Probing biochemical interactions between the phage protein paratox and its binding partners in *Streptococcus pyogenes*

By:

Tasneem Hassan Muna

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 over half a million deaths every year by causing diseases like necrotizing fasciitis and rheumatic heart disease. Streptococcal toxins that cause such diseases are encoded by prophages found in the *S. pyogenes* genome. The prophage encoded toxin genes are always found adjacent to a highly conserved gene termed paratox (*prx*). Past work by our lab discovered that Prx disrupts quorum sensing and natural competence in *S. pyogenes* by directly binding to the DNA binding domain (DBD) of quorum sensing receptor ComR. In doing so, Prx protects prophage DNA from being replaced by potential homologous recombination. To probe for additional Prx functions, we aimed to explore streptococcal biochemical pathways that may be regulated by Prx. Using a pull-down assay combined with mass-spectrometry, the lab previously identified SpyM3_0890 or EndoS1 and SpyM3_1246 or P1246 as potential Prx binding partners. To study the Prx partners, we purified both proteins to study their biochemical interactions with Prx. First, we show that both proteins form stable complexes with Prx *in vitro*. Next, an X-ray crystal structure of EndoS1 shows that it resembles the specificity (S) subunit of type I restriction modification (RM) enzymes. A cocrystal structure of an EndoS1:Prx complex revealed that Prx binds the EndoS1 at its potential DBD. P1246 is located in a cluster of phage genes that regulate DNA processing and Alpha Fold predicts that P1246 adopts a DBD-fold identical to ComR. Based on our structures we have identified a helical motif conserved in ComR, EndoS1, and P1246 that is critical for Prx interaction. This motif could serve as a different approach for pursuing more Prx interaction partners. Since both EndoS1 and P1246 are potential DNA binding proteins, Prx likely blocks their ability to interact with a specific promoter or repressor region in *S. pyogenes*. Overall, our study demonstrates how small phage proteins like Prx can regulate host biochemistry for the preservation of the phage.

Acknowledgements

Not many people are able to say that they have enjoyed every day of their time in graduate school, but I am fortunate to say that I can. And this was possible because of some amazing people around me. To begin with, I was fortunate to have a super supportive PI like Dr Gerd Prehna. Gerd is an excellent mentor, always going to great lengths to accommodate a student's need, both inside and outside the lab. For many reasons I was never comfortable talking about my work and always unnecessarily feared people's judgement. I particularly remember my first few months in the lab when I was struggling to generate data, thinking I might have disappointed my PI. Gerd instead was always very kind and encouraging, appreciating me for the effort I put in. Over the last few years this inspired me a lot and has helped me gain so much confidence! Gerd creates such an inclusive laboratory environment, which invites the friendliest people. I am very grateful to all the members of the Prehna Lab, especially Nicole Rutbeek who gave me a wonderful head start on this project. I learned a lot from all my lab mates and survived through tough times with the help of their strong humor! I am very thankful to my committee members for their valuable suggestion and kindness from the beginning of my program. I would not be here today without the support of all my teachers in Bangladesh, USA and Canada, and I want to take this opportunity to thank them all.

Lastly, I want to appreciate the people who decided to stick with me even after seeing me at my worst! My husband Zisanur Rahman has been my biggest cheerleader during the whole time I have known him, and he has a huge contribution in all my achievement. I truly appreciate all the sacrifice my parents in Bangladesh have made to make sure I have this life. Indeed, I would be nothing without these people, and I am so thankful for what I have accomplished!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

β-Me	B-mercaptoethanol
ChIP-Seq	Chromatin immunoprecipitation assay with sequencing
DBD	DNA binding domain
DLS	Dynamic light scattering
GAS	Group A Streptococcus
GST	Gltutathione-S-transferase
HGT	Horizontal gene transfer
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LB	Luria-Bertani
M	Methylation
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethylene glycol
PICI	Phage inducible chromosomal island
Prx	Paratox
R	Restriction
RM	Restriction modification
S	Specificity
SD	Superdex
SDS	Sodium dodecyl sulfate
SEC	Size-exclusion chromatography
SEC-MALS	Size-exclusion chromatography with multi-angle light scattering
THY	Todd Hewitt broth (supplemented with 0.2% yeast extract)
TPR	Tetratricopeptide repeat
WT	Wild-type
XIP	SigX inducing peptide

1 Introduction

1.1 *Streptococcus pyogenes*

1.1.1 Classification and Clinical relevance

Streptococcus pyogenes, also known as Group A *Streptococcus* (GAS), is a Gram-positive bacterium within the order of Lactobacillales and belongs to the Streptococcaceae family[1]. The Lancefield classification system classifies *Streptococci* in groups based on the type of carbohydrate antigens on their cell surface[2]. GAS serotypes are characterized by the presence of group A carbohydrate (GAC), which is composed of rhamnose-rhamnose backbones with repeated N-acetyl-glucosamine (GlcNAC) side chains[3,4]. The GlcNAC constitutes 50% of the GAS cell wall and acts as the epitope that serves as the basis for rapid diagnostic testing of GAS infections[5]. Within GAS there is more than 200 serotypes which are characterized by the expression of unique surface proteins called M proteins[6]. The M proteins are an important contributor of GAS virulence, and can resist host response by interacting with different host proteins like fibrinogen[7], IgG, and complement regulatory proteins[8,9].

GAS can exhibit a wide spectrum of infections, which can range from asymptomatic commensalism to symptomatic and severe infections. GAS colonizes throat and soft tissues, where they can start superficial or invasive infections[10]. Milder forms of diseases caused by GAS include pharyngitis, scarlet fever, impetigo and cellulitis. More severe infections caused by GAS can lead to necrotizing fasciitis, bacteremia and streptococcal toxic shock syndrome[11]. Chronic GAS infections can also lead to immune-mediated diseases like acute rheumatic fever, rheumatic heart disease, and acute glomerulonephritis[10,12]. GAS causes almost 700 million superficial and throat infections and nearly 700,000 severe invasive infections. Altogether GAS diseases cause over

500,000 deaths every year[12], which make them one of the top ten pathogens responsible for infection-related deaths[12].

1.1.2 Horizontal gene transfer

Horizontal gene transfer (HGT) is a unique phenomenon that allows prokaryotes to acquire novel genetic features and hence is a key driver of bacterial evolution[13,14]. In *Streptococcus*, HGT occurs through three main processes: transformation, transduction, and conjugation. Transformation is the ability of bacteria to acquire exogenous DNA and incorporate it into their genome[15], transduction is transfer of genetic material through infection by bacteriophage[16], and conjugation is the transfer of genetic elements between bacteria using conjugative plasmids[17,18]. HGT allows introduction of entirely new genetic traits by acquisition of new DNA sequence or adds variability by homologous recombination between related DNA[19]. Genetic transfer among same or closely related species gives rise to mosaicism, where a certain area in the genome constitutes nucleic acids from both the donor and the recipient[20]. Overall, HGT can incorporate critical resistant determinants and virulence factors that often provide survival fitness to the bacterial population. A well-documented case of mosaicism providing survival fitness is the evolution of penicillin resistance in *Streptococcus*. Low affinity penicillin binding proteins (PBPs) were introduced to *Streptococcus* species by homologous recombination between closely related species[21,22]. In *S. pyogenes*, contribution of HGT is highly prominent in the *emm* gene family[23,24]. The *emm* genes encode M or M-like proteins, which can interact with proteins involved in the immune system[25,26] and act as an inhibitor of phagocytosis[10].

1.1.3 Quorum sensing and natural competence

In order to communicate with their siblings and other non-kin members of the community, bacteria have evolved a complex mode of communication called quorum sensing (QS). QS involves the

production and detection of chemical signal molecules which lead to alteration of gene expression and eventually influence phenotypic behaviour. The signal molecules are usually secreted as inactive pro-peptides which undergo a proteolytic cleavage after their secretion[27–30]. After reaching a threshold concentration, the processed active peptides can either bind to membrane receptors or be imported into the cytoplasm[31]. Once inside, peptides can bind specific transcriptional regulators and trigger various signaling cascades. In Gram-positive bacteria QS is known to drive key cellular processes like virulence, competence, sporulation, bacteriocin production and biofilm formation[32–35]. For *S. pyogenes*, imported signal peptides are bound by ComR, which is a transcription factor of the Rgg (regulator gene of glucosyltransferase) family[31].

Competence in GAS is a transient state that allows uptake of exogenous DNA from the environment and requires expression of proteins that facilitate DNA uptake and recombination[36]. The master regulator of competence is called SigX, whose expression is controlled by the QS receptor ComR[37] (Figure 1A). The competence cascade is achieved through ComS, the pro-peptide that is secreted and processed to the active 7-8 amino acid long XIP (sigX inducing peptide)[38]. XIP is imported back into the cell and binds to ComR, leading to its conformational change and dimerization[39]. ComR:XIP can then bind to the promoters of *comS* and *sigX*, inducing their expression[40]. SigX is an alternative sigma factor of RNA polymerase which can recognize a conserved DNA sequence known as a CIN-box[41]. This sequence is found upstream of the competence regulon or set of genes that express machinery for DNA uptake. SigX binding to the promoter of competence regulon results in expression of these genes and induces natural competence (Figure 1A). Despite containing all genes necessary for natural competence, GAS are not known to exhibit natural competence in laboratory setting. This suggests two things: the

existence of a negative regulator of natural competence and the inability to effectively mimic the natural environment in the laboratory.

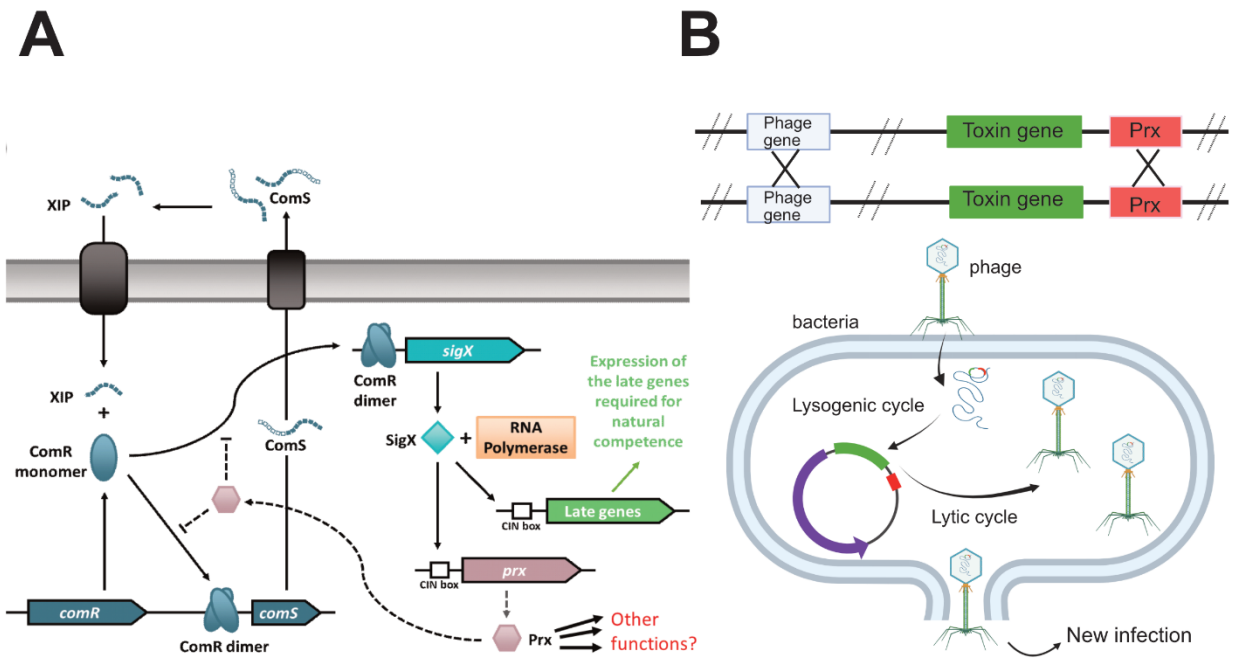


Figure 1:Paratox is an inhibitor of quorum sensing and is linked to toxin genes. (A) Quorum sensing and natural competence pathway in *S. pyogenes*. Figure was modified from the master's thesis of Nicole Rutbeek[42]. (B) The *prx* gene is genetically linked to toxin gene and both are mobilized together during infection. Prx is also one of the hot spots for genetic recombination (shown with cross).

1.2 Presence of prophages

1.2.1 Transduction and bacteriophage infection in GAS

Bacteriophages are viruses that replicate within bacteria and can be grouped according to the life cycle they adopt. Phages that follow a strictly lytic life cycle hijack the host and utilize that to manufacture more phage components[43]. After viral packaging, newly formed phages can lyse the host cell to release new virions, which go on to infect new hosts to continue the same life cycle.

The other group of phages, known as lysogenic (or temperate) phages have the potential to undergo either a lytic or a lysogenic cycle. During the lysogenic cycle, phages can integrate into host genome by site specific recombination and become a stable genetic element (prophage state), where they lie dormant and replicate with their host[43]. Whenever prophages sense host SOS responses (like activation of the host SOS response), they switch to a lytic cycle and escape the compromised host. When packaging phage components during new phage production, erroneous packing of some bacterial DNA can occur as well[44–46]. Phages containing the bacterial DNA can infect and transfer this genetic material to a new host by a process called generalized transduction[47]. The integrated DNA replicates with the host and is transferred to its progeny during cell division. In *S. pyogenes*, transduction is known to play essential role in dissemination of streptococcal genes that encode for virulence and antibiotic resistance[43]. *S. pyogenes* acts as a host of both lytic and lysogenic phages and over 90% of *S. pyogenes* serotypes are known to contain prophages[48]. The lysogenic phages of GAS are mostly T12 phages belonging to the *Siphoviridae* family[49].

