial X-ray crystal structure of a type II ComR

Because each ComR type is highly variable (only 30-50% identical to one another), and there are currently no existing type II structures (besides the DBD in complex with Prx) it was of interest to obtain a structure of ComR *mutans.* A number of datasets of ComR *mutans* were collected, however the most successful dataset currently is of Se-Met expressed ComR (no His-tag). Crystals were obtained from 1:1 gel filtration buffer and PACT G3 crystallization buffer (0.2 M NaI, 0.1 M Bis-Tris propane pH 7.5, 20% PEG 3350). The crystal diffracted to a resolution of 2.50 Å, and could only be solved to an R_{free} of 28.5. All the relevant data collection and refinement statistics are listed in Table 3. Although the structure (Figure 15A) requires further refinement and building, it serves as a useful model of a type II protein for downstream experiments, such as the design of amino-acid substitutions and targets for FRET labelling.

Table 3. X-ray data collection and refinement statistics for ComR *mutans*

	Se-Met ComR *mutans*
Data Collection	
Wavelength (nm)	0.979590
Space group	C2
Cell dimensions	
a, b, c (Å)	142.0, 52.4, 99.0
α, β, γ(°)	90.0, 128.2, 90.0
Resolution (Å)	50.0 - 2.50
R_{meas} (%)	10.1 (127.3)
CC(1/2)	99.8 (50.9)
$I/\sigma I$	9.94 (1.02)
Completeness (%)	98.6 (95.8)
Redundancy	
Current Refinement	
Resolution (Å)	49.2 - 2.50
Number of monomers in the asymmetric unit	2
R_{work}/R_{free}	24.5/28.5

Values in parentheses represent statistics for the highest resolution shell

3.7 FRET provides additional insight into the ComR:Prx model

To help understand the conformational change taking place when Prx binds to ComR, ComR *mutans* was double-labelled with two separate fluorophores for a fluorescence resonance energy transfer (FRET) experiment. AlexaFluor555, which has a maximum excitation wavelength of 555 nm and emission of 580, was added to the N terminus of ComR using a succinimidyl ester. AlexaFluor647, with a maximum excitation at 650 and emission at 655, was added to residue C199 (which is the only surface-exposed/free cysteine of ComR *mutans*) with a C2 maleimide. The theoretical distance between each dye (based on the partially solved structure of ComR *mutans*) is 35 Å, and is shown in Figure 15B. Based on the Forster distance of these dyes, when labelled ComR is not bound to Prx we would expect to observe some fluorescence in the emission range. When Prx binds to ComR, the fluorescence intensity is expected to either increase or decrease, depending on whether the two fluorophores move closer or further apart.

First, FRET using ComR alone was observed followed by ComR alone in 7 M urea to denature the protein (Figure 16A). ComR at 0.1 µM showed a strong fluorescence emission at 670 nm after excitation at 545 nm, which can be attributed to efficient FRET. When urea was added, the observed fluorescence decreased, indicating that the fluorophores were further apart due to denaturation of ComR. This observable change in FRET suggests that the double labelling of ComR was successful. Next, Prx was added to ComR at increasing concentrations but no significant change in FRET (ratio between fluorescence intensity at 570 and 670 nm) was observed (Figure 16B). However, a proportional increase fluorescence intensity was observed with the addition of Prx. It is already known that Prx binds near the N terminus of the ComR DBD, so it is possible that Prx binding has a direct effect on the fluorescence intensity by changing the environment of the fluorophore conjugated to the N-terminus of ComR.

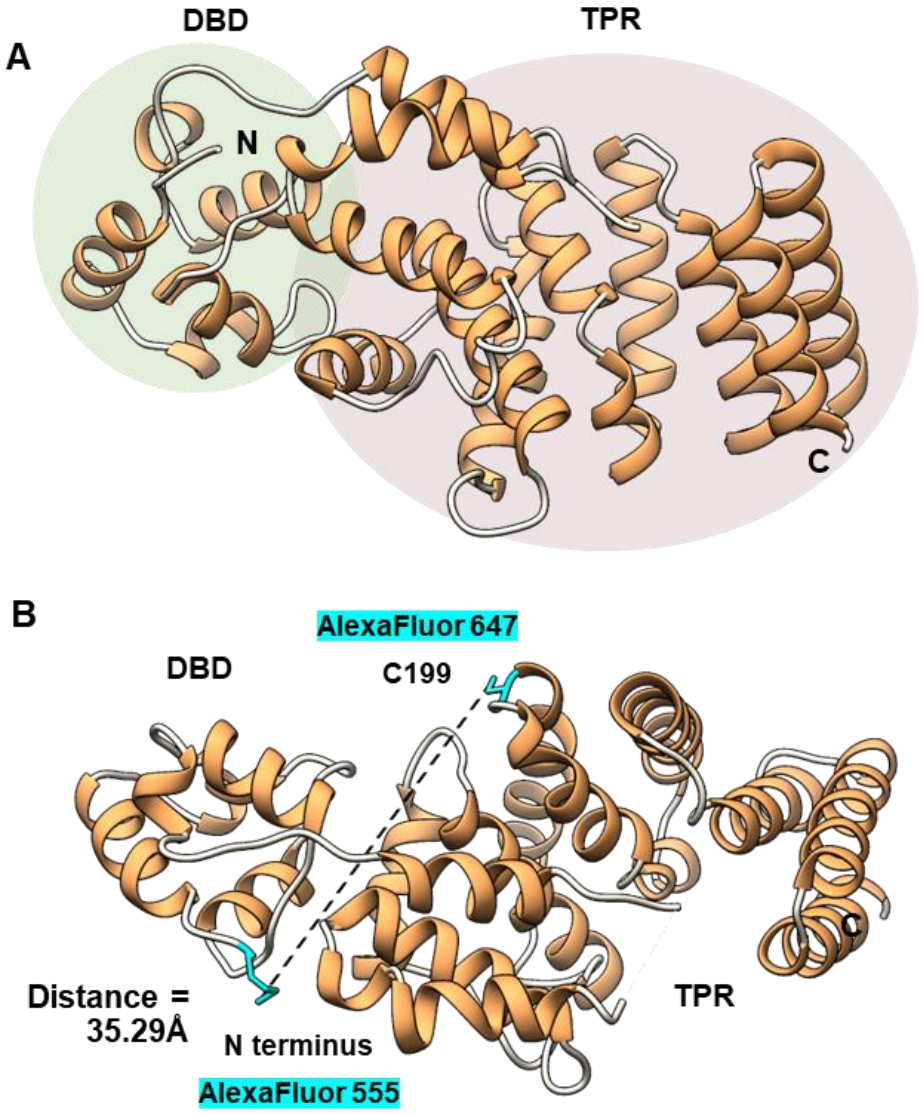

Figure 15. Partially solved X-ray crystal structure of ComR *mutans*. A) Partial structure of ComR *mutans* with the DBD and TPR domains highlighted. **B)** ComR structure with the residues labelled for FRET (N terminus and C199) highlighted. The approximate distance between these two points is 35Å.

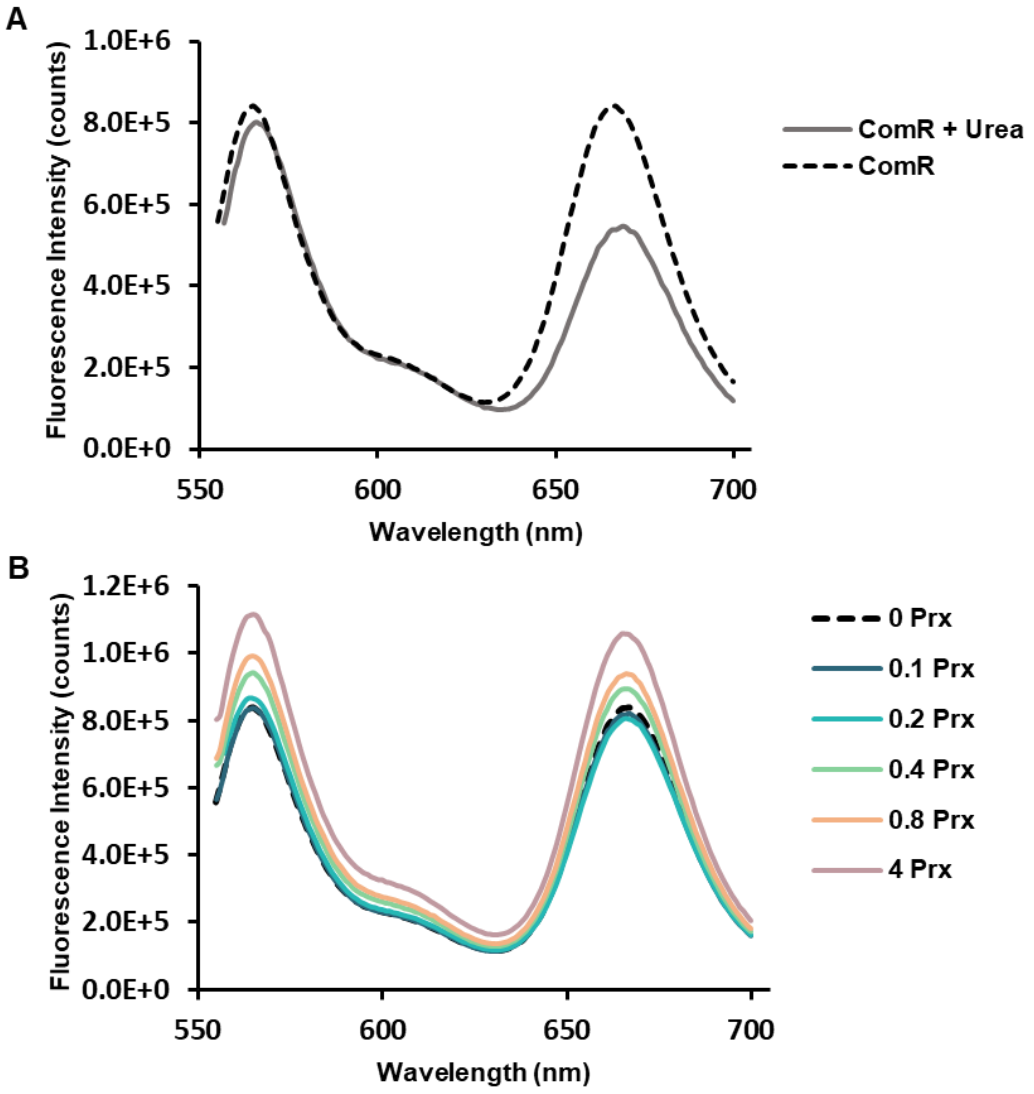

Figure 16. Fluorescence resonance energy transfer between two points within ComR with the addition of Prx. A) Fluorescence intensity from 550 to 700 nm of double-labelled ComR and labelled ComR with 6 M Urea. Less FRET (fluorescence intensity at 670nm) is observed when ComR is unfolded (in 6 M Urea), vs folded. **B)** Fluorescence intensity measured from 550 to 700 nm of double labelled ComR with increasing concentrations of Prx (in µM).

While these results are inconclusive, the DBD could rotate in both the X and Z directions in such a way that there is no significant change in distance between the two fluorophores. Based off the placement of the DBD and ComR *suis* TPR in the SAXS calculated density maps, there is only a change in distance of approximately 5 Å between the two residues. If this experiment is to be repeated, new ComR protein variants would need to be made by removing the free cysteine and introducing a new cysteine at an alternate location for labeling.

3.8 Prx prevents ComR from binding DNA without interfering with XIP recognition

3.8.1 Electro-mobility shift assays

Electro-mobility shift assays (EMSAs) were performed similar to previously described[19], with FAM-labelled *comS* promoter DNA, ComR, XIP, and increasing concentrations of Prx. When ComR and XIP are added to the *comS* promoter region, ComR binds the DNA and a shift can be observed (due to an increase in size) on a native PAGE gel. However, when Prx is added at increasing concentrations Prx inhibits the ability of ComR to bind DNA and no shift is observed (Figure 17).

The first set of experiments (Figure 17A) was performed similarly to published EMSAs with ComR and Prx[19], but also included allowing for the pre-formation of different complexes. ComR, Prx, DNA, and XIP were added in different orders in various experiments to observe if XIP and Prx could out-compete each other. First, ComR, XIP, and Prx were incubated together followed by the addition of DNA (left panels), where a 1:1 molar ratio of Prx appeared to inhibit ComR from binding DNA. Next the ComR:XIP complex was allowed to form, followed by the addition of Prx (middle panels), and the ComR:Prx complex was allowed to form followed by the addition of XIP. Interestingly (although strictly qualitative), Prx appeared to be much better at inhibiting DNA binding when ComR:XIP was incubated together first, rather than when the

ComR:Prx complex formed first. Each of these experiments were also performed with PrxD32A (which is unable to bind ComR based off of SEC and ITC results) as a control.

The second experiment attempted to out-compete Prx with excess XIP (Figure 17B). First ComR:Prx complexes were allowed to form with a sufficient inhibitory concentration of Prx (2 fold molar excess to ComR), followed by the addition of increasing concentrations of XIP, up to an 8 fold molar excess relative to Prx. Even at the highest concentration of XIP relative to Prx, no shift (ComR binding DNA) was observed. This suggests that the mechanism by which Prx inhibits ComR could be unrelated to XIP recognition. Additional experiments (section 3.7.2) were later performed to further test this hypothesis. Controls were also performed for ComR and Prx individually and together, to ensure that no non-specific DNA-binding interactions were taking place (Figure 17C). .

3.8.2 Fluorescence & fluorescence anisotropy competition assays

The XIP competition EMSA (Figure 17B) suggested that the mechanism by which Prx inhibits ComR could be independent of XIP binding. In order to test this hypothesis, fluorescence competition assays were performed using dansyl labelled XIP. In solution dansyl-XIP has a low fluorescence intensity, however the fluorescence intensity significantly increases once the XIP is bound to ComR. This increase in intensity is shown in Figure 18A. Buffer background shows no peak at 535, dansyl-XIP alone has a very slight and shifted peak, and increasing concentrations of ComR results in a significant increase in fluorescence intensity.

The first experiment shown (Figure 18B) is a binding curve of dansyl-XIP with increasing concentrations of ComR. This was done in order to determine the point at which half of the XIP would be bound to ComR (which is the ideal concentration of ComR to be used for the competition assays). The K_d determined from the binding curve was 0.320 µM +/- 0.040, which is a much

higher affinity than the previously reported value of 3.6 μM determined by ITC[19] for the un-labelled XIP. This difference could potentially be attributed to the added hydrophobicity of the dansyl label. Furthermore, the next competition experiment (Figure 18C) with 0.2 μM dansyl-XIP, 0.5 μM ComR, and increasing amounts of un-labelled XIP show that the un-labelled XIP is able to displace dansyl-XIP in a concentration dependent manner, confirming that the ComR:dansyl-XIP interaction is specific and equivalent to XIP. The same competition assay was repeated with increasing concentrations of Prx added, rather than un-labelled XIP. Regardless of the concentration of Prx, there was no decrease in fluorescence intensity meaning that Prx was unable to displace dansyl-XIP from ComR. This shows that Prx either was either unable to bind to the ComR:XIP complex or that Prx could bind ComR:XIP without displacing the XIP and form a ternary complex.

Based on the EMSA results, Prx being unable to interact with ComR:XIP was highly improbable. However, to validate the hypothesis that Prx can bind ComR:XIP complexes, fluorescence anisotropy experiments were performed. The same Prx competition assay was repeated, however the steady-state fluorescence anisotropy was measured instead of the fluorescence intensity (Figure 18D). This allowed for the determination of a change in size of the overall complex, based on the rotational correlation time of the fluorescently labelled peptide[68]. Although Prx is a small protein (8 kDa), accounting for only around 18% of the molecular mass of a possible ternary complex, a noticeable increase in fluorescence anisotropy of the dansyl:XIP was observed in a concentration dependent manner. This indicates Prx is able to bind to the activated ComR:XIP, forming a ternary complex. The resulting dissociation constant was 0.7 +/- 0.3 μM, which is relatively close to the K_d (determined by ITC), for Prx and apo ComR.