1.2.2 Phage encoded virulence genes

Prophage genomes can encode for gene products that greatly influence the host virulence phenotype. *S. pyogenes* prophages are known to encode for phospholipase A2 (*sla*), DNases (*spd1*) and superantigens like *speA*, *speC*, *speH*. Superantigens like SpeA are toxins that can cause diseases like scarlet fever or lead to streptococcal toxic shock syndrome[50]. DNases can dampen the host immune system by degrading the defensive trap created by neutrophils[51]. Phospholipases enhance bacterial adhesion to their niche and aggravate host inflammatory response[52,53]. Overall, the pathogenicity of *Streptococcus* can be directly attributed to the phage encoded toxins. In fact,

genome sequencing revealed that the presence of prophages is a distinct feature that differentiates between more invasive and less invasive GAS strains[54].

1.3 Paratox

1.3.1 Conserved phage protein paratox

All prophages of pathogenic GAS strains have a highly conserved open reading frame termed *paratox* (*prx*). In the prophage, *Prx* is always located at the 3' end of a toxin gene and just before the phage attachment site[54]. The toxin genes and *prx* are known to be in a linkage disequilibrium, meaning they are transferred as a single genetic cassette[55]. GAS prophages lacking virulence genes also lack *prx* suggesting an evolutionary advantage to this linkage, although this link is not well understood (Figure 1B). The location of *prx* beside the gene that encodes viral integrase also makes *prx* a hotspot for recombination. Therefore, *prx* may play an important role in the dissemination of phage encoded toxins between GAS prophages by recombination. *Prx* genes are known to contain a CIN box in their promoter, hence *prx* expression is induced by the signal peptide XIP[41].

1.3.2 Prx inhibits natural competence and quorum sensing

A known function of the protein Paratox (Prx) is to inhibit quorum sensing (QS) and natural competence in GAS[38,41]. Past work in our lab has discovered that this inhibition is due to the interaction of Prx with ComR (Figure 1A)[38]. ComR consists of a DNA binding domain (DBD) and a tetratricopeptide repeat (TPR) domain. In the monomer form, the DBD is shielded by the TPR domain. XIP binding to the TPR domain of ComR induces a conformational change that unhinges the DBD from the TPR. This allows ComR dimerization and also exposes the DBD for interaction with its cognate promoter sites[56]. A cocrystal structure of the ComR DBD bound to Prx (PDB ID: 7N10) revealed that Prx interacts with the DBD by forming salt bridges with residues that are important for DNA binding[38]. This interaction also induces a conformation change that releases

the DBD from ComR, decoupling it from activation by XIP[38]. Overall, this prevents ComR from associating with its target promoters to inhibit GAS quorum sensing and natural competence. This in part explains why natural competence is difficult to observe in GAS, although it contains all genes needed for competence. Additionally, in a strain where *prx* was deleted GAS was found to be 100,000 times more electrocompetent[41]. By inhibiting natural competence, Prx protects itself from being replaced by recombination with newly acquired DNA[38].

1.3.3 Similarity to Aqs1

Connection between phage proteins and quorum sensing has been observed in many bacterial species that include members of *Pseudomonas* and *Vibrio*[57]. *Pseudomonas* is infected by the phage DMS3, that encodes for an anti-activator of quorum sensing. The phage derived protein Aqs1 is known to interact with LasR, the master regulator of quorum sensing in P*seudomonas aeruginosa*[58]. However, in addition to inhibiting quorum sensing, it also interacts with PilB, the type IV pilus assembly ATPase. This interaction blocks pilus assembly, but at the same time blocks superinfection of *Pseudomonas* by other pilus-dependent phages[58]. These actions are similar to what is observed with Prx, as both the phage encoded proteins are leveraging the host quorum sensing pathway to limit new DNA entry. Therefore, it is rational to hypothesize that Prx, like Aqs1, also performs multiple biochemical functions to maximize prophage economy.

1.4 Anti-phage defense in *S. pyogenes*

1.4.1 Type I restriction modification enzymes

All bacteria, even the ones that carry prophages in their genome, have multiple defense mechanisms that protect them from phage infection. In most cases, these defense systems employ proteins that can interact with phage DNA. A classic example of anti-phage defense system is the type I restriction modification (RM) enzymes. Type I RM are the first type of restriction

endonucleases that were discovered[59] . They are involved in recognition of self DNA from foreign DNA[60,61]. The type I RM system in *S. pyogenes* is composed of three different subunits: HsdS (specificity or S subunit), HsdM (methylation or M subunit), and HsdR (restriction or R subunit)[62,63]. In a holoenzyme, the subunits can exist in three stoichiometries: 2R+2M+S, 1R+2M+S and 2M+S[64]. 2R+2M+S state allows translocation and cleavage of DNA, 1R+2M+S is able to only translocate DNA but unable to cleave, and 2M+S has the potential to act as an independent methyl transferase[65,66]. In principle, type I RM recognizes the methylation states of DNA[59]. When they encounter unmethylated DNA, which is the case with phage DNA and invading mobile genetic elements, they bind to its target DNA sites through the S subunit[62,67]. After DNA binding, the motor domains of the R subunit allow ATP dependent translocation of the DNA. This results in a loop formation while the whole enzyme still remains attached to the recognition site[68,69]. The nuclease domains of the R subunit then cleaves the DNA at a site distant from the recognition site, which is followed by methylation by the M subunit[70,71]. Type I RMs have also evolved the technique to prevent erroneous cleavage of self DNA: the holoenzyme recognizes hemimethylated DNA, a characteristic of newly replicated DNA. Upon encountering such hemimethylation, the enzyme methylates the other strand and falls off the DNA[72]. In addition to protecting cells against invading DNA, methylation performed by type I RMs also play critical roles in expression of important virulence genes[73].

1.4.2 Specificity (S) subunit

Type I RMs bind to their target DNA sites with the help of the specificity (S) subunit. A typical S subunit itself is a multi-domain protein built of four domains: two globular terminal recognition domains (TRDs) and two helical conserved regions (CRs)[74] (Figure 2A). X-ray crystal structure of the S subunit of *Methanococcus jannaschii* is shown as an example (Figure 2B). The domains

are arranged such that the active conformation has two TRDs at either ends, separated by the antiparallel CRs[75] . They recognize non-palindromic, bipartite DNA which are 3-5 base pairs of nucleotides separated by 6-8 base pairs of non-specific nucleotide repeats. The molecular arrangement of the protein allows the two TRDs to be in contact with their target DNA residues upstream and downstream of the repeats while the CRs serve as a molecular ruler for their separation[76–78]. The two TRDs within the subunit show high conservation (nearly 40%)[78]. Recombination between related TRD genes generate new TRD specificities, enriching the DNA recognition repertoire[79,80]. This allows Type I RMs to evolve against phage mutations that may have changed their recognition sites[81,82].

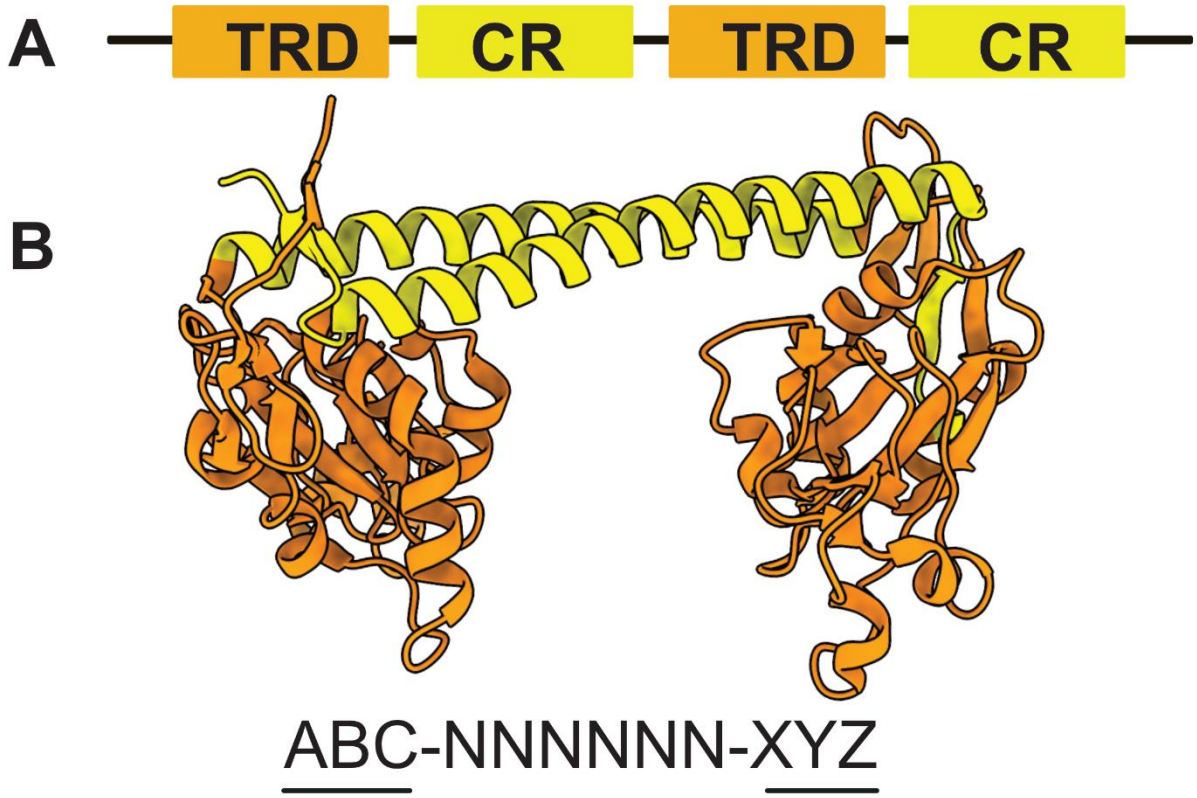


Figure 2: Structure of a specificity (S) subunit of type I restriction modification enzyme. (**A**) Domain arrangement of specificity (S) subunit of type I RM. It is composed of two globular

terminal recognition domains (TRD) (orange) and two helical conserved regions (CR) (yellow). (**B**) X-ray crystal structure of S-subunit of type I RM of *Methanococcus jannaschii* (PDB ID 1YF2)[78]. The globular TRDs bind bipartite DNA. The recognition site usually contains about 3-5 base pairs (ABC and XYZ, underlined) separated by six to eight non-specific nucleotide repeats (N). Structure was downloaded from PDB and modified using ChimeraX[83].

1.5 Objectives

1.5.1 OBJECTIVE 1: Validate paratox interactions with streptococcal proteins

Phage proteins like paratox often perform multiple biological functions and interact with multiple host proteins[58,84]. Mass-spectrometry performed on streptococcal lysates using Prx as a bait revealed a number of proteins that could potentially interact with Prx[42]. After expressing these proteins in inducible recombinant expression vectors, their interaction with Prx will be validated using various biochemical methods. This will also indicate versatility of Prx in shaping streptococcal biology.

1.5.2 OBJECTIVE 2: Elucidating the structure and function of Prx interaction partners

The potential interaction partners short-listed in mass-spectrometry analysis are uncharacterized. After their interaction with Prx is validated, obtaining their molecular structures alone and in complex with Prx will help in determining the mechanism of that interaction. The structures will also give insight into their function. Overall, characterizing the interaction partners will help uncover how Prx influences multiple GAS cellular pathways.

2 Materials and Methods

2.1 Bacterial growth conditions

Escherichia coli was grown in either Luria-Bertani (LB) broth or in plates supplemented with 1.5% agar at 37°C with adequate aeriation. When appropriate, ampicillin was used at a final concentration of 100 µg/mL. *Streptococcus pyogenes* strain MGAS315 was grown in Todd Hewitt media supplemented with 5% yeast extract (THY) at 37°C. For GAS growth in broth, 13 mL culture was grown in 15 mL plastic tubes, which were sealed tight and left standing still in an incubator. For growth on plates, THY was supplied with 1.5% agar and the plates were sealed with parafilm for carbon dioxide enrichment. To make *E. coli* freezer stocks, liquid culture and sterile glycerol was mixed to achieve a final concentration of 25% glycerol and stored in cryogenic vials at -80°C. MGAS315 freezer stocks were made by adding 400 µL of 80% glycerol to 1.2 mL of liquid overnight culture and stored in cryogenic vials at -80°C.

Because WT MGAS315 had no selection markers, the freezer stocks were periodically checked for purity by streaking on blood agar plates. Briefly, frozen culture was picked up by a sterile loop and streaked on the surface of blood agar plates and incubated overnight as described above. The plates were checked for complete hemolysis after 24-48hr, which validated the purity of the freezer stocks.

2.2 Generation of constructs

2.2.1 DNA extraction

Expression vectors used for the study included pET22b(+), pGEX6P-1 and pETDuet-1 which were purchased from GenScript. Nisin inducible plasmid pMSP3535 was purchased from Addgene. Plasmids were extracted from overnight cultures of *E. coli* DH5α using commercial plasmid extractions kits (ThermoFisher Scientific K0502). For genome extraction of MGAS315, overnight

MGAS315 cultures were processed using commercial genome extraction kit (ThermoFisher Scientific K0721).

2.2.2 Cloning in expression vectors

All constructs were generated using the restriction cloning method. *endoS1*, *hsdS* and *hsdM* were amplified by polymerase chain reaction (PCR) from MGAS315 genome and cloned in expression vector pET22b (+) (GenScript) between NdeI and XhoI sites. *endoS1* (*spyM3_0890*) was amplified using primers p140 and p141. *p1246* (*spyM3_1246*) was amplified using primers p202 and p203 and was cloned in pGEX6P-1 (GenScript) between *BamHI* and *NotI* sites. Co-expression of HsdS and HsdM was achieved by cloning the individual proteins in pETDuet-1 (GenScript) in multiple cloning site 1 and 2 respectively. HsdS was amplified from MGAS315 genome using primers p152 and p206 and cloned between *NdeI* and *XhoI* sites. HsdM was amplified using primers p183 and p207 and cloned between *EcoRI* and *NotI* sites. Wild type Prx constructs and all other Prx variants were generated by Nicole Rutbeek. EndoS1 variants were generated using Q5 mutagenesis (New England Biolabs) using the full length Prx and EndoS1 plasmids as templates. Briefly, primers with appropriate linkers were designed to amplify upstream and downstream of the nucleotide to be mutated. Q66A EndoS1 was amplified using primers p211 and p212. R69A EndoS1 was amplified using primers p213 and p214. The amplified product was ligated using T4 DNA ligase and polynucleotide kinase (NEB). Ligated vectors were transformed into competent *Escherichia coli* BL21-Gold (DE3). For generation of truncated gene versions, Q5 mutagenesis was used. Using plasmids with full length gene as template, primers were designed to amplify upstream and downstream of the unwanted region. Amplified vector was then ligated overnight using ligase and PNK (polynucleotide kinase). For cloning EndoS1 in pMSP3535, the gene with a 6His-tag was amplified off EndoS1::pET-22b(+) using primers p223 and p224. Amplified gene product was

ligated to digested pMSP3535 between BamHI and SphI. The ligated plasmid was amplified in DH5α and the resultant EndoS1::pMSP3535 plasmid was transformed in competent wild type MGAS315. All constructed vectors were verified by Sanger sequencing (Sickkids, Toronto). All primers used for generation of the constructs are listed in Table A1.

2.2.3 Competent cell generation and transformation

For generation of chemically competent *E. coli* cells, overnight cultures were diluted 1:50 in 50 mL LB broth which was grown until it reached an $OD_{600\ nm}$ of 0.4. The culture was centrifuged for 10 minutes at 4000 rpm (in a 24x1.5/2.0 mL rotor in a microcentrifuge) at 4°C. After discarding the supernatant, cells were resuspended in cold sterile 30 mL 0.1 M calcium chloride and chilled on ice for 30 minutes. They were centrifuged as above, and the calcium chloride was removed. Finally, cells were resuspended in sterile 15% glycerol and frozen at -80°C. Electrochemically competent MGAS315 cells were generated by growing overnights in THY + 20 mM glycine at 37°C. Overnight culture was diluted 1:50 in fresh THY + 20 mM glycine and grown until OD_{600nm} of 0.8. Cells were centrifuged as before and washed twice with 50 mL cold 15% glycerol. Finally, cells were resuspended in 15% glycerol and stored at -80°C.

For transformation of *E. coli*, 10-50 ng of DNA was added to thawed competent cells and chilled on ice for 30 minutes. Immediately after heat shocking the cells at 42°C for 45 seconds, they were resuspended in 1 mL LB. Cells were recovered by shaking for 1 hour at 37°C before plating on appropriate media. For transformation of MGAS315, 3-4 µg of DNA was added to thawed cells on ice. Cells were immediately pulsed at 1.75 kV using a Bio-Rad MicroPulser™. After resuspension in 1 mL fresh THY, cells were recovered at 37°C for 2-3 hours before plating on media.

2.3 Protein expression and purification

2.3.1 Small-scale induction trials.

To check for protein expression, generated constructs were grown overnight in 3 mL LB broth with appropriate antibiotic. The following day, overnight cultures were sub-cultured in fresh 3 mL LB plus antibiotic at a 1:100 dilution. The cultures were grown at 37°C until the optical density reached around 0.6 at $OD_{600\ nm}$. At this point the cultures were equally divided into two tubes, one of which received isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM. Both tubes were grown overnight in a shaker maintained at 20°C. On the following day, cultures from two tubes were transferred to separate microcentrifuge tubes and spun at 4000 rpm for 1 minute at room temperature. The pelleted cells were resuspended in 300 µL lysis buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 25 mM imidazole) and kept on ice during the rest of the procedure. Resuspended cells were sonicated for 5 seconds and centrifuged at 20000xg for 20 minutes at 4°C. Soluble fragment of the lysates were transferred to new centrifuge tubes while the insoluble pellets were resuspended in 300 µL of fresh lysis buffer. 10 µl of soluble lysates and insoluble pellets were collected from both induced and non-induced tubes and run through 15% SDS PAGE gel for about 50 minutes at 200V. Gels were stained with Coomassie and satisfactory expression of a protein was determined when distinct bands of appropriate molecular weight were observed in the soluble lysates of induced samples.

2.3.2 Large-scale protein purification

After verification of protein expression, large-scale purification was performed by growing the appropriate constructs in 30 mL LB broth with antibiotic. The following day it was sub-cultured in 3 litres of fresh LB broth at a 1 to 100 dilution and grown in a shaker at 37°C until the optical density reached 0.6. At that point the culture was supplemented with 1 mM IPTG and moved to

20°C for growth overnight. The next day induced cells were collected by centrifuging at 4200 rpm for 30 minutes and resuspended in 100 mL lysis buffer. At this point, pellets were either frozen at -80°C or prepared for lysis.

All steps performed after the lysis required the samples to be on ice. Before lysis, resuspended cells were treated with 1 mM phenylmethylsulfonyl fluoride (PMSF), 10 mM $MgCl_2$, and a small amount of DNase I (approximately 1-5 mg). The cells were then lysed by passing through an Emulsiflex-C3 (Avestin). The cell lysate coming out of the Emulsiflex was centrifuged again at 17000 rpm (30,000xg), 4°C for 30 minutes to separate the soluble portion of the lysate from the insoluble pellet. The soluble fraction should ideally contain the protein of interest and is treated in accordance with the tag it was cloned with.

2.3.3 Affinity protein purification

For proteins that were cloned with a 6His-tag, the proteins were purified using Nickel-affinity chromatography. The columns were prepared by adding 8 mL of Nickel NTA agarose resin (GoldBio). The ethanol in which the resin was stored was first washed off by passing several column volumes of water through the column. The resins were charged by passing ~15 mL of 0.2 M $NiSO_4$. The nickel was washed off with water and the column was equilibrated with a commonly used lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl, 25 mM imidazole). The soluble lysate was loaded on the column and the flow-through was collected. The column was washed with five column volumes of the same lysis buffer (wash). Finally, the protein bound to the resins was eluted using an elution buffer that composed of 50 mM Tris pH7.5, 500 mM NaCl and 500 mM imidazole. Samples collected at all stages of the purification (lysate, flow-through, wash, elution) were saved for running through an SDS-PAGE gel. Eluted protein was concentrated and purified further using size exclusion chromatography (SEC) using AKTA pure (GE). Depending on molecular weight,

proteins were run over either the HiLoad 16/600 Superdex 75 or the Superdex 200 size-exclusion column.

2.3.4 GST-tagged protein purification

GST tagged proteins were purified by gravity column using glutathione-sepharose resin (Cytiva). Briefly, the lysates were first passed through a Q-sepharose column (Cytiva) for removal of debris, some proteins, and DNA. The Q-sepharose column was first washed with water, and then equilibrated with lysis buffer (50 mM Tris pH 7.5, 200 mM $NaCl_2$, 2 mM β-Me). The partially purified lysate was collected to be loaded on the glutathione column. To clean the Q-sepharose column, 1 column volume of regeneration buffer 1 (50 mM Tris pH 8.8, 1 M NaCl) was passed through it, which was followed by several column volumes of water. Finally, the resins were preserved by 20% ethanol. For preparing the glutathione-sepharose column, 8 mL resin in a column was washed with water to remove the residual ethanol, and the column was equilibrated with the same lysis buffer. Lysate collected from the Q-sepharose column was run through the GST column, followed by wash with 200 mL of wash buffer (50 mM Tris pH 7.5, 500 mM NaCl, 2 mM β-Me). Finally, the proteins that were bound to the glutathione resins were eluted by passing 50 mL of elution buffer (50 mM Tris pH 7.5, 500 mM NaCl, 20 mM glutathione). The elution buffer was made just prior to use and the pH of the buffer was adjusted after addition of the glutathione. After elution the glutathione column was washed with 50 mL of regeneration buffer 1, then regeneration buffer 2 (20 mM sodium acetate pH 4.5, 1 M NaCl), and finally water before storing the resins in 20% ethanol.

To cleave off the GST-tag, HRV-3C protease was added to the eluted protein at a weight ratio of 1:50 (enzyme:protein). The mixture was placed in Snakeskin (ThermoFisher) dialysis tubing and dialyzed overnight at 4°C in digestion buffer that composed of 50 mM Tris pH 7.5, 200 mM NaCl,

2 mM β-Me. The following day dialyzed protein was centrifuged for 15 minutes at 4200 rpm at 4°C and re-run over the glutathione-sepharose column. Digested protein in the flow-through was collected while cleaved GST and undigested protein was removed by binding the column. To further remove residual cleaved GST the flow-through was purified using SEC to completely isolate the protein of interest.

2.3.5 Size exclusion chromatography (SEC)

For obtaining pure protein suitable for biochemical analysis, both 6His-tagged and GST-tagged proteins were purified using size exclusion chromatography (SEC). This also allowed complete buffer exchange of the proteins into their suitable buffers. Gel filtration buffer for Prx was composed of 20 mM Tris at pH7.5, 100 NaCl, and 1 mM β-Me. Gel filtration buffer for EndoS1, HsdM and HsdS composed of 50 mM HEPES at pH 7.5, 150 mM NaCl and 1 mM β-Me. Before performing SEC, all eluted proteins were concentrated to a volume of 2-5 mL using Amicon centrifugal filters with appropriate molecular weight cut off. Cut off value for Prx was 3 kDa and for all other proteins it was 10 kDa.

All SEC was performed at 4°C using an AKTA pure in a HiLoad 16/600 Superdex (SD) 75 or HiLoad 16/600 Superdex (SD) 200 gel filtration column (GE Healthcare). When not in use, the columns are stored in 20% ethanol to prevent any contamination. Before use, the columns are prepared by passing through two column volumes of Milli-Q water and then equilibrated by passing two column volumes of appropriate gel filtration buffer. All buffer passed through SEC column was de-gassed by vacuum filtration. Prior to loading, protein was centrifuged for 5 min at 4°C, 17000 g to remove any precipitation. Protein eluted was saved for running on an SDS-PAGE gel to check for purity by Coomassie stain. After satisfactory protein yield was obtained, pure protein was used immediately, or flash frozen in liquid nitrogen and stored in -80°C for future use.

2.4 Protein:protein interaction assays

2.4.1 SEC binding assay

SEC protein binding assays were performed using an AKTA pure, using either a Superdex75 increase 10/300 column (for EndoS1, Prx, P1246) or Superdex200 increase 10/300 column (for EndoS1, HsdM). The proteins for each assay were incubated together or resuspended alone in buffer on ice for half an hour before loading on a SEC column. After column equilibration, each protein alone was run though a SEC column, followed by the protein:protein mixture, keeping all SEC parameters constant for accurate comparison. As a general rule, protein with higher molecular weight was incubated with protein of smaller molecular weight at a ratio of 1:2. In all cases, the gel filtration buffer was used (50 mM HEPES at pH 7.5, 150 mM NaCl, I mM β-Me).

2.4.2 Isothermal titration calorimetry

Before performing isothermal titration calorimetry (ITC), EndoS1 and Prx were dialyzed overnight in the same buffer (50 mM HEPES at pH 7.5, 150 mM NaCl, I mM β-Me). ITC was performed using a MicroCal ITC200 (Malvern). Each experiment required ~40 µL of 600 µM of Prx injected to ~240 µL of 50 µM EndoS1 at a constant temperature of 25°C. Controls were run with buffer injected into EndoS1 and Prx injected into buffer. Data was evaluated using the Origin Software (Malvern) and binding constant was calculated using a one-site model.

2.5 Protein crystallization and data collection

2.5.1 Crystallization screening

Crystal trays were set up using commercially available crystal screens from Nextal by sitting vapor diffusion in Intelli-Plate 96-3 LVR plates. Briefly, 60 µL mother liquor was dispensed into each reservoir that had 3 sample wells. Each sample well was tested for a specific protein concentration. For both EndoS1 and (Δ1-9)EndoS1:Prx crystals, proteins were screened at a concentration of 12

mg/mL, 8 mg/mL and 4 mg/mL. Proteins were mixed with the mother liquor at a 1:1 ratio and the final drop dispensed in the sample well contained 1.5 µL liquor and 1.5 µL protein. Prior to setting up EndoS1:Prx cocrystal trays, (Δ1-9)EndoS1 and Prx were mixed at a 1:2 ratio and incubated on ice for 30 minutes. Crystal trays for EndoS1 were set up using a Crystal Gryphon Robot (Art Robbins Instrument). Trays for (Δ1-9)EndoS1:Prx complex were set up manually using a multichannel pipette. After setting up trays, they were incubated at 4°C and checked for crystal formation periodically. EndoS1 crystals were formed in in JCSG+ (NeXtal): D7 (0.2 M Lithium sulfate, 0.1 M Tris pH 8.5, 40% PEG 200). (Δ1-9)EndoS1:Prx cocrystals were formed in JCSG+ (NeXtal) D3 (0.2 M Sodium chloride, 0.1M Na/K phosphate pH6.2, 50% PEG 200).

2.5.2 Data collection

Crystals were first tested for cryoprotection. The cryoprotectant for EndoS1 was composed of the crystallization buffer supplemented with 35% PEG. The crystals were flash frozen in liquid nitrogen and immediately mounted on the in house MicroMax-007 HF X-ray source and R-axis 4++ detector (Rigaku). (Δ1-9)EndoS1:Prx cocrystal complex did not require addition of cryoprotectant as the crystallization buffer already contained 50% PEG 200. Frozen crystals were also shipped to the Canadian Light Source (CLS) in Saskatoon, SK for remote data collection on beamline 08B1.

2.5.3 Data processing and refinement

All the collected data was processed by XDS[85] and CCP4[86]. Phases for EndoS1 were obtained by molecular replacement using specificity subunit of *Methanococcus jannaschii* (PDB ID 1YF2). The structures were built with Coot[87] and refined with Phenix[88]. Graphical manipulation of the structures were performed using Chimera X[83].