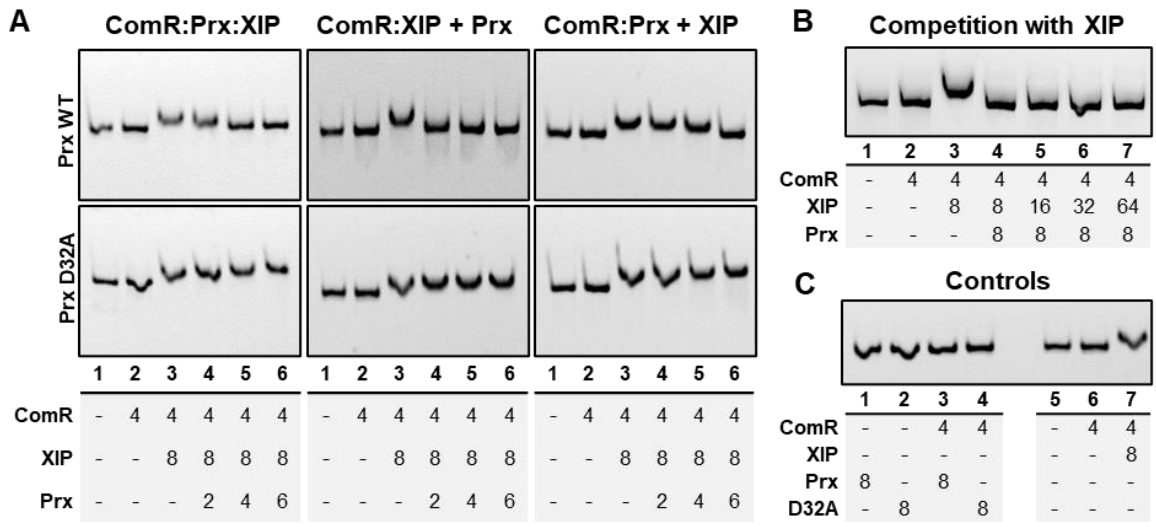

Figure 17. Electromobility shift assays of ComR binding the *comS* promoter. A) EMSAs with ComR, Prx, and XIP incubated together (left), ComR:XIP:DNA complex pre-formation (middle), and ComR:Prx complex pre-formation (right). Experiments were also repeated with PrxD32A (lower gels). **B)** Competition of Prx with increasing concentrations of XIP. **C)** Control samples to ensure there were no non-specific interactions with DNA. For each sample, 100 ng of FAM-labelled p*comS* was added along with the specified amounts of protein or XIP (in μM) below each gel.

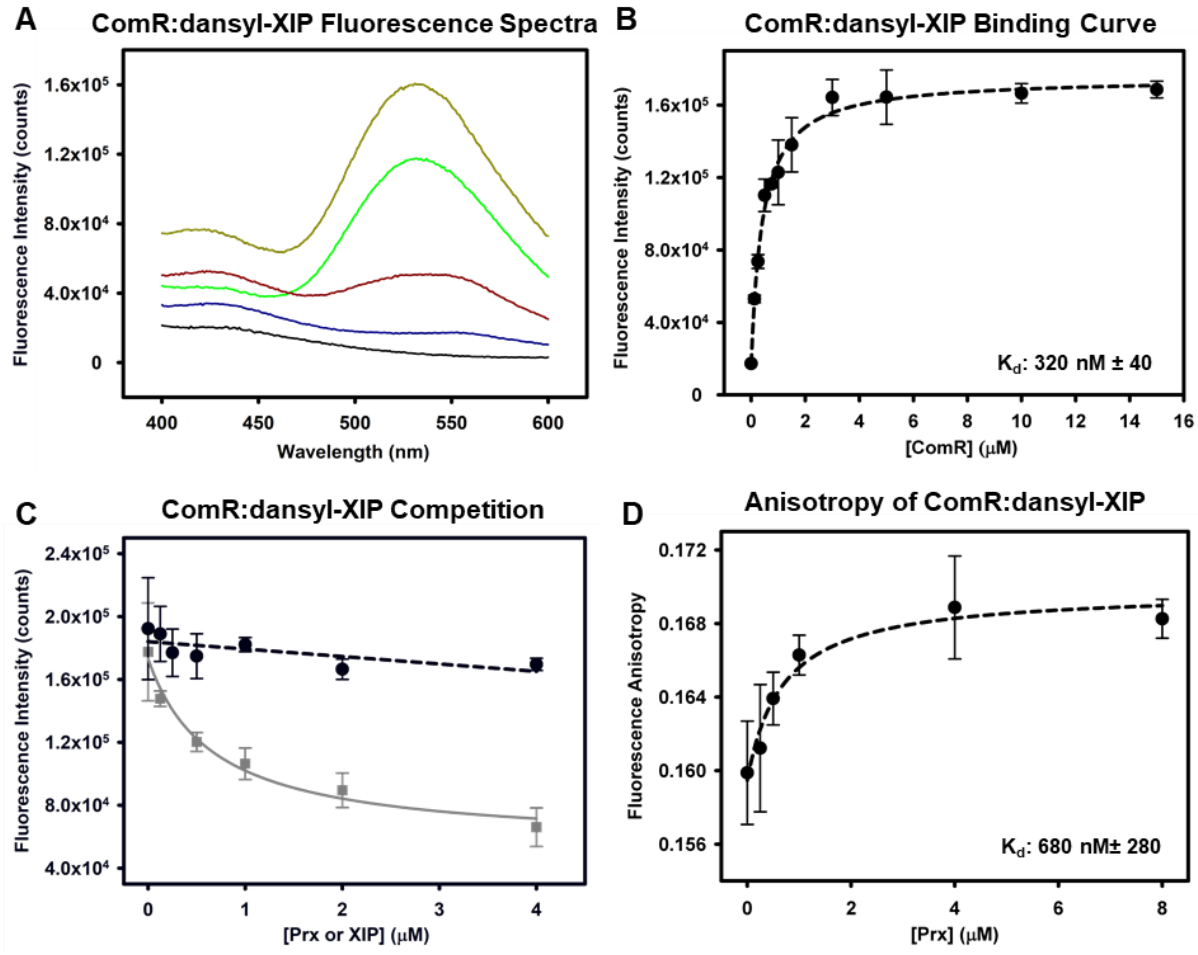

Figure 18. Prx competition fluorescence binding assays. A) Observed increase in dansyl emission as dansyl-XIP bound to ComR. The black line is the buffer background, blue: 0 μM ComR, red: 0.125 μM ComR, green: 0.75 μM ComR, and dark green: 15 μM ComR. **B)** Resulting binding curve of dansyl-XIP at a constant concentration of 0.2 μM with increasing concentrations of ComR ranging from 0.125 to 15 μM. **B)** Competition of dansyl-XIP (0.2 μM) bound to ComR (0.5 μM) with increasing concentrations of non-labelled XIP (grey) or Prx (black). **D)** Fluorescence anisotropy of dansyl-XIP (0.2 μM) bound to ComR (0.5 μM), with the addition of increasing amounts of Prx. Error bars represent replicates, where each experiment was prepared in triplicate.

3.9 Other potential Prx binding partners

Because Prx is genetically linked to a phage encoded toxin gene, and there also exists an example of a similar small phage protein having adapted to perform multiple functions, it was hypothesized that Prx could have a second function within the phage. The first step to finding a new potential biological role for Prx involved searching for possible binding partners. To find new protein interaction partners, a proteomics approach was taken by performing pull-down assays with *S. pyogenes* MGAS315 lysates using Prx (also from MGAS315) as bait. As the goal was to search for any potential binding partner of Prx, multiple experiments were conducted using cell lysates from cells grown in variable conditions. The conditions included: no stimulus, stimulation of cells with XIP to induce the ComRS pathway and subsequently Prx, and stimulation by mitomycin C to induce the prophage[24].

Mass-spectrometry was initially only performed with the XIP induced and mitomycin C induced samples for which a large number of proteins (around 300 to 500) were identified in both the background controls (no Prx added to beads) and Prx pull-down samples. Many of the proteins found in the Prx pull down samples were also in the background controls, rendering them background contaminants. It is common in these experiments to have a number of highly-expressed proteins that bind commercial affinity resins. After filtering the pull-down lists to remove proteins found in both the background control and experiment samples, a list of around 30 potential protein hits was created. This list was further refined to remove known metal-binding proteins, as well as essential and constitutively expressed proteins (for example proteins involved in ribosomal maturation or tRNA synthesis). The final list (Table 4) shows five potential proteins of interest chosen based on their significant presence in the experimental pull-down samples relative to the background samples, as well as their annotated functions. These potential hits

include: the s subunit of a Type 1 endonuclease, a protein of unknown function, an XRE family transcriptional regulator, a putative LexA or CI transcriptional regulator, hyaluronate lyase, and an unannotated phage protein. Future experimentation will involve generating protein expression constructs and determining whether any of these proteins are able to interact with Prx.

Table 4. Top hits from Prx pull-downs with MGAS315 lysates

		Log(E)[a]			
Protein Identifier	**Notes**	**XIP[b]**	**XIP-Prx[c]**	**MitC**	**MitC-Prx**
gi|21910426|	**SpyM3_0890** - Type I restriction endo-nuclease subunit S	/	-81.6	/	-85.3
gi|21909843|	**SpyM3_0307** - hypothetical phage protein (no homology to known structures, and not encoded within a known MGAS315 or MGAS5005 prophage)	/	-10.9	/	-47
gi|21910500|	**SpyM3_1346 -** XRE family transcriptional regulator	/	/	/	-25.8
gi|21910513|	**SpyM3_0964 -** LexA family transcriptional regulator, putative phage CI repressor	/	-63.5	-1.5	-1.6
gi|21910201|	**SpyM3_0665 (*hylA*)** - extracellular hyaluronate lyase	/	-10.9	/	-2.5
gi|21910782|	**SpyM3_1246 –** hypothetical phage protein (absolutely no homology to known structures)	/	/	-1.9	-23.1

[a] The log of the expectation values (E) are displayed, such that a larger negative number corresponds to a lower E value and a more probable protein hit. [b] Background control sample of lysate without Prx added. [c] Pull-down sample with Prx and lysate.