2.6 Protein detection

2.6.1 Western blot

Cell lysates for western blots were prepared by sonicating 300 µL cell suspension for 5 seconds. After sonication, the samples were centrifuged 20 minutes at 4°C and 21000g. The soluble and insoluble cell fractions were separated and mixed with 4X laemmli SDS buffer at a 3:1 ratio. 15 µl of each sample was run over 15% SDS-PAGE gel for 50 minutes at 200V. Buffer used for running buffer was commonly used 1X Tris with SDS (Bio-Rad). After separation of proteins, they were transferred to a nitrocellulose membrane at 100V and 350 mAmps at 4°C. The transfer buffer was composed of 25 mM Tris at pH 8.3, 192 mM glycine, 20 % methanol and 0.025% SDS. After transfer the membrane was washed for 5 minutes with TBST (25 mM tris, 190 mM NaCl, 0.1% tween) and blocked for 1 hour at room temperature by incubating with 5% milk in TBST. After blocking the membrane was washed for 5 minutes with TBST and incubated with primary antibody (NEB) at 4°C overnight. Next day the membrane was washed with TBST and incubated with HRP-conjugated secondary antibody (NEB) for 1 hour at room temperature. After washing the membrane with TBST, immunoblotting was performed with ECL™ (GE Healthcare) and membrane was imaged in ChemiDoc™ (Bio-Rad).

3 Results

3.1 Purification of potential Prx interaction partners identified from pull-down assays

3.1.1 Searching for potential Prx interaction partners

We hypothesized that if Prx has multiple biological functions, this would require interaction with different host proteins. To search for potential Prx binding partners, a former graduate student Nicole Rutbeek performed a pull-down assays with cell lysates of MGAS315 using immobilized Prx as a bait[42]. As Prx is a phage protein and we know that Prx expression is induced by XIP [41] the cells were treated as follows: stimulation with XIP to induce the quorum sensing pathway ComRS, treatment with mitomycin C to induce phage genes, and untreated MGAS315 to express constitutive proteins. Proteins that co-precipitated from the lysates with immobilized Prx were analyzed by mass-spectrometry. After omitting essential and constitutively expressed proteins, a final curated list of five potential Prx binding proteins[42] was created (Table 1). These uncharacterized proteins included SpyM3_0890 (similar to the specificity subunit of type I restriction enzymes), SpyM3_0307 (hypothetical phage protein), SpyM3_1346 (HTH cro/C1-type domain-containing protein), SpyM3_0964 (HTH cro/C1-type domain-containing protein), SpyM3_0655 (similar to extracellular hyaluronate lyase), and SpyM3_1246 (hypothetical phage protein). Interestingly, SpyM3_1346 and SpyM3_0964 were seen to have same amino acid sequence but were expressed from different loci. SpyM3_1346 was more prevalent under XIP induction while SpyM3_0964 was more prevalent under mitomycin induction, suggesting the genes for these proteins are under the regulation of different promoters[42].

Table 1: Probable Prx binding partners shortlisted in pull-down assay with MGAS315 lysate.

Locus tag	Annotation
SpyM3_0890 (EndoS1)	Specificity subunit of type I restriction modification enzyme
SpyM3_0307	Putative phage protein
SpyM3_1346	HTH cro/C1-type domain-containing protein
SpyM3_0964	HTH cro/C1-type domain-containing protein
SpyM3_0665	Extracellular hyaluronate lyase
SpyM3_1246 (P1246)	Phage protein

Listed proteins had the highest log of expectation value (E), or highest probability of being present when induced by mitomycin C or XIP. The proteins are listed in order of their E value, with the highest hit on top.

3.1.2 Purification of potential interaction partners

The next step in identifying Prx binding partners was to purify the proteins that were shortlisted. Initially, the following proteins were selected for construct generation: SpyM3_0890 for having the highest probability, SpyM3_0307 and SpyM3_1246 for their phage origin, and SpyM3_0946 for its redundancy. The genes were amplified by polymerase chain reaction (PCR) and cloned in expression vector pET22b(+). SpyM3_0307 could not be cloned into the expression vector and was not pursued further. Expression vectors with the rest of the candidate genes were transformed in *E. coli* BL21 (DE3) Gold cells and over expression induced using Isopropyl-β-D-Thiogalactopyranoside (IPTG). The proteins were purified by nickel-affinity chromatography by taking advantage of their 6His-tags. SpyM3_0890 appeared quite stable and soluble (while SpyM3_0964 and SpyM3_1246 were less stable). At this point SpyM3_0964 was not pursued any

further. Adding a GST (glutathione-S-transferase) tag is known to increase stability of some proteins, so SpyM3_1246 was cloned in a pGEX6P-1 vector. The protein was purified using a glutathione-sepharose column and this resulted in more solubility and higher yield of SpyM3_1246. After affinity chromatography, both proteins were further purified using size exclusion chromatography (SEC) using an AKTA pure. From this point SpyM3_0890 is referred to as EndoS1 and SpyM3_1264 is referred to as P1246. The eluted EndoS1 (SpyM3_0890) appeared as a protein with a molecular weight of about 23 kDa on an SDS-PAGE gel (Figure 3A, top gel) and produced a single uniform peak in SEC, showing that the protein is one species in solution. GST-P1246 (SpyM3_1246) appeared to have a molecular weight of around 35 kDa although it was predicted to be a protein of 9 kDa. The higher-than-expected molecular weight was due to the attachment of the GST-tag which has a molecular weight of about 26 kDa. The GST-tag was removed by a proteolytic cleavage, and the free P1246 was re-purified using SEC. As expected, tag-free P1246 appeared as a 9 kDa protein on a Coomassie stained SDS PAGE gel (**Figure 3B** top gel).

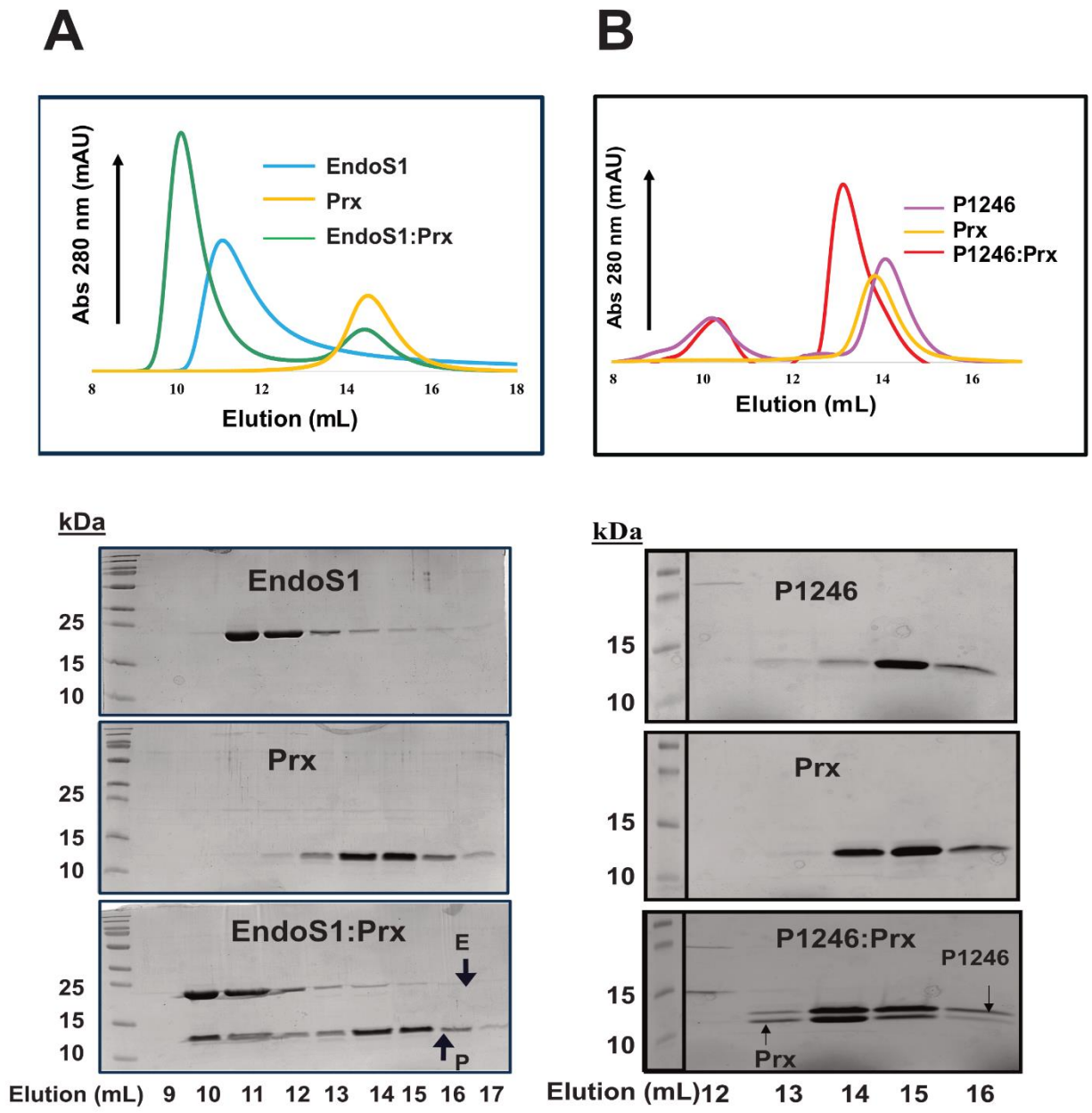


Figure 3: EndoS1 and P1246 directly bind Prx. (**A**) Size-exclusion chromatography (SEC) binding assay of EndoS1 with Prx. The chromatogram of EndoS1:Prx complex (green) produced a shift in elution volume which is indicative of complex formation. Bottom panel shows SDS-PAGE gels of the representative volumes of each protein run in the assay. (B) SEC binding assay

of P1246 with Prx. P1246:Prx complex produces a shift in elution volume (red). Bottom panel shows SDS-PAGE gels of individual protein runs.

3.2 Validation of Prx interaction partners

3.2.1 Size-exclusion chromatography binding assay of EndoS1 and P1246

Since the mass-spectrometry only predicted the probability of these proteins being present in the pull-down sample with Prx, the next step was to perform a binding assay of these proteins with Prx. Each potential interaction partner was incubated with molar excess of Prx and loaded onto a SEC column for separation by size. The proteins were also loaded on their own as controls. In SEC, larger proteins elute before smaller proteins. Therefore, if two proteins interact with each other and create a complex, the complex will have a greater molecular weight than the individual proteins and will elute earlier. When mixed and loaded together, EndoS1:Prx complex produced a shift from 11.6 mL to 10 mL (Figure 3A, top panel). This shift in elution volume is indicative of stable complex formation. Fractions collected from SEC were further verified by SDS-PAGE gels, which showed co-elution of Prx and EndoS1 at the volumes corresponding to the elution shifts (Figure 3A, bottom panel).

However, P1246 had to be removed from the GST-tag before performing the binding assay. A pull-down binding assay was attempted with GST-P1246 but that construct was unable to show any binding. Interaction between the proteins may have required access to an interface that was shielded by the large N-terminal GST-tag. Untagged P1246 and Prx were loaded together over SEC, and a shift in elution volume was observed. P1246:Prx eluted at 13.6 mL while P1246 alone eluted at 15 mL, showing P1246 directly binds Prx (Figure 3B, top panel). In SEC, Prx alone elutes earlier (14.4 mL) than P1246 (15 mL), despite being smaller (8 kDa and 9 kDa respectively) . This is because in absence of a binding partner, Prx adopts a more relaxed partially folded conformation

that increases its hydrodynamic radius[38]. However, after the complex fractions from SEC were visualized by a Coomassie stained SDS-PAGE gel, co-elution could be observed at 13 mL, which showed the presence of two distinct bands of different sizes (Figure 3B, bottom panel)

3.2.2 Prx residues E6 and D32 are critical for binding EndoS1 and P1246

After validating that EndoS1 and P1246 directly bind Prx, we proceeded to identify Prx residues that could be critical for this interaction. Similar binding assays were performed with different Prx variants that were generated previously[38]. These variants included E6A, E9A, D12A, F31W, and Y33A. In these variants, the original amino acids were substituted to alanine, except F31W, where it was substituted to tryptophan. The individual variants were mixed separately with EndoS1 and loaded over SEC column to check for a shift in their elution volumes. However, the yield of P1246 was low after the GST-tag removal, and it was only tested with Prx variant D32A. In the case of EndoS1, the characteristic shift in elution volume that was observed with the wild type (WT) Prx complex was obtained with all Prx variants except E6A and D32A (Figure 4A). A similar absence of shift in elution volume was observed with P1246:D32APrx complex (Figure 4B). This indicates that Prx residue E6 is critical for interaction with EndoS1 and residue D32 is critical for interacting with both EndoS1 and P1246. These finding is especially interesting because Prx residues E6, E9 and D32 are known to be critical for interacting with the DBD (DNA binding domain) of ComR. Given that two out of three Prx residues important for binding ComR are also important for binding EndoS1, Prx may be using a similar, although not identical, interface to interact with these three binding partners.