4 DISCUSSION

4.1 Mechanism of natural competence inhibition

The results shown in this thesis comprehensively illustrate the mechanism of natural competence inhibition in *Streptococcus* by the phage protein Prx. Prx can interact with various ComR proteins, including orthologs of each type. Despite this broad range of activity Prx showed a clear preference for type II ComR proteins and was not able to form a complex with the similar transcription factor Rgg3. When assessing the ability of Prx to interact with the individual domains of ComR it was revealed that Prx could interact with the minimal DBD with a higher affinity than the full-length ComR protein. A crystal structure of the ComR DBD in complex with Prx highlights the electrostatic and conserved interaction surface, of which many structural contacts were verified to be biologically relevant. Furthermore, it was determined that Prx functions by directly preventing DNA binding at it has no effect on the ability of ComR to recognize its XIP. Moreover, Prx is capable of binding to ComR in both its inactive or active (XIP-bound) state.

The version of Prx used throughout this thesis is from a strain of *S. pyogenes* (MGAS315), which encodes a type II ComR protein. Therefore, a type II ComR protein was used for the majority of the experiments. Unfortunately, the ComR proteins from MGAS315 and MGAS5005 did not express and purify well, so ComR from *S. mutans* was used. These ComR proteins are approximately 40-50 % identical to one another and the residues shown to be involved in the Prx interaction (R33, Q34, R37, K49) are conserved between ComR *mutans* and the other ComR *S. pyogenes*. Prx interacts with ComR *mutans* with high affinity so it was of interest to see if Prx could interact with the other ComR types, which can share less than 40% amino acid sequence identity. Pull downs as well as gel filtration binding assays showed that Prx was able to interact with type I and III ComRs, yet appeared to show a clear preference for type II ComRs. The type

III ComR (*suis*) and type I (*thermophilus*) both were pulled down by Prx, and showed a slight shift by gel filtration, but not to the same extent as ComR *mutans* and MGAS5005. Although Prx interacts with ComR *thermophilus*, in SEC experiments the complex eluted corresponding to a size much larger than the ComR:Prx complex. This could indicate that the complex is forming a large aggregate. Interestingly, small variations in ComR *suis* and ComR *thermophilus* relative to ComR *mutans* and ComR *pyogenes* could help explain the preference of Prx for type II ComR proteins. All ComR types share most of the important residues with other type II ComR proteins, including residues R33, Q34, and R37. However, ComR *thermophilus* has a larger arginine in place of K49 found in ComR *mutans*,. ComR *suis* has an alanine instead of the Q42 found in other ComR proteins. Q42 also appears to be important for the Prx interaction as it forms hydrogen bonds with Y15 in Prx. These variations, among other differences in the DNA binding surface of the ComR DBD could help explain the observed preference for the type II proteins used.

Overall, the DBD of ComR is highly conserved relative to the TPR domain (Figure 19) and Prx specifically interacts with the conserved residues of the DBD that are essential for recognizing DNA. The increased preference for type II proteins however suggests that this Prx may have specifically evolved to inhibit the ComR type expressed by the phage's host. However, Prx is highly prevalent in different species of *Streptococcus* (including species containing type I and III ComR proteins) and Prx is often over 60-80% identical between the different species. In order to gain a more complete picture of the evolutionary relationship between Prx and ComR, a co-evolutionary analysis with Prx and ComR from the same strains/species would be beneficial, in addition to determining the specificity of various Prx proteins (for example *S. suis*, which contains a type III ComR) for different ComR proteins.

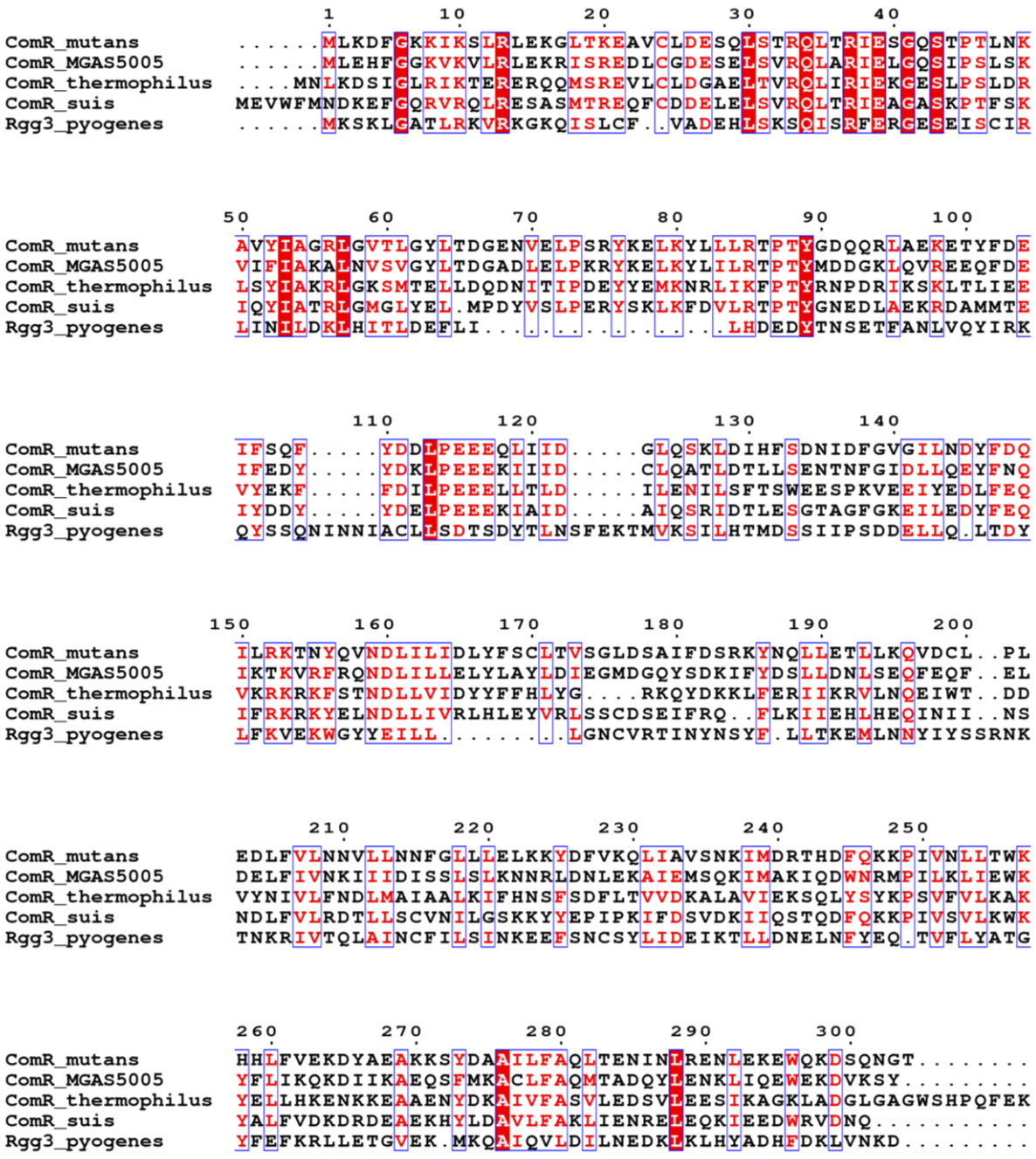

Figure 19. Multiple sequence alignment of various ComR proteins and Rgg3. Proteins were aligned using MAFFT[69] and sequence identities and similarities were visualized using ESPRIPT3[70].

ComR, belongs to the Rgg sub-family of RRNPP transcription factors, which also includes other Rgg proteins known to regulate virulence in *Streptococcus*. It was therefore of interest to see if Prx could also interact with a more distantly related Rgg protein. Rgg3 from *S. pyogenes* shares only 17% sequence identity with ComR *mutans*, however since the DBDs share a similar structure (RMSD of 1.4 $Å^2$) and Prx interacts with conserved and functionally significant residues present in this protein, it was hypothesized that Prx may still be able to bind Rgg3. A gel filtration binding assay with Prx and Rgg3 from *S. pyogenes* showed no shift or complex formation, suggesting that Prx does not bind to Rgg3. Rgg3 does contain a number of residues known to form contacts with Prx including residues corresponding to Q34 and R37, however has a serine instead of the conserved R33 in all assayed ComR proteins. This Rgg protein also has a phenylalanine in place of I38, potentially further interfering with the Prx interaction surface. The inability of Prx to complex with Rgg3 despite its highly conserved interactions, suggests that Prx may have evolved specifically to inhibit ComR, and subsequently prevent natural competence without interfering with the roles of the Rgg proteins.