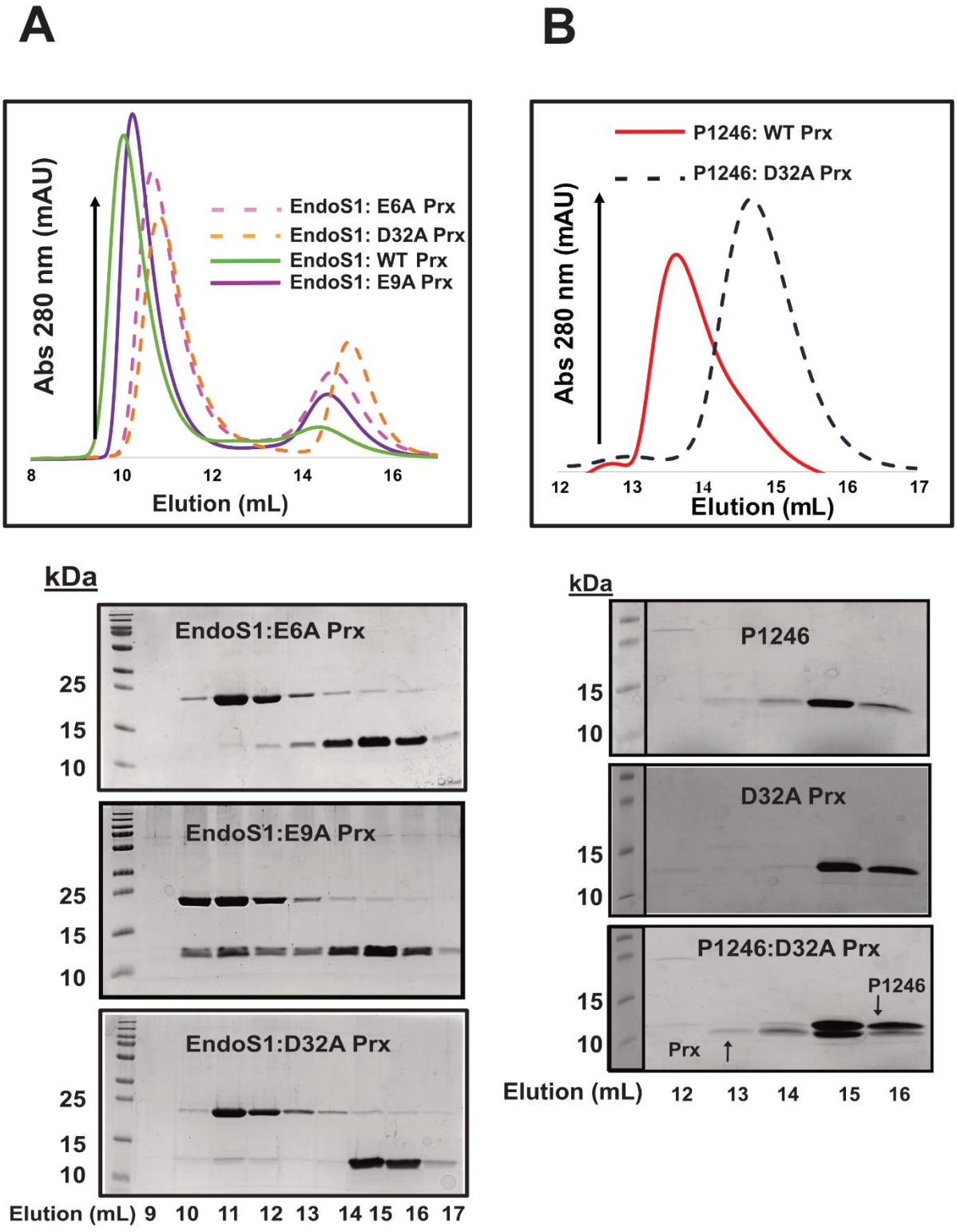

A
B
Abs 280 nm (mAU)
EndoS1: E6A Prx
EndoS1: D32A Prx
EndoS1: WT Prx
EndoS1: E9A Prx
8
10
12
14
16
Elution (mL)
P1246: WT Prx
P1246: D32A Prx
12
13
14
15
16
17
kDa
EndoS1:E6A Prx
EndoS1:E9A Prx
EndoS1:D32A Prx
25
15
10
Elution (mL) 9 10 11 12 13 14 15 16 17
P1246
D32A Prx
P1246:D32A Prx
Prx
Elution (mL) 12 13 14 15 16

Figure 4: Complex formation requires key Prx residues. (**A**) Top panel shows SEC binding assay of EndoS1 with different Prx variants. EndoS1 was mixed with each Prx variant separately and loaded over SEC column. Shift in elution volume produced with wild type Prx (green) was absent with E6A Prx (pink dotted) and D32A Prx (orange dotted). Bottom panel shows SDS-PAGE gels of EndoS1:Prx binding assay. Co-elution of EndoS1 with Prx was not observed with Prx variant E6A and D32A. **(B)** SEC binding assay of P1246 with Prx variant D32A. D32A failed to produce the shift in elution volume (black dotted) that was observed with wild type Prx (red). Bottom panel shows SDS-PAGE gel images of individual protein runs.

3.2.3 Isothermal titration calorimetry of Prx and EndoS1

To make a quantitative assessment of the binding between Prx and its partners, isothermal titration calorimetry (ITC) was performed with Prx and EndoS1. P1246 yield was too low for ITC experiments. For ITC, 50 μM of EndoS1 was placed in the thermal cell to which 600 μM Prx was periodically injected. In ITC there is a change in heat when two proteins interact, and the change in heat is proportional to the strength of this interaction. Each Prx injection caused a downward peak in the calorimeter, indicating the release of heat or an exothermic reaction, which subsided as EndoS1 reached saturation. Multiple thermodynamic parameters were calculated from this experiment: The interaction between Prx and EndoS1 yielded an enthalpy (ΔH) of -24.8 $\pm$ 2.2 Kcal/mol. The change in entropy (ΔS) of the interaction was -61.4 cal/mol/deg, which was unfavorable and therefore indicates a decrease in disorder. This could be due to the interaction being electrostatic which causes less solvent reordering as compared to a hydrophobic interaction. This also could be due to the loss of paratox conformational flexibility after binding EndoS1. Using a one-site binding model, Prx binds EndoS1 with a K_d of 18 μM as a 1:1 complex (Figure 5A). ITC was also performed with Prx variant E9A (Figure 5B), which was of interest because E9A Prx

interrupted binding with ComR but not with EndoS1. It was observed that E9A binds EndoS1 with a binding constant of 50 µM. This indicated that although not critical, mutation in residue E9 decreased the binding affinity.

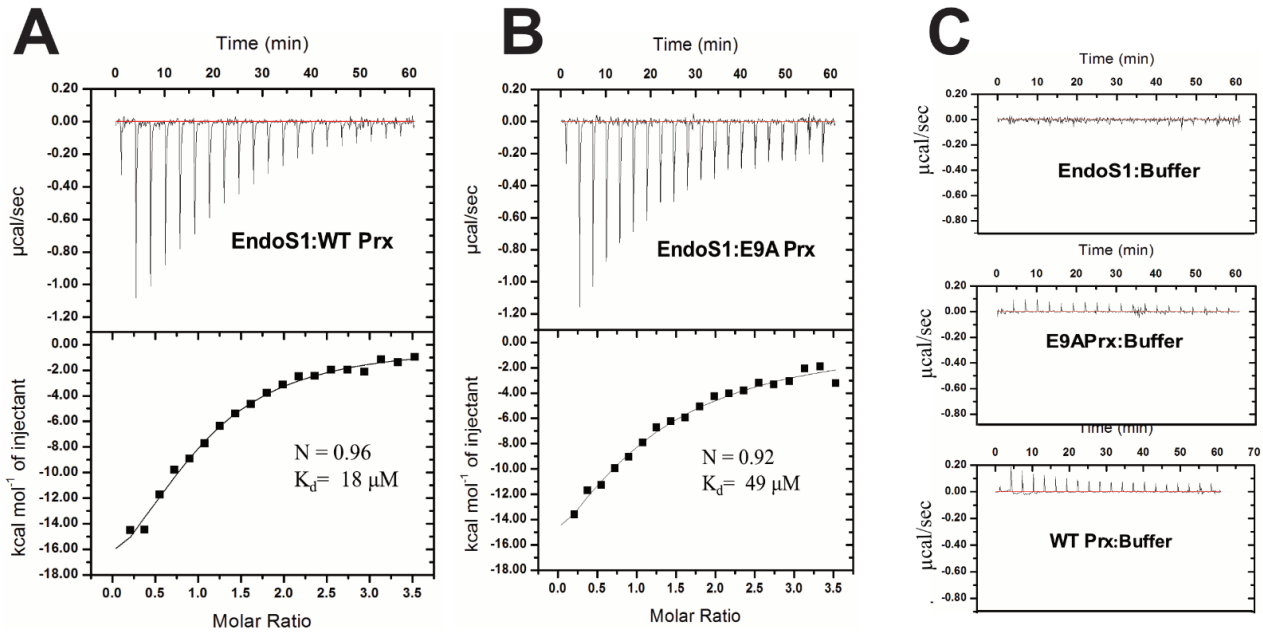


Figure 5: Isothermal titration calorimetry (ITC) of EndoS1 with Prx. (A) EndoS1 binds wild type (WT) Prx with an affinity of 18 µM and a stoichiometry of 1:1. **(B)** EndoS1 binds Prx variant E9A that is lower than observed with WT Prx, although no disruption in binding was observed in SEC binding assay. **(C)** ITC controls of EndoS1, WT Prx and E9A Prx with buffer.

3.3 EndoS1 is a homodimer

3.3.1 Dynamic light scattering of EndoS1

In SDS-PAGE gels, EndoS1 appeared as a protein of about 23 kDa, which was in agreement with its predicted size obtained from bioinformatic servers. However, in SEC EndoS1 elutes earlier than what was expected for a protein of this size, indicating that it may be an oligomer. Although the elution volume of a protein that has passed through SEC can give an idea about the molecular

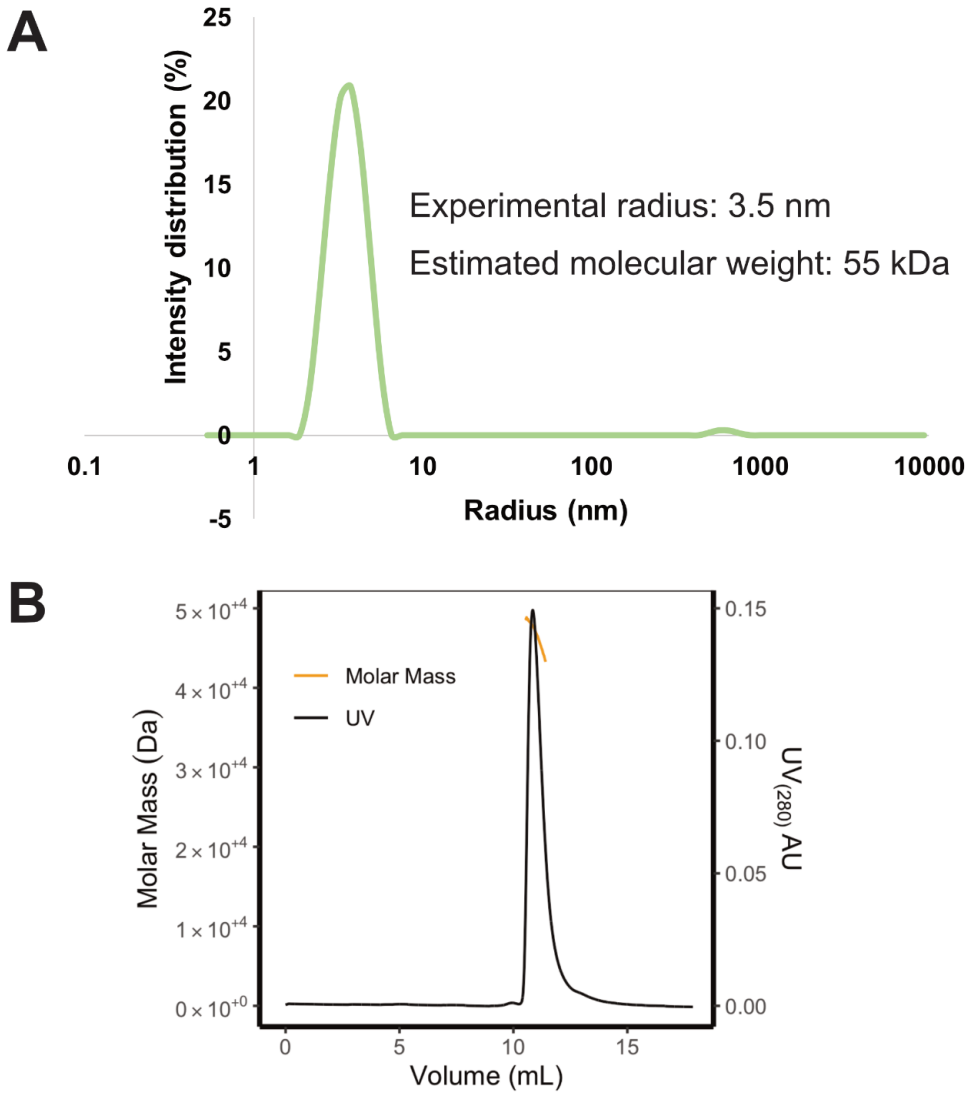
weight of the protein, it is based on the hydrodynamic radius of the particle and not its actual mass. Since the protein gels were run in denaturing conditions (due to addition of SDS), any higher order oligomers formed by EndoS1 would not be detected. So, to determine the molecular weight and the oligomeric state of EndoS1, DLS (dynamic light scattering) was performed. In DLS, the protein sample is illuminated by a laser and light scattering by the particles in the sample can be used to calculate is hydrodynamic radius or its apparent molecular weight. DLS was performed using a Prometheus Panta (NanoTemper). Purified EndoS1 was loaded in capillaries and 3 replicates were carried out at 25º C.

In DLS EndoS1 appeared as mostly one species in solution, meaning it was homogenous with minimum aggregates. The cumulant hydrodynamic radius obtained by the majority species in the sample was 3.5 nm (Figure 6A). To calculate the molecular weight of EndoS1, this value was plotted in a standard curve generated with hydrodynamic radii of proteins with known molecular weight. These proteins included cytochrome c, myoglobin, carbonic anhydrase, ovalbumin, albumin, alcohol dehydrogenase, aldolase, catalase, ferritin and thyroglobulin. The molecular weight of the EndoS1 was calculated to be around 55 kDa, which is significantly larger than the 23 KDa band observed in an SDS-PAGE gel. Overall, DLS supported that EndoS1 existed as an oligomer.

3.3.2 SEC-MALS (multi-angle light scattering)

For absolute characterization of the molecular weight and conformation of EndoS1 in solution, we combined multi-angle light scattering with size exclusion chromatography (SEC-MALS). In this assay, EndoS1 eluted as a single peak which composed almost 100% of the mass fraction that was injected. The single peak indicated the predominance of a single species, or oligomeric state, of EndoS1 in solution. This peak yielded a molecular weight of 46 kDa with hydrodynamic radius of

3.236 nm. This is close to the molecular weight and the hydrodynamic radius obtained in DLS (55 kDa and 3.55 nm respectively). Taken together, the SEC-MALS is in agreement with our previous observations that determined that EndoS1 exists as a homodimer in solution (Figure 6B).



	Peak 1
Hydrodynamic radius (nm)	3.236 (±0.232%)
Mass Fraction (%)	100.0
Mw (kDa)	46 (±0.2%)
Polydispersity (Mw/ Mn)	1.003 (±0.231%)

Figure 6: EndoS1 forms a homodimer. (A) Dynamic light scattering of purified EndoS1 reveals a hydrodynamic radius of 3.5 nm. Conversion of the radius to molecular weight indicates EndoS1

oligomerization. (**B**) SEC-MALS depiction of the single peak produced by EndoS1. The molecular mass fit of the peak is shown in yellow. The mass-analysis produced by ASTRA software listed below shows that EndoS1 dimer has a molecular weight of 46 kDa.

3.4 Crystal structures of EndoS1 and an EndoS1:Prx complex

3.4.1 Crystal structure of EndoS1

After the discovery of EndoS1 as a Prx binding partner, the amino acid sequence of EndoS1 was entered into prediction server PHYRE[89] to gain some insights into its possible structure. The results predicted its similarity to the specificity (S) subunit of type I restriction modification (RM) enzymes. S subunits are used for binding to specific recognition sites on DNA and are characterized by two globular heads separated by two helices[64]. Interestingly, the PHYRE, along with AlphaFold, predicted EndoS1 to have one globular domain and one helical arm, thereby resembling a partially assembled S subunit.

We were able to pursue an X-ray crystal structure of full-length EndoS1. EndoS1 crystallized as a homodimer (Figure 7A). However, the crystal had translational non-crystallographic symmetry, limiting the ability to fully refine the model. Keeping the AlphaFold and PHYRE predictions in mind, the crystal structure of the S subunit of *Methanococcus jannaschii* (PDB ID 1YF2) was used for molecular replacement while building the EndoS1 crystal structure. Much similar to the predictions, the partially built crystal structure revealed that each monomer consisted of an N-terminal α+β mixed globular head (M1- L153) and a C-terminal arm made of 11-turn-helices (V159-K195). The linker region between the globular head and the helical arm is composed of amino acids that create a hydrophobic patch (155-VIPL-158). This region forms hydrophobic interactions with Y117 and L113 in the globular domain, stabilizing the head-arm arrangement. The two units of EndoS1 arrange themselves in an antiparallel way, which positioned the two

globular domains of the protein at two ends. The arms arrange themselves in a coiled-coil structure and act as the dimer interface. The helical arms of EndoS1 are aligned in a way that positions amino acids from one arm packed into the amino acids of the other in a knob-hole fashion[90] (Figure 7A, right).

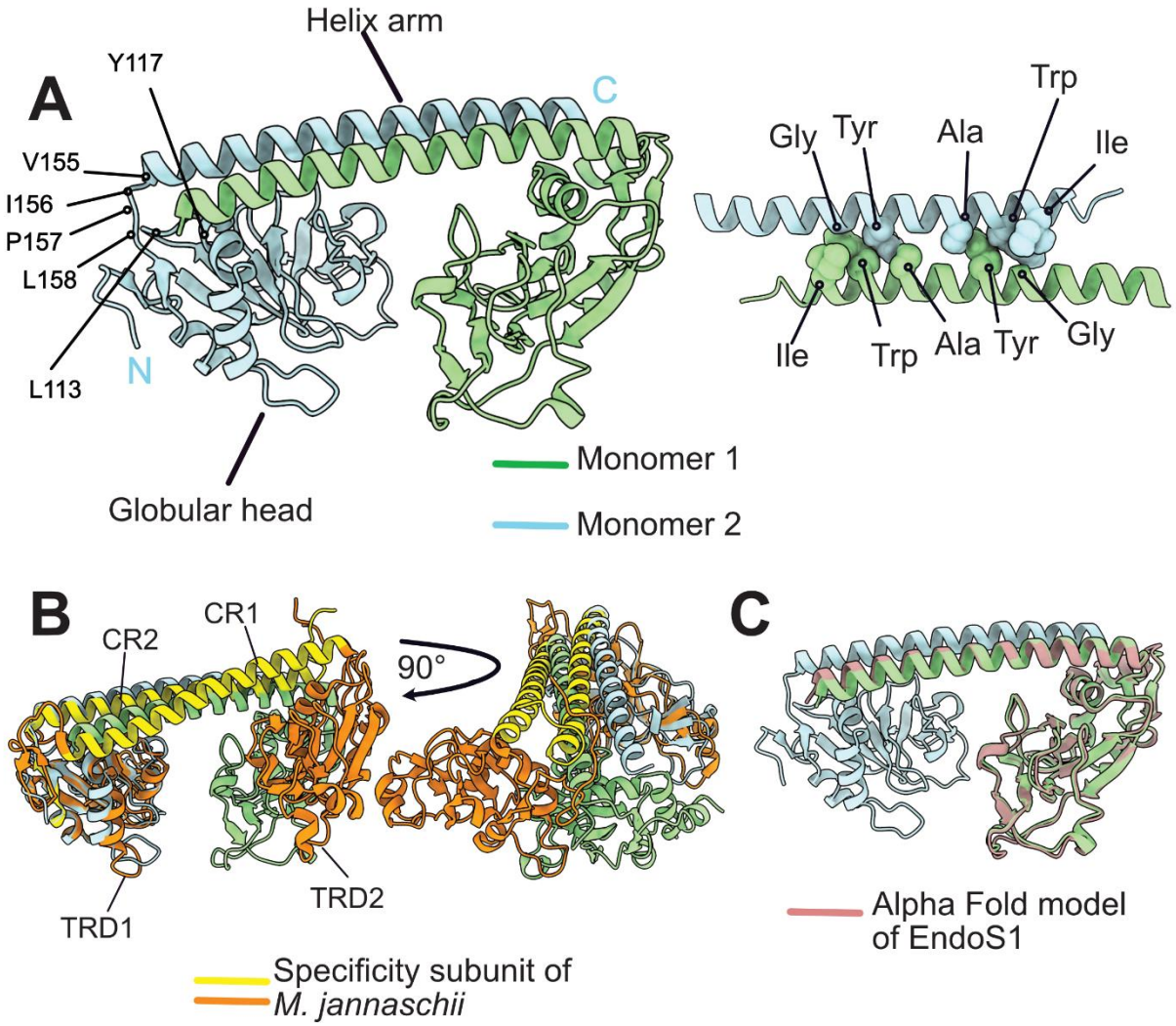


Figure 7: Partial X-ray crystal structure of EndoS1. (**A**) Two monomers of EndoS1 (green and blue) arrange anti-parallel to each other to form a homodimer. The helices contain hydrophobic residues that are packed into a knob-hole arrangement. (**B**) Alignment of EndoS1 crystal structure with the crystal structure of specificity (S) subunit of *M. jannaschii* (PDB ID 1YF2). The DNA binding domains (TRD1 and TRD2) of the S subunit are colored orange, and the helices (CR1 and CR2) are yellow. Variation in their conformation causes one of the EndoS1 monomers to be deviated away from TRD1 of the S subunit. (C) Alignment of EndoS1 crystal structure with AlphaFold model of an EndoS1 monomer (rosy brown).

Table 2: Data collection and refinement statistics for partially built EndoS1

Data collection	
Wavelength (Å)	1.1808
Space group	P 1 2 1
Cell dimensions	
a, b, c (Å)	82.362 43.872 85.8406
α, β, γ (°)	90 118.668 90
Resolution (Å)	42.87-2.7 (2.797-2.7)
Total Reflections	50243 (6629)
Unique reflections	14936 (1971)
CC(1/2)	0.999 (0.636)
R_{merge}	0.080 (1.157)
R_{pim}	0.051 (0.738)
$I/\sigma I$	8.2 (0.9)
Completeness (%)	98.9 (98.3)
Redundancy	3.4
Refinement	
R_{work}/R_{free} (%)	0.4823/0.4849
Average B-factors (Å^2)	91.81
Protein	91.83
Ligands	85.85
No. of atoms	
Protein	392
Ligand	10
Water	0
Rms deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.76
Ramachandran plot (%)	
Total favored	90.46
Total allowed	8.51

3.4.2 EndoS1 resembles the specificity (S) subunit of the type I restriction modification (RM) enzyme

Crystal structure of homodimer EndoS1 was superimposed over the crystal structure of S-subunit from *M. jannaschii* (PDB ID 1YF2). Inside cells, S subunits form a critical component of the Type I RM that are important for mediating defense against phage DNA. An S subunit is a multidomain protein composed of two globular heads, also known as terminal recognition domains (TRD1 and TRD2) and 2 helical regions known as conserved regions (CR1 and CR2)[78]. The domains fold in an arrangement that positions TRD1 and TRD2 at either ends with the two CRs in between. The overall structure of the dimerized EndoS1 show a strong resemblance to the S subunit with some variation in their conformations (Figure 7B). TRD1 of the S subunit aligned well with one monomer of EndoS1 but the α-helices of EndoS1 have a different conformation than the CRs of the S subunit. This difference in helical regions led to variable globular head arrangement in both proteins. TRD2 in the S subunit is rotated away from the EndoS1 globular head by almost 180°. (Figure 7B, right). The TRDs recognize non-palindromic bipartite DNA, that is groups of specific nucleotides that are separated by a stretch of 6-8 non-specific nucleotides. The resemblance of the globular head of EndoS1 to the TRDs of S subunit indicates the possibility of the heads being DNA binding domains. The most distinctive feature of the S subunit is the fact that all four domains (2 TRDs and 2 CRs) are part of a single polypeptide chain. EndoS1 on the other hand is characterized by one globular head and one helical arm. The dimerized EndoS1 reconfirms the data provided by DLS and SEC-MALS. However, the most critical attribute of the dimerization of EndoS1 maybe to closely mimic the structure of an actual S subunit.

It is thought that the CRs of S subunit serve as a molecular ruler between the target recognition sites. The helical region in the S subunit is also known to be important for mediating binding with

other proteins, but it is not known if the helices in EndoS1 have the same function. However, it was seen that they are essential for the solubilization of the protein, as attempt of cloning the minimal globular domain resulted in a protein that is highly insoluble.

The crystal structure of the EndoS1 dimer was also compared to the predicted AlphaFold model (Figure 7C). Each monomer of EndoS1 shows high similarity to the predicted model, with only slight variation at positions K28-V39, A134-D139 of the globular head and I189-G198 of the helical arm.

3.4.2 Cocrystal structure of EndoS1:Prx complex

Crystal trays were set up at different conditions using full-length EndoS1 and WT Prx to obtain a crystal structure of the EndoS1:Prx complex. However, full-length EndoS1:Prx complex failed to yield any crystals. To overcome this problem, a new EndoS1 construct was generated. It was seen that the N-terminal end consisted of a small, disordered region as predicted by bioinformatic servers. Disordered regions are known to hinder crystallization due to their dynamic and flexible nature[91]. Nine amino acids from the disordered region were trimmed and the new construct was termed (Δ1-9)EndoS1. Before proceeding to setting up crystal trays, first the ability of (Δ1-9)EndoS1 to bind Prx was tested by performing a binding assay as described previously (Figure 8A). After validation of Prx binding, crystal trays for the complex were set up using commercially available crystal screens. The new complex yielded crystals and they were tested for their cryo-conditions and diffraction capability. From this point (Δ1-9)EndoS1:Prx is referred to as EndoS1:Prx. A structure of an EndoS1:Prx complex was solved to a resolution of 2.69 Å, with an Rfree of 0.2576. The data collection statistics are displayed in Table 2.

Table 3: Data collection and refinement statistics for EndoS1:Prx complex.

	Data collection
Wavelength (Å)	1.5418
Space group	P61
Cell dimensions	
a, b, c (Å)	67.494 67.494 232.208
α, β, γ (°)	90 90 120
Resolution (Å)	33.75-2.69 (2.786-2.69)
Total Reflections	1793759 (17965)
Unique reflections	16564 (2073)
CC(1/2)	0.998 (0.652)
R_{merge}	0.141 (0.872)
R_{pim}	0.045 (0.308)
$I/\sigma I$	13.3 (2.2)
Completeness (%)	99.47 (95.82)
Redundancy	10.8 (8.7)
Refinement	
R_{work}/R_{free} (%)	0.2323/0.2576
Average B-factors (Å^2)	58.29
Protein	58.79
Ligands	25.88
Water	46.81
No. of atoms	
Protein	486
Ligand	27
Water	123
Rms deviations	
Bond lengths (Å)	0.014
Bond angles (°)	1.27
Ramachandran plot (%)	
Total favored	97.51
Total allowed	2.49

The cocrystal structure shows that EndoS1 is a homodimer (blue, green), with one Prx (gold) directly bound to each monomer in the globular end domain (Figure 8B). Analysis of the interacting surface reveals that this interaction is mostly electrostatic (Figure 8C). The Prx interface consists of acidic residues that impart an overall negative charge to this region. This interface interacts with a region of EndoS1 that is mostly electropositive due to the presence of basic residues. Several critical hydrogen bonds are seen to be forming at the binding interface. Prx

residues E6 and D32, which were found to be critical for interacting with EndoS1 through SEC binding assays, are seen at this binding interface. Additionally, Prx residue E6 stabilizes one end of the complex by forming salt bridges with EndoS1 residue K28. Prx residue D32 plays a key role by forming the core of the hydrogen bond network (Figure 8D). D32 also forms a salt bridge with EndoS1 residue R69 and this interaction is critical for positioning R69 to form hydrogen bond with Prx residue E38. Additionally, Prx residue K13 is seen to form a hydrogen bond with EndoS1 residue D23. The other end of the complex interface is stabilized by a hydrogen bond formed between Prx residue E44 and EndoS1 residue R65.