The interaction of Prx with ComR is not only conserved, but highly electrostatic. Prx has an electronegative surface with many acidic residues, which it uses to interact with the largely electropositive DNA binding surface of the ComR DBD. This is further supported by the ITC data where the enthalpy values for ComR:Prx and DBD:Prx were favorable (largely negative) and the entropy values were slightly unfavorable (negative), indicating that Prx binding to ComR is not primarily driven by hydrophobic interactions. The ITC data also revealed additional insights when comparing the ComR:Prx complex to the DBD:Prx complex. The DBD:Prx complex had a much higher affinity with a binding constant of approximately 50 nm compared to 400 nm, and also a less favorable entropy value relative to the ComR:DBD complex. The slightly more favorable

entropy value of the full ComR:DBD complex could either indicate that more hydrophobic interactions are taking place, or that Prx binding to ComR results in a less restricted (more flexible) conformation relative to Prx binding the DBD[71]. The latter was further supported by X-ray crystallography as well as SAXS. A crystal structure of the DBD:Prx complex revealed that the DBD must be released by the TPR domain in order for Prx to bind. Furthermore, SAXS data showed that ComR:Prx is more flexible than the ComR on its own, and the estimated electron density map had additional density in which the DBD:Prx complex along with the ComR TPR domain could be easily aligned.

The exact orientation of Prx and the DBD relative to the TPR could not be determined based on SAXS and the limited crystallization construct, so FRET was used to provide a better estimation of the conformation of the full ComR:Prx complex. By labelling two residues of ComR, the N terminus and C199 with two different fluorophores, a change in distance or conformation may be observed. However, despite successful labelling, the results were inconclusive with no observable change in distance with the addition of Prx to ComR. While it is possible that Prx could cause a conformational change where the DBD is rotated in such a way that results in no change in distance between the N terminus and C199, this could not be confirmed as Prx had a concentration dependent effect on the donor fluorophore. Future experiments using different labelling sites, may provide valuable information into the complex dynamics and allow for a more accurate alignment of the ComR:Prx complex into the SAXS data.

With the ComR:Prx interaction surface well defined, the next step involved understanding the overall mechanism of Prx in relation to the ComRS pathway. Although Prx bound specifically to the DNA binding residues of ComR, it was not yet clear whether Prx had an effect on the ability of ComR to recognize its signal molecule (XIP). Since ComR is a dynamic protein which

undergoes extensive conformational changes once activated[41], it was also of interest to see if XIP had an effect on the ability of Prx to bind ComR. To assess the relationship between the XIP and Prx, EMSAs were performed similarly to previously[19] with the added parameters of allowing either the ComR:Prx complex to pre-form before the addition of XIP and DNA, or allowing the ComR:XIP:DNA complex to form, followed by the addition of Prx. While strictly qualitative, less Prx was required to prevent ComR from binding DNA when the ComR:XIP:DNA was allowed to incubate first. This could either be attributed to Prx having a higher affinity for the for the activated form of ComR, which has a more accessible DBD, already freed from the TPR, or potentially that Prx could disrupt the activated ComR:XIP dimer. A second EMSA experiment aimed to determine whether the XIP could have an inhibitory effect on the interaction and outcompete Prx. Regardless the concentration of XIP added in excess to Prx, Prx was still able to prevent ComR from binding DNA suggesting that the ComR:Prx interaction is unrelated to the XIP.

To further support the EMSAs, competition assays were performed using a fluorescently labelled XIP (dansyl-XIP). Competition of ComR:dansyl-XIP with increasing amounts of Prx revealed that Prx was unable to displace the XIP, despite its concentration. This meant that Prx was either binding to ComR:XIP forming a ternary complex, or that Prx was unable to bind ComR:XIP. The same competition assay was then repeated looking at the anisotropy of dansyl-XIP rather than fluorescence intensity, which showed that Prx could bind to the activated ComR:XIP and form a ternary complex.

Overall, this mechanism of quorum sensing inhibition is highly efficient. Prx binds ComR with nanomolar affinity and is able to recognize either the apo (inactive) ComR, or XIP-bound (active) ComR. This results in the uncoupling of ComR and its signal molecule, allowing Prx to

inhibit ComR in either state. Furthermore, by not displacing the XIP Prx effectively sequesters XIP, reducing the effective concentration available to interact with other ComR proteins.

This mechanism of quorum sensing inhibition, while structurally unique, is analogous to a recently described mechanism of quorum sensing inhibition in a Gram-negative bacterium[48]. The DMS3 phage of *Pseudomonas aeruginosa* uses a small protein, Aqs1, to inhibit host quorum sensing[48]. Similar to Prx, Aqs1 binds to the DBD of the transcriptional regulator LasR preventing it from binding DNA. Both proteins use electronegative surfaces to interact with conserved and functional DNA binding resides, however they share no structural similarities. Aqs1 consists of two helices and binds its respective DBD as a dimer, and shares no sequence homology with Prx. The DBDs of ComR and LasR while both containing helix-turn-helix domains, also share limited structural homology with a RMSD of 5.1 $Å^2$. A comparison of each complex and their respective interaction surface is displayed in Figure 20. This similar mechanism of quorum sensing inhibition, mediated by completely different proteins demonstrates an example of convergent evolution between a Gram-positive and Gram-negative bacteriophage. Furthermore, both these phages inhibit a host quorum sensing pathway as a means of self-preservation. By inhibiting LasR, the DMS3 phage prevents the expression of downstream quorum sensing systems, which are thought to play a role in resistance to phage infection[48,72]. Prx in turn inhibits natural competence expression, which could ultimately prevent the loss or damage the prophage due to homologous recombination with foreign DNA.

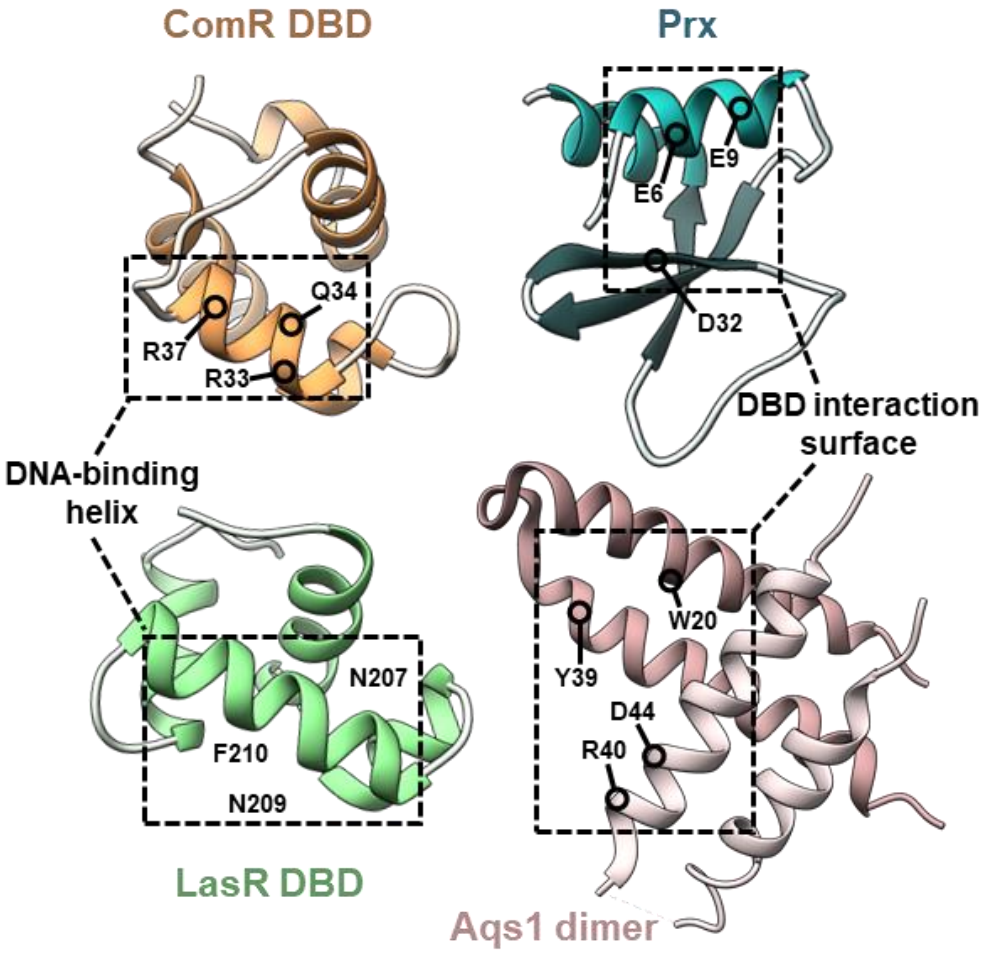

Figure 20. Comparison of the ComR:Prx and LasR:Aqs1 interactions. On the left the DNA binding, as well as phage protein interaction surfaces of the ComR DBD and LasR DBD (PDBid: 6V7W) are shown, with residues that contact the respective phage protein highlighted. On the right, Prx and the Aqs1 dimer (6V7W) are shown with the DBD binding surfaces, along with a few important residues highlighted.