Prx bound EndoS1 was also overlayed on unbound EndoS1 (Figure 8). Difference in conformation was seen in the CR regions, indicating that Prx interaction can cause some conformational change in EndoS1. This change may be facilitated by the proline rich linker that can provide flexibility between the arm and the head.

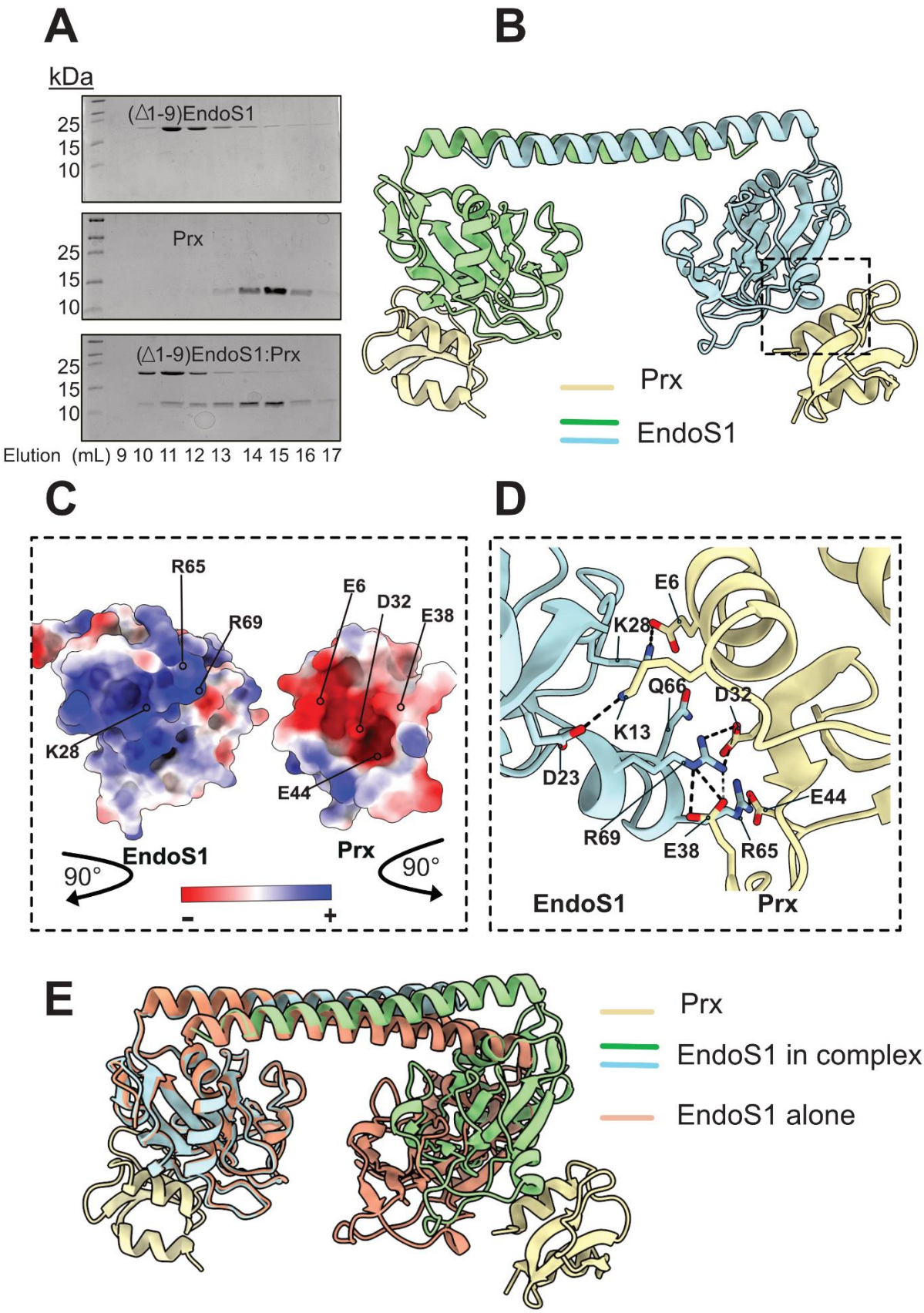

A
kDa
(Δ1-9)EndoS1
25
15
10
Prx
(Δ1-9)EndoS1:Prx
Elution (mL) 9 10 11 12 13 14 15 16 17
B
Prx
EndoS1
C
R65
R69
K28
E6
D32
E38
E44
EndoS1
Prx
90°
90°
-
+
D
E6
K28
Q66
D32
K13
D23
E44
R69
E38
R65
EndoS1
Prx
E
Prx
EndoS1 in complex
EndoS1 alone

Figure 8: Cocrystal structure of EndoS1:Prx complex. **(A)** SDS-PAGE gels of SEC binding assay that confirms (Δ1-9) EndoS1 binds Prx. **(B)** Cocrystal structure of EndoS1 dimer (blue, green) with 2 Prx (gold). Prx directly binds each EndoS1 monomer via the globular head. Boxed area shows the binding interface and is enlarged below. **(C)** Surface analysis of EndoS1:Prx binding interface. EndoS1 uses an electropositive surface to interact with an electronegative surface of Prx. **(D)** Key hydrogen bonds and salt bridges between EndoS1 and Prx residues that stabilize complex formation. (**E**) Alignment of EndoS1 in complex (green and blue) and unbound EndoS1 alone (salmon). The two states have different conformation in the CR regions.

3.4.3 Identification of a conserved helical motif

The EndoS1:Prx cocrystal structure was aligned with a co-crystal structure of ComR-DBD:Prx (PDB ID 7N10) to check for any similarities (Figure 9A). No significant variation was seen in the Prx conformation in both complexes, which supported our binding assays that indicated Prx uses a similar interface to interact with its binding partners. Interestingly, a helical motif was identified in both EndoS1 and the ComR-DBD that served as the interface for Prx binding. Sequence alignment of the helical regions in both proteins identified four residues in EndoS1 (R65, Q66, L67 and R69) that were also conserved at the same positions at ComR-DBD (R33, Q34, L35, R37) (Figure 9A). It is known that some residues (R33, Q34, R37) are strictly conserved in ComR orthologues among different species of *Streptococcus*, and are directly involved in DNA binding[56]. The arginine residues also form salt bridges and hydrogen bonds with several Prx residues and therefore are critical for stabilizing ComR-DBD:Prx complex[38].

To see if a similar helix is observed in P1246 and the potential binding partner SpyM3_0964, an AlphaFold model of both proteins were overlayed with ComR-DBD:Prx structure. Indeed, a

similar helix was observed in both P1246 and SpyM3_0964 (Figure 9B, 9C), indicating that Prx interacts with its binding partners through a helical motif. Interestingly, both proteins are predicted to have a helix-turn-helix DNA binding domain identical in fold to ComR-DBD. However, conservation of all the key DNA binding residues was not observed in P1246 and SpyM3_0964, indicating that the helical motif is only part of the Prx recognition surface.

To test if the conserved residues in the helical region of EndoS1 were also important for forming EndoS1:Prx complex, EndoS1 mutants were generated where residues at position 66 and 69 were substituted to alanine. Binding assay with Prx and EndoS1 variants Q66A and R69A were performed using SEC. The characteristic shift in elution volume seen after a complex formation was observed with Q66A EndoS1:Prx showing that Q66A EndoS1 is capable of binding Prx (Figure 9D). However, it was seen that when R69A EndoS1 was mixed with Prx and loaded on SEC column, the proteins eluted separately and no shift in elution volume was observed (Figure 9E). This indicated that R69 in EndoS1 is critical for forming stable complex with Prx. Overall, these binding assays corroborate the cocrystal structure, which showed that no critical hydrogen bonds were contributed by Q66 in EndoS1 during complex formation (Figure 8D). It is possible that manipulating Q66 led to a change in binding affinity compared to wild type EndoS1, but that quantitative assessment can only made by performing assays like isothermal titration calorimetry. R69, however, sits at the core of the hydrogen bonding network, forming direct bonds with Prx residue E38 and D32. This interaction fixes the EndoS1 conformation that exposes critical EndoS1 residues to the Prx interface, which is needed for successful Prx binding. Overall, these binding assays indicate that both EndoS1 and ComR contain a helical motif that is composed of residues that are important for stable complex formation with Prx. This also indicates that the helical motif in EndoS1 may be part of a DNA binding domain.

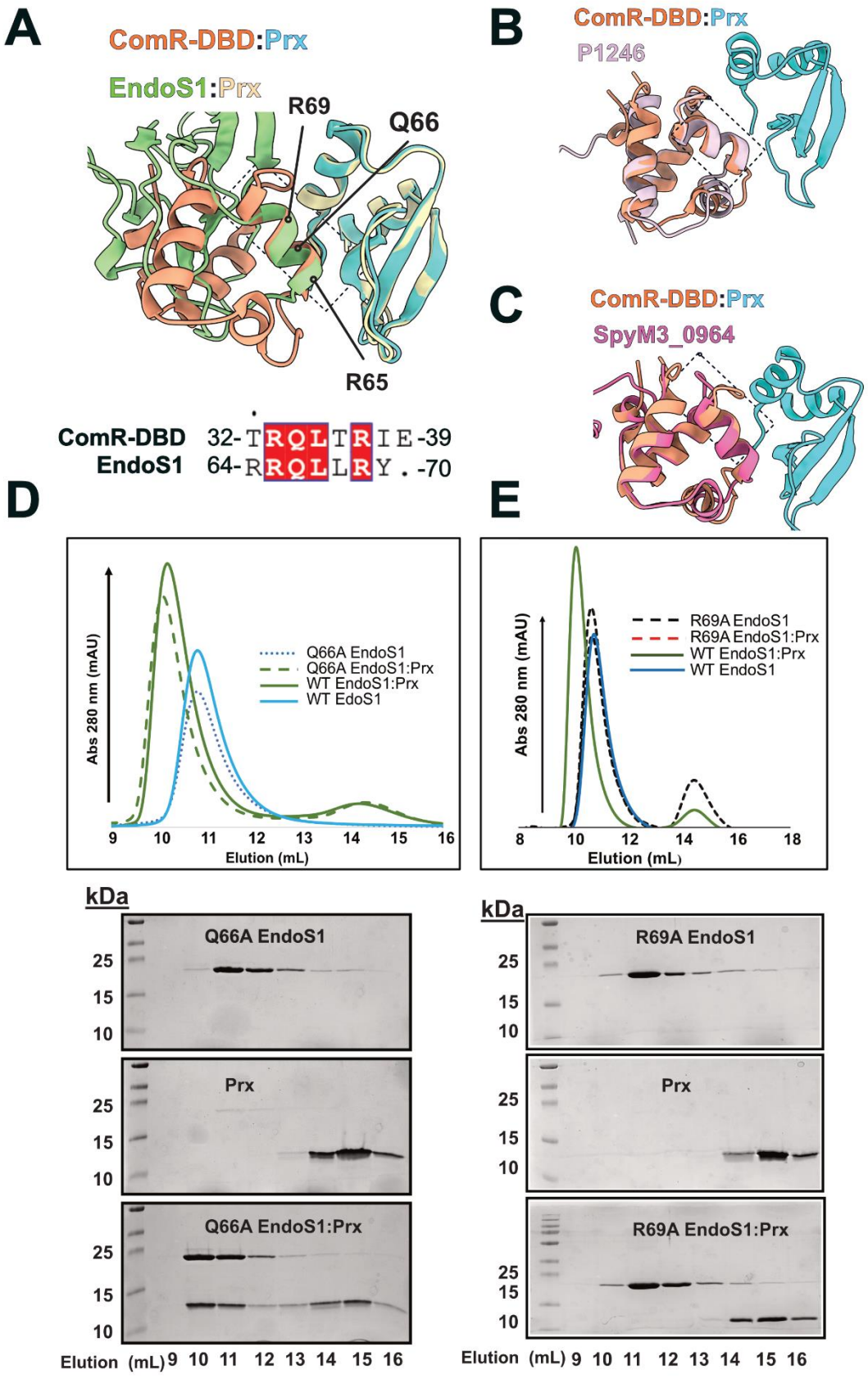

A
ComR-DBD:Prx
EndoS1:Prx
R69
Q66
R65
ComR-DBD 32-TRQLTRIE-39
EndoS1 64-RRQLLRY.-70
B
ComR-DBD:Prx
P1246
C
ComR-DBD:Prx
SpyM3_0964
D
Abs 280 nm (mAU)
Q66A EndoS1
Q66A EndoS1:Prx
WT EndoS1:Prx
WT EdoS1
Elution (mL)
kDa
Q66A EndoS1
Prx
Q66A EndoS1:Prx
Elution (mL) 9 10 11 12 13 14 15 16
E
Abs 280 nm (mAU)
R69A EndoS1
R69A EndoS1:Prx
WT EndoS1:Prx
WT EndoS1
Elution (mL)
kDa
R69A EndoS1
Prx
R69A EndoS1:Prx
Elution (mL) 9 10 11 12 13 14 15 16

Figure 9: A helical motif is required for interaction with Prx. (**A**) Structure alignment of EndoS1:Prx (green and gold) and ComR-DBD:Prx complex (orange and cyan) revealed a helical motif (dotted box) at the Prx interaction interface. Highlighted residues were found to be conserved in both ComR and EndoS1 after sequence alignment. (**B**) Overlay of P1246 (purple) and (**C**) SpyM3_0964 (pink) with ComR-DBD:Prx showed the presence of the helix in both proteins (dotted box). **(D)** SEC binding assay of Q66A EndoS1 (left) and (**E**) R69A EndoS1 (right) with Prx. SDS-PAGE gels of respective assays are shown below. Q66A EndoS1 binds Prx and the complex produces a shift in elution volume (green dotted) which was not observed by R69A EndoS1 (black dotted).