4.2 Other potential roles of Prx

It is of interest to further study Prx in its role in relation to the biology of the phage for several reasons. First, Prx has a novel fold[19] and appears to adopt a different conformation in the working buffer solutions relative to the crystal structure. Based on the SAXS data, Prx has an Rg corresponding to a molecular weight larger than it is, is slightly flexible, and the calculated electron density map contained excess density into which the crystal structure could not be placed. Interestingly however, the crystal structure of Prx alone (6CKA[19]) was nearly identical to Prx when in complex with the ComR DBD, suggesting that the crystal structure is at least in part biologically relevant. By better understanding the structure of Prx in solution, it may draw similarities between other known protein structures and provide insight towards any potential additional functions.

There are also a number of details which help contribute towards the hypothesis that Prx has more than one role, the primary being the linkage of Prx to the prophage encoded toxins[45]. Prx also shares distant structural similarities with the lambda phage GpW protein[47]. Finally, there are numerous examples of bacteriophage economizing their limited genomes by encoding proteins evolved to perform multiple functions. For example, the T7 lysozyme also acts as a regulator by binding the T7 RNA polymerase and inhibiting further transcription[73]. Even Aqs1 which inhibits LasR, also binds PilB using a separate surface and inhibits type IV pilus assembly[48,74].

The first step towards determining any additional functions of Prx involved the use of pull-down assays to search for potential protein binding partners. Pull-down experiments with *S. pyogenes* MGAS315 lysates revealed a number of interesting hits. However, due to the abundance of identified peptides present in the background control samples it is difficult to initially infer any clear binding partners. If repeated, future experiments could involve even more stringent wash steps with higher concentrations of imidazole in the wash buffer. There were a significant amount

of metal-binding proteins detected (which could have the potential to bind to the Ni-NTA resin) so an additional option could involve the use of glutathione resin and GST-Prx as bait. However, this option was avoided due to the size of GST relative to Prx and its potential to sterically hinder potential interactions (as observed with ComR). Despite the significant background contamination, there were a few proteins that were identified as potential binding partners. These proteins were present in one or two of the experimental pull-down samples and neither of the background controls, not known to bind to metals, are not normally constitutively expressed in high abundance, and have interesting roles potentially having some relation to the bacteriophage. These 5 hits are worth designing protein expression constructs for to assess whether or not they are capable of interacting with Prx.

Of these five potential Prx binding partners, the most interesting hit was the putative s subunit of a Type I restriction enzyme. Type I restriction enzymes consist of two R (restriction endonuclease) subunits, two M (methylation) subunits, and a single S (DNA recognition) subunit, and unlike Type II restriction enzymes, recognize separated DNA sequences and translocate DNA, cleaving at a random site far from the recognition sequences[75,76]. These enzymes cleave non-methylated DNA protecting the bacterium from bacteriophage and other foreign genetic elements[75]. There already exists a few examples of bacteriophage having evolved mechanisms to avoid restriction. The T7 phage uses Ocr, which binds and inhibits the R and M subunits by mimicking DNA[77]. Another mechanism employed by the lambda phage involves the RalI protein, which is thought to bind the M subunit causing a conformational change leading to methylation rather than restriction of its DNA[78]. Currently there are no known examples of bacteriophage inhibitors solely targeting the S subunit.

Another protein of interest is a completely uncharacterized protein hypothesized to be "phage protein" based off the annotations of nearly identical sequences on BLAST. However, this gene was not located within a known prophage region in MGAS315. When searching for this gene in another strain of *S. pyogenes*, MGAS5005, there were two hits with only 26-29% sequence identity M5005_Spy0354 and M5005_Spy0357. These genes were also not found within a known prophage, however they were found on either side of an *speJ* exotoxin gene. A Phyre2 search with this hypothetical protein turned up no potential structural homologues. Other protein hits include a hypothetical XRE family transcriptional regulator, and a putative LexA-like, CI phage repressor, both found in separate MGAS315 prophages. These hypothetical transcriptional regulators share 100 % amino acid sequence identity to one another. Two other interesting hits are HylA, an extracellular hyaluronate lyase and another unannotated phage protein SpyM3_1246. This unannotated phage protein is found within a MGAS315 prophage, however a Phyre2 search revealed no obvious structural homologues.

These five potential hits provide a good starting point for assaying potential Prx binding partners. The next steps involve cloning these genes into protein expression constructs, purification, and assessing for binding to Prx either via pull-down or SEC binding assays. In the event that none of these proteins interact with Prx, MS can be performed with the media pull-down samples (already digested and processed for MS), or pull-down assays can be re-attempted using a less harsh method of lysis compared to sonication, or a different buffer condition containing mild detergents to aid in the solubility of different proteins.

In order to acquire a complete picture of the role of Prx in relation to the phage, proteomics could highlight the changes in global protein expression due to the presence of Prx. The proposed experiments involve comparing the induction of *S. pyogenes* with XIP, mitomycin C, and no

induction of a wild type strain (MGAS5005) to a mutant strain lacking a *prx* (MGAS5005Δ*prx3*). These six conditions could then be quantitatively compared by tandem-mass-tag mass spectrometry. However, the growth and induction of these strains were highly variable and inconsistent and would therefore require substantial optimization before proceeding with quantitative mass spectrometry.

5 CONCLUSIONS & FUTURE DIRECTIONS

The primary focus of this thesis involved describing the mechanism of natural competence inhibition in *Streptococcus* by the bacteriophage protein Prx using a variety of orthologous biochemical techniques. Both the protein-protein interaction surface of Prx with ComR, as well as the overall mechanism of quorum sensing inhibition by Prx were described in detail. Prx inhibits the natural competence pathway by binding to the DBD of ComR, the key transcription factor of the pathway, and specifically recognizes conserved and functional residues involved in recognizing DNA. Prx functions by directly preventing ComR from binding DNA without effecting XIP recognition by ComR, and is able to recognize and inhibit both the inactive ComR monomer as well as the activated XIP bound ComR.

By inhibiting natural competence expression, Prx is ultimately preventing the uptake of foreign DNA. This could serve to protect the bacteriophage in a number of ways. Inhibiting DNA uptake could prevent the streptococcal host from acquiring new CRISPR cassettes and associated genes, or more importantly prevent the loss or damage of the prophage due to homologous recombination. This mechanism while structurally unique, is similar to a mechanism of quorum sensing inhibition mediated by the small DMS3 phage protein Aqs1 in *P. aeruginosa*[48]. Prx and Aqs1 share no structural homology, however both proteins have evolved to bind and inhibit the DBDs of their respective transcription factors. This provides an example of convergent evolution between a Gram-positive and Gram-negative bacteriophage modulating their host pathways for the benefit of self.

Although this mechanism is well characterized, future work could assess the specificity of Prx proteins from different species of *Streptococcus* with various ComR proteins in order to better understand its evolution. Also, since the structure of the full complex remains ambiguous

crystallization could be attempted with other ComR and Prx orthologs. Additional FRET assays could also be performed to better understand the interaction dynamics between ComR, Prx, and XIP.

The primary focus of any future experimentation would involve the investigation of any other potential roles of Prx in the biology of the phage. Pull-down assays revealed a few interesting protein hits for which expression constructs need to be designed and proteins successfully purified to assess their ability to bind Prx *in vitro*. Finally, understanding the role of Prx also involves understanding its global effect on the proteome. Quantitative proteomics of a wild-type strain of *S. pyogenes* compared to a strain lacking *prx* induced under a variety of conditions, will help reveal any downstream effects of Prx on protein expression and overall life-cycle of the bacterial cell and the bacteriophage.

6 REFERENCES

1. Lancefield, R. C. A serological differentiation of human and other groups of hemolytic Streptococci. *J. Exp. Med.* **57**, 571–95 (1933).

2. Van Sorge, N. M. *et al.* The classical lancefield antigen of group A Streptococcus is a virulence determinant with implications for vaccine design. *Cell Host Microbe* **15**, 729–740 (2014).

3. World Health Organization. The Current Evidence for the Burden of Group A Streptococcal Diseases Department of Child and Adolescent Health and Development World Health Organization. *Dep. Child Adolesc. Heal. Dev.* (2005).

4. Ralph, A. P. & Carapetis, J. R. Group A Streptococcal Diseases and Their Global Burden. *Curr. Top. Microbiol. Immunol.* **368**, (2013).

5. Fischetti, V. A. Streptococcal M protein: molecular design and biological behavior. *Clin. Microbiol. Rev.* **2**, 285–314 (1989).

6. Walker, M. J. *et al.* Disease manifestations and pathogenic mechanisms of Group A Streptococcus. *Clin. Microbiol. Rev.* **27**, 264–301 (2014).

7. Finley, C. R. *et al.* What are the most common conditions in primary care? Systematic review. *Can. Fam. Physician* **64**, 832–840 (2018).

8. Pfoh, E., Wessels, M. R., Goldmann, D. & Lee, G. M. Burden and economic cost of group A streptococcal pharyngitis. *Pediatrics* **121**, 229–34 (2008).

9. Centers for Disease Control and Prevention. Antibiotic resistance threats in the United States. 1–113 (2019).

10. Beres, S. B. *et al.* Genome sequence of a serotype M3 strain of group A Streptococcus: Phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10078–10083 (2002).

11. Sumby, P. *et al.* Evolutionary origin and emergence of a highly successful clone of serotype M1 group a Streptococcus involved multiple horizontal gene transfer events. *J. Infect. Dis.* **192**, 771–82 (2005).

12. Shanker, E. *et al.* Pheromone Recognition and Selectivity by ComR Proteins among Streptococcus Species. *PLoS Pathog.* **12**, 1–32 (2016).

13. McShan, W. M., McCullor, K. A. & Nguyen, S. V. The Bacteriophages of Streptococcus pyogenes . *Gram-Positive Pathog.* 158–176 (2019). doi:10.1128/9781683670131.ch11

14. Johnston, C., Martin, B., Fichant, G., Polard, P. & Claverys, J. P. Bacterial transformation: Distribution, shared mechanisms and divergent control. *Nat. Rev. Microbiol.* **12**, 181–196 (2014).

15. Leonard, C. G., Colón, A. E. & Cole, R. M. Transduction in Group A streptococcus. *Biochem. Biophys. Res. Commun.* **30**, 130–135 (1968).

16. Penadés, J. R., Chen, J., Quiles-Puchalt, N., Carpena, N. & Novick, R. P. Bacteriophage-mediated spread of bacterial virulence genes. *Curr. Opin. Microbiol.* **23**, 171–178 (2015).

17. Dubnau, D. & Blokesch, M. Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annu. Rev. Genet.* **53**, 217–237 (2019).

18. Marks, L. R., Mashburn-Warren, L., Federle, M. J. & Hakansson, A. P. Streptococcus pyogenes biofilm growth in vitro and in vivo and its role in colonization, virulence, and genetic exchange. *J. Infect. Dis.* **210**, 25–34 (2014).

19. Mashburn-Warren, L., Goodman, S. D., Federle, M. J. & Prehna, G. The conserved mosaic prophage protein paratox inhibits the natural competence regulator ComR in Streptococcus. *Sci. Rep.* **8**, 1–15 (2018).

20. Thierauf, A., Perez, G. & Maloy, S. Generalized transduction. *Methods Mol. Biol.* **501**, 267–86 (2009).

21. Chiang, Y. N., Penadés, J. R. & Chen, J. Genetic transduction by phages and chromosomal islands: The new and noncanonical. *PLoS Pathog.* **15**, 1–7 (2019).

22. Hyder, S. L. & Streitfeld, M. M. Transfer of erythromycin resistance from clinically isolated lysogenic strains of Streptococcus pyogenes via their endogenous phage. *J. Infect. Dis.* **138**, 281–6 (1978).

23. McCullor, K., Postoak, B., Rahman, M., King, C. & McShan, W. M. Genomic Sequencing of High-Efficiency Transducing Streptococcal Bacteriophage A25: Consequences of Escape from Lysogeny. *J. Bacteriol.* **200**, (2018).

24. Euler, C. W., Juncosa, B., Ryan, P. A. & Deutsch, D. R. Targeted Curing of All Lysogenic Bacteriophage from Streptococcus pyogenes Using a Novel Counter-selection Technique. *PLoS One* 1–22 (2016). doi:10.1371/journal.pone.0146408

25. Nguyen, S. V & McShan, W. M. Chromosomal islands of Streptococcus pyogenes and related streptococci: molecular switches for survival and virulence. *Front. Cell. Infect. Microbiol.* **4**, 109 (2014).

26. Campbell, A. M. Chromosomal insertion sites for phages and plasmids. *J. Bacteriol.* **174**, 7495–9 (1992).

27. Desiere, F., McShan, W. M., van Sinderen, D., Ferretti, J. J. & Brüssow, H. Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic Streptococci: evolutionary implications for prophage-host interactions. *Virology* **288**, 325–41 (2001).

28. Shannon, B. A., McCormick, J. K. & Schlievert, P. M. Toxins and Superantigens of Group A Streptococci. *Microbiol. Spectr.* **7**, (2019).

29. Moon, A. F., Krahn, J. M., Lu, X., Cuneo, M. J. & Pedersen, L. C. Structural characterization of the virulence factor Sda1 nuclease from Streptococcus pyogenes. *Nucleic Acids Res.* **44**, 3946–3957 (2016).

30. Smoot, L. M. *et al.* Characterization of two novel pyrogenic toxin superantigens made by

an acute rheumatic fever clone of Streptococcus pyogenes associated with multiple disease outbreaks. *Infect. Immun.* **70**, 7095–104 (2002).

31. Papageorgiou, A. C. *et al.* Structural basis for the recognition of superantigen streptococcal pyrogenic exotoxin A (SpeA1) by MHC class II molecules and T-cell receptors. *EMBO J.* **18**, 9–21 (1999).

32. Sitkiewicz, I. *et al.* Emergence of a bacterial clone with enhanced virulence by acquisition of a phage encoding a secreted phospholipase A2. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16009–14 (2006).

33. Brenciani, A. *et al.* Phim46.1, the main Streptococcus pyogenes element carrying mef(A) and tet(O) genes. *Antimicrob. Agents Chemother.* **54**, 221–9 (2010).

34. Banks, D. J., Lei, B. & Musser, J. M. Prophage induction and expression of prophage-encoded virulence factors in group A Streptococcus serotype M3 strain MGAS315. *Infect. Immun.* **71**, 7079–86 (2003).

35. Bergé, M., Mortier-Barrière, I., Martin, B. & Claverys, J.-P. Transformation of Streptococcus pneumoniae relies on DprA- and RecA-dependent protection of incoming DNA single strands. *Mol. Microbiol.* **50**, 527–36 (2003).

36. Fontaine, L. *et al.* A novel pheromone quorum-sensing system controls the development of natural competence in Streptococcus thermophilus and Streptococcus salivarius. *J. Bacteriol.* **192**, 1444–1454 (2010).

37. Cheng, Q., Campbell, E. A. & Naughton, A. M. The com locus controls genetic transformation in Streptococcus pneumoniae. *Mol. Microbiol.* **23**, 683–692 (1997).

38. Fontaine, L., Wahl, A., Fléchard, M., Mignolet, J. & Hols, P. Regulation of competence for natural transformation in streptococci. *Infect. Genet. Evol.* **33**, 343–360 (2015).

39. Mashburn-Warren, L., Morrison, D. A. & Federle, M. J. The cryptic competence pathway in Streptococcus pyogenes is controlled by a peptide pheromone. *J. Bacteriol.* **194**, 4589–4600 (2012).

40. Mashburn-Warren, L., Morrison, D. A. & Federle, M. J. A novel double-tryptophan peptide pheromone controls competence in Streptococcus spp. via an Rgg regulator. *Mol. Microbiol.* **78**, 589–606 (2010).

41. Talagas, A. *et al.* Structural Insights into Streptococcal Competence Regulation by the Cell-to-Cell Communication System ComRS. *PLoS Pathog.* **12**, 1–26 (2016).

42. Campbell, E. A., Choi, S. Y. & Masure, H. R. A competence regulon in Streptococcus pneumoniae revealed by genomic analysis. *Mol. Microbiol.* **27**, 929–939 (1998).

43. Neiditch, M. B., Capodagli, G. C., Prehna, G. & Federle, M. J. Genetic and Structural Analyses of RRNPP Intercellular Peptide Signaling of Gram-Positive Bacteria. *Annu. Rev. Genet.* **51**, 311–333 (2017).

44. Capodagli, G. C. *et al.* Structure-function studies of Rgg binding to pheromones and target promoters reveal a model of transcription factor interplay. *Proc. Natl. Acad. Sci. U. S. A.*

117, 24494–24502 (2020).

45. Aziz, R. K. *et al.* Mosaic prophages with horizontally acquired genes account for the emergence and diversification of the globally disseminated M1t1 clone of Streptococcus pyogenes. *J. Bacteriol.* **187**, 3311–3318 (2005).

46. Holm, L. Using Dali for Protein Structure Comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020).

47. Hauser, R. *et al.* Bacteriophage protein-protein interactions. *Adv Virus Res* **83**, 219–298 (2012).

48. Shah, M. *et al.* A phage-encoded anti-activator inhibits quorum sensing in Pseudomonas aeruginosa. *Mol. Cell* **81**, 571-583.e6 (2021).

49. Zhang, X. & Studier, F. W. Multiple roles of T7 RNA polymerase and T7 lysozyme during bacteriophage T7 infection. *J. Mol. Biol.* **340**, 707–30 (2004).

50. van de Rijn, I. & Kessler, R. E. Growth characteristics of group A streptococci in a new chemically defined medium. *Infect. Immun.* **27**, 444–8 (1980).

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

52. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).

53. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).

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

55. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).

56. Meier-Stephenson, V. *et al.* Identification and characterization of a G-quadruplex structure in the pre-core promoter region of hepatitis B virus covalently closed circular DNA. *J. Biol. Chem.* **296**, 1–16 (2021).

57. Manalastas-Cantos, K. *et al.* ATSAS 3.0 : expanded functionality and new tools for small-angle scattering data analysis . *J. Appl. Crystallogr.* **54**, 343–355 (2021).

58. Panjkovich, A. & Svergun, D. I. CHROMIXS: Automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34**, 1944–1946 (2018).

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

60. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).

61. Rambo, R. P. & Tainer, J. A. Characterizing flexible and intrinsically unstructured

biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559–571 (2011).

62. Grant, T. D. Ab initio electron density determination directly from solution scattering data. *Nat. Methods* **15**, 191–193 (2018).

63. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

64. Lasarre, B., Aggarwal, C. & Federle, M. J. Antagonistic Rgg regulators mediate quorum sensing via competitive DNA binding in Streptococcus pyogenes. *MBio* **3**, (2013).

65. Fenyö, D. & Beavis, R. C. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal. Chem.* **75**, 768–74 (2003).

66. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–10 (1990).

67. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–58 (2015).

68. James, N. G. & Jameson, D. M. Steady-state fluorescence polarization/anisotropy for the study of protein interactions. *Methods Mol. Biol.* **1076**, 29–42 (2014).

69. Rozewicki, J., Li, S., Amada, K. M., Standley, D. M. & Katoh, K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res.* (2019). doi:10.1093/nar/gkz342

70. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).

71. Brown, A. Analysis of Cooperativity by Isothermal Titration Calorimetry. *Int. J. Mol. Sci.* **10**, 3457–3477 (2009).

72. Moreau, P., Diggle, S. P. & Friman, V.-P. Bacterial cell-to-cell signaling promotes the evolution of resistance to parasitic bacteriophages. *Ecol. Evol.* **7**, 1936–1941 (2017).

73. Stano, N. M. & Patel, S. S. T7 Lysozyme Represses T7 RNA Polymerase Transcription by Destabilizing the Open Complex during Initiation. *J. Biol. Chem.* **279**, 16136–16143 (2004).

74. Chung, I. Y., Jang, H. J., Bae, H. W. & Cho, Y. H. A phage protein that inhibits the bacterial ATPase required for type IV pilus assembly. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11503–11508 (2014).

75. Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A. & Wilson, G. G. Type I restriction enzymes and their relatives. *Nucleic Acids Res.* **42**, 20–44 (2014).

76. Studier, F. W. & Bandyopadhyay, P. K. Model for how type I restriction enzymes select cleavage sites in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 4677–81 (1988).

77. Kennaway, C. K. *et al.* The structure of M.EcoKI Type I DNA methyltransferase with a

DNA mimic antirestriction protein. *Nucleic Acids Res.* **37**, 762–70 (2009).

78. Loenen, W. A. & Murray, N. E. Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. *J. Mol. Biol.* **190**, 11–22 (1986).

7 APPENDIX

Table A1. Bacterial strains used in this study

Genus & species	Strain	Used for	Source
Escherichia coli	Bl21-Gold (DE3)	Protein expression, cloning	Agilent Techn-ologies
Streptococcus mutans	UA159	Generating FAM-labelled *comS* promoter	[19]
Streptococcus pyogenes	MGAS315	Prx pull downs with cell lysates	[39]
Streptococcus pyogenes	MGAS5005	Proteomics	[19]
Streptococcus pyogenes	MGAS5005Δ*prx3*	Proteomics	[19]

Table A2. Primers used in this study

Primer	Sequence (5'-3')	Uses	Source
MGAS5005_*rimP*_F	ATCTCAGGCGTTACTGTCTTGG	MGAS5005	This
MGAS5005_*prx3*_R	CGAATTTAAGCAAGCAATCGACC	strain	study
MGAS5005_*sdn1*_R	AGCTAGAATGGTCAGAGCAGAG	verification	
UA159_P*comS*_F	AAGCAGGTAGACTGCCTTCCATTGG	FAM-p*comS*	19
FAM-UA159_P*comS*_R	5'FAM-CCTGTTATTCTCCTTTCTTT	for EMSAs	

Table A3. Original protein expression constructs used in this study

Construct Name	Strain and species	Expression Vector	Notes	Source
ComR *mutans*	*Streptococcus mutans*	pET15b	N-terminal, thrombin removable 6-His tag, cloned into NdeI and BamHI sites	[19]
ComR MGAS5005	*Streptococcus pyogenes* MGAS5005	pET15b	N-terminal, thrombin removable 6-His tag, cloned into NdeI and BamHI sites	Federle lab
ComR MGAS315	*S. pyogenes* MGAS315	pET15b	N-terminal, thrombin removable 6-His tag, cloned into NdeI and BamHI sites	Federle lab
ComR *suis*	*Streptococcus suis*	pET15b	N-terminal, thrombin removable 6-His tag	[12]
ComR *thermophilus*	*Streptococcus thermophilus*	pET21A	C-terminal HRV-3C removable 6-His tag, codon optimized	This study
Prx	*S. pyogenes* MGAS315	pET21A	C-terminal, non-removable 6-His tag	[19]
GST-Prx	*S. pyogenes* MGAS315	pGEX6p1	N-terminal HRV-3C removable GST tag, codon optimized, cloned into BamHI and XhoI sites	This study
6His-SUMO-Rgg3	*S. pyogenes* NZ131	pET15b	N-terminal removable SUMO tag with N terminal 6-His tag	Federle lab

Table A4. Primers for the generation of protein expression constructs

Expression Construct	Original construct	F Primer (5' to 3')	R Primer (5' to 3')	Notes
TPR	ComR *mutans*	ATTACATATGAGTCGTTATAAAGAACTAAAGTATTTATTATTACG	ATTAGGATCCTTATGTCCCGTTCT	S74 to T304
DBD	ComR *mutans*	ACATTTTCTCAGTCCGTTAAATAGC	TGAGTTACCCAGTCGTTATAAAG	M1 to D66
PrxD5A	Prx	TTATATATAGCTGAGTTTAAAGAAGCG	CATATGTATATCTCCTTCTTAAAG	
PrxE6A	Prx	TATATAGATGCGTTTAAAGAAGCGATTG	TAACATATGTATATCTTCTTCTTAAAG	
PrxK8A	Prx	AGATGAGTTTGCAGAAGCGATTG	ATATATAACATATGTATATCTCCTTC	
PrxE9A	Prx	GAGTTTAAAGCAGCGATTGATAAGG	ATCTATATATAACATATGTATATCTCC	
PrxD12A	Prx	ATACCCCTTAGCAATCGCTTC	ATTTTAGGTGACACAGTAGC	
PrxT20A	Prx	TTTAGGTGACGCAGTAGCGATAG	ATATACCCCTTATCAATCGCTTC	
PrxF31A	Prx	CGGAAAGATAGCTGATTATGTGTTACCAC	TTTTTACGCACTATCGCTGTG	
PrxD32A	Prx	AAGATATTTGCTTATGTGTTACCAC	TCCGTTTTTACGCACTATC	
PrxY33F	Prx	ATATTTGATTTTGTGTTACCACAC	CTTTCCGTTTTTACGCAC	
PrxH37A	Prx	TGTGTTACCAGCCGAAAAAGTGAGAG	TAATCAAATATCTTTCCGTTTTTACGC	
PrxV48A	Prx	GTTGTGACGGCAGAGAGAGTGG	TTCATCATCTCTCACTTTTTCGTG	