3.5 Investigating biological function of EndoS1

3.5.1 MGAS315 HsdS purification

The crystal structures verify that the EndoS1 dimer closely resembles a completely assembled specificity (S) subunit of type I restriction modification (RM) enzymes. Although the true S-subunit was not detected in the pull-down assays performed for Prx binding partners, the similarity in the structures prompted our investigation on whether or not the S-subunit of MGAS315 also binds Prx. The S subunit of type I RM of MGAS315 is encoded by the *hsdS* gene. *HsdS* was amplified from the genome and cloned in the expression vector pET-22b(+) upstream of a 6His-tag. During purification, it was observed that HsdS was very insoluble, and very little quantity was yielded in the soluble cellular fraction (Figure 10 A). This amount was not sufficient to perform binding assays with Prx. In a physiological environment, the S subunit remains in complex with the methylation (M) subunit, forming 1S-2M, which is a methyl transferase. To see if the solubility of HsdS is dependent on simultaneous expression of M, a co-expression construct was created. The gene encoding for M subunit, *hsdM*, was cloned along with *hsdS* in a pETDuet-1 vector. IPTG

induction of this plasmid led to expression of both proteins. After protein induction and purification, it was seen that both HsdS and HsdM were present in the soluble cell fraction (Figure 10B). This indicated that the HsdS is soluble only when expressed with the M subunit. Due to the complication of purifying HsdS alone, a binding assay with Prx was not pursued. Instead, the predicted structure of the HsdS was analyzed to extract information about potential Prx binding. After overlaying the HsdS AlphaFold with EndoS1, the conserved helix motif that was observed in EndoS1 and ComR, was not observed in HsdS (Figure 10C). As such, we conclude that HsdS does not bind Prx.

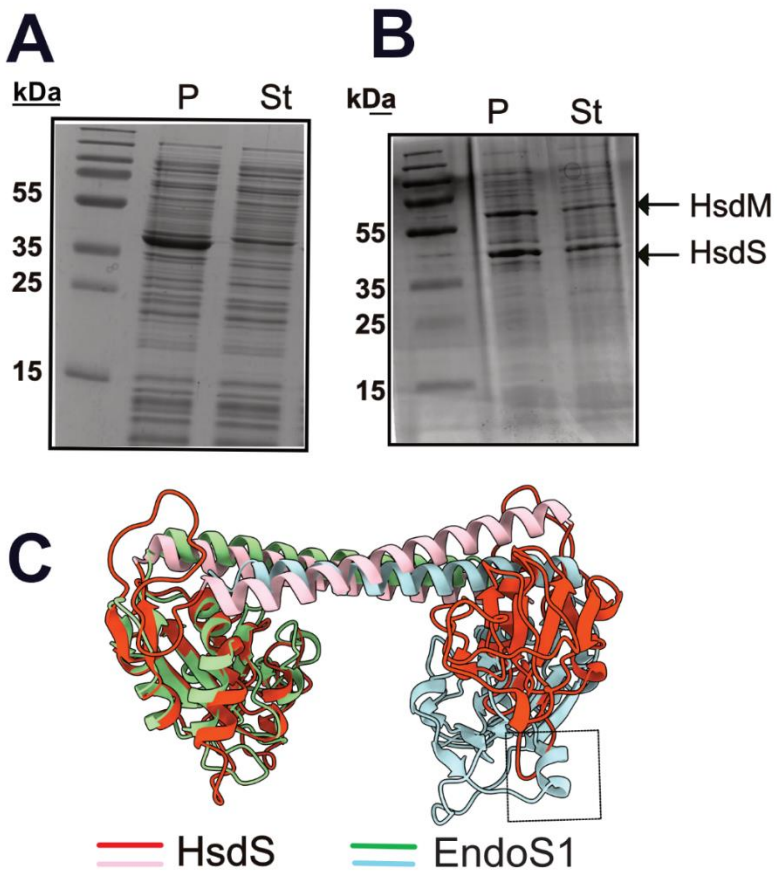


Figure 10: Helix for Prx binding not observed in HsdS. (A) Induction trial of S-subunit (HsdS) of MGAS315. 43kDa HsdS had low solubility and was mostly expressed in cellular pellets (P),

with a small amount in the supernatant (St). **(B)** Induction trial of cells co-expressing HsdS and methylation subunit (HsdM). Higher amount of HsdS and HsdM are seen in the soluble fraction (St) compared to pellet (P), showing HsdS is more soluble when expressed with HsdM. **(C)** HsdS from MGAS315 overlayed (pink and red) on EndoS1 (blue and green). Conserved helix needed for Prx binding (boxed) is not seen in HsdS.

3.5.2 EndoS1 does not bind the methylation (M) subunit HsdM

Normally, the S subunit combines with two methylation (M) subunits to form a methyl transferase (MTase)[92]. To test if EndoS1 also functions like an S subunit, we wanted to first test if EndoS1 interacts with the M subunit. The methylation subunit in MGAS315 is encoded by the *hsdM* gene. The gene was cloned in the expression vector pET-22b(+) (GenScript) and purified for a binding assay with EndoS1. The proteins were incubated together for the intention of loading on a size-exclusion chromatography (SEC) column to see if they form a complex and co-elute. However, incubation of purified EndoS1 and HsdM led to precipitation and SEC binding assay with them could not be pursued. Interaction between the proteins could require unknown cellular cofactors, therefore a co-expression construct for EndoS1 and HsdM (EndoM) was created where both proteins were expressed from a single vector (pET-DUET-1). In this configuration, the proteins would have an opportunity to interact intracellularly compared to *in vitro* incubation as attempted before. The proteins were co-expressed and purified. Both the proteins were soluble; however, co-elution of EndoS1 and HsdM was not observed in either the affinity chromatography step or by SEC, indicating that EndoS1 does not bind HsdM. To see if Prx could assist binding, EndoS1 was incubated with Prx to form an EndoS1:Prx complex, which was mixed with HsdM to perform the binding assay. However, no co-elution of EndoS1 and HsdM was observed (Figure 11A). This led to the conclusion that perhaps EndoS1 does not form part of a functional methyl transferase.

The 1S-2M methyl transferase (MTase) adopts a closed conformation upon encountering DNA, where the two M subunits undergo conformational change to form a clamp like structure. The M subunits are known to contain an N-terminal catalytic pocket and a C-terminal helix. The two helices from two M subunits combine with CRs of the S subunit to form a 4-helix bundle that stabilizes the 1S-2M structure[92]. It was seen previously that the AlphaFold model of HsdS in MGAS315 adopts a different conformation than EndoS1, especially in the helical region (Figure 10C). It cannot be confirmed if that difference is responsible for disrupting the EndoS1:HsdM complex. Since a crystal structure of HsdS:HsdM complex is not available for MGAS315, we looked at the X-ray crystal structure of the DNA bound MTase of *Thermoanaerobacter tengcongensis* (PDB ID 5YBB). The four-bundle helix necessary for S:M interaction could be identified in the structure (Figure 11B). Overlaying the crystal structure with EndoS1 revealed that the CRs of EndoS1 are slightly deviated (Figure 11C). Although this particular MTase belongs to a different species, it still emphasizes the importance of the properly aligned helical regions of the three proteins to form a stable complex. A different conformation adopted by the EndoS1 CRs may not be ideal for forming a 4-helix bundle with the two HsdM subunits.

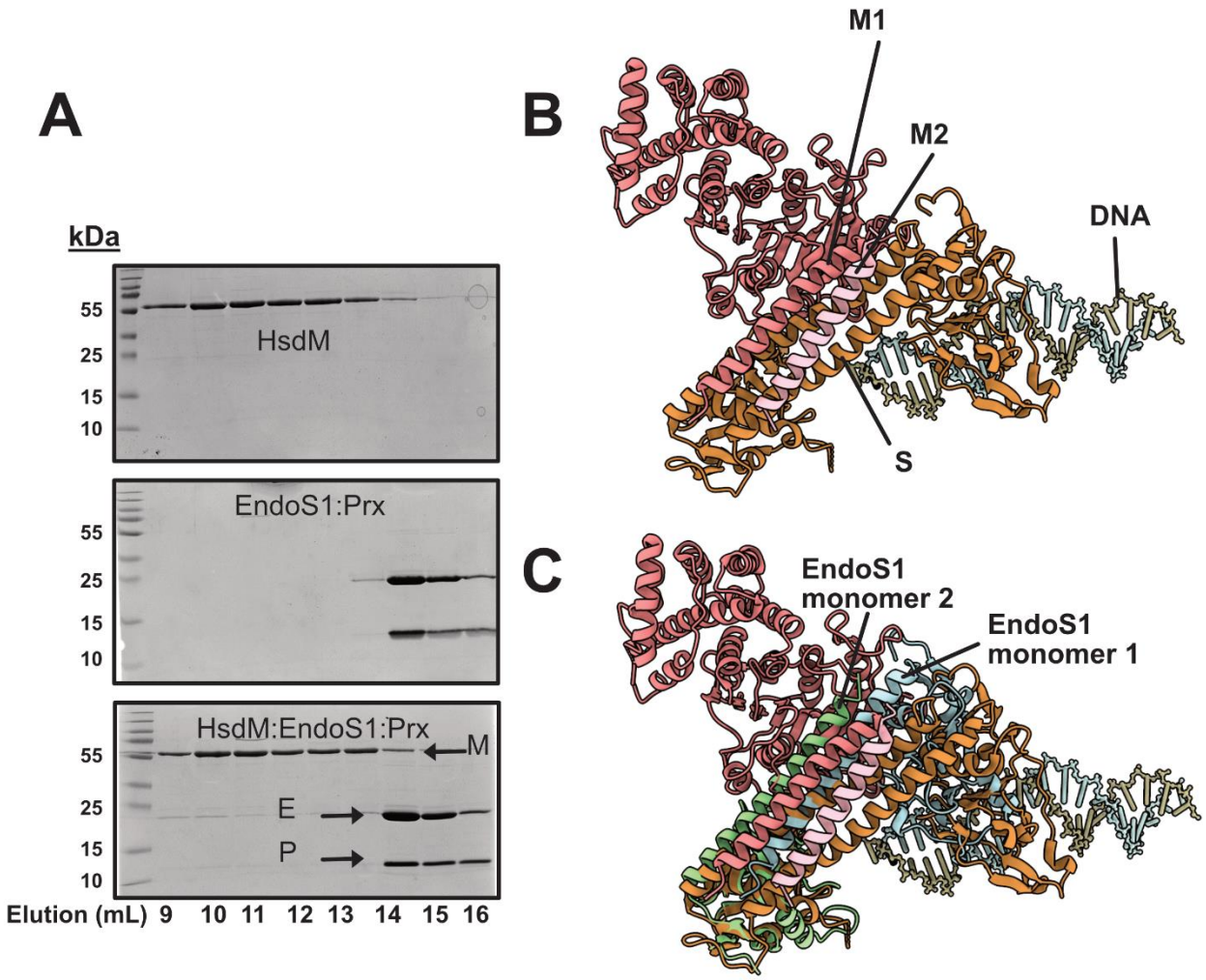


Figure 11: Methylation subunit (HsdM) does not interact with EndoS1. (A) SEC binding assay of HsdM with EndoS1:Prx complex. Co-elution of the three proteins were not observed on SDS-PAGE gels. Position of each protein is marked with arrows. (**B**) Crystal structure of methyl transferase of *Thermoanaerobacter tengcongensis* (PDB ID 5YBB). 2 methylation (M) subunits (coral and pink) interact with one specificity (S) subunit to form a four-bundle helix that stabilizes this interaction. Only a partial structure of M2 is depicted in the complex. (**C**) Structure comparison between the methyl transferase and EndoS1 dimer (green and blue). Alignment of the EndoS1 helix arms may not be sufficient to stabilize interaction with the M subunits.

3.6 Investigating the DNA binding ability of EndoS1

3.6.1 Overexpression of EndoS1 in MGAS315 for ChIP-seq

Since EndoS1 resembles HsdS and HsdS binds DNA, we wanted to test if EndoS1 also binds DNA. For this we decided to do a ChIP-seq (Chromatin immunoprecipitation combined with sequencing) assay. Briefly, in ChIP seq the protein of interest is fused to a tag and expressed in the cells where it has access to the genome. The cells are lysed and sonicated for shearing DNA. However, if the protein binds a specific region of the DNA, that DNA is shielded and protected from shearing. The protein, along with the bound DNA, is then pulled down using an antibody specific for the tag. DNA can be eluted from the protein and sequenced to learn about the DNA site where the protein binds. As the first step, a construct was designed to overexpress tagged EndoS1 with MGAS315 (Figure 12A). For this, pMSP3535, an inducible vector for Gram-positive bacteria was chosen. This vector contained an erythromycin (Erm) resistant gene as a selection marker and a PnisA promoter that could be induced by the addition of nisin (an antibacterial peptide). EndoS1 fused to 6His-tag was cloned downstream PnisA and transformed into MGAS315. The construct was termed EndoS1::3535. To make sure that the MGAS315 transformation was successful, transformants and wild type MGAS315 were grown in liquid media containing 1.5 µg/mL of erythromycin. Growth by transformants and no growth by wild type strains confirmed the presence of EndoS1::3535 in the cells (Figure 12B). To ensure that the EndoS1::3535 encoded for EndoS1, a colony PCR was performed. Briefly, genome from transformed cells were used as template to amplify a portion of EndoS1. PCR and gel-electrophoresis confirmed the presence of EndoS1 expressing plasmid in the transformants. The transformants were also streaked on sheep blood agar to test for their ability to hemolyze. Hemolysis by the transformants confirmed pure MGAS315 culture (Figure 12C). After

verification, EndoS1 containing MGAS315 was grown and induced with nisin. After 6 hours of induction the cell pellets were lysed by sonication. To make sure that over expression of EndoS1 was achieved, a western blot was performed. Antibody against the His tag was used to detect the expression of EndoS1. Unfortunately, 6His-tag, and hence EndoS1 expression from PnisA, was not observed. There are two reasons that can explain this unexpected result: An improper induction time of the MGAS315 by nisin that was insufficient to turn on pNisA and mutation of the plasmid inside *S. pyogenes*. Next attempts for this assay would require optimization of the induction in *E. coli* to see if the plasmid is indeed inducible, and plasmid extraction from the *S. pyogenes* transformants to be analyzed by Sanger sequencing.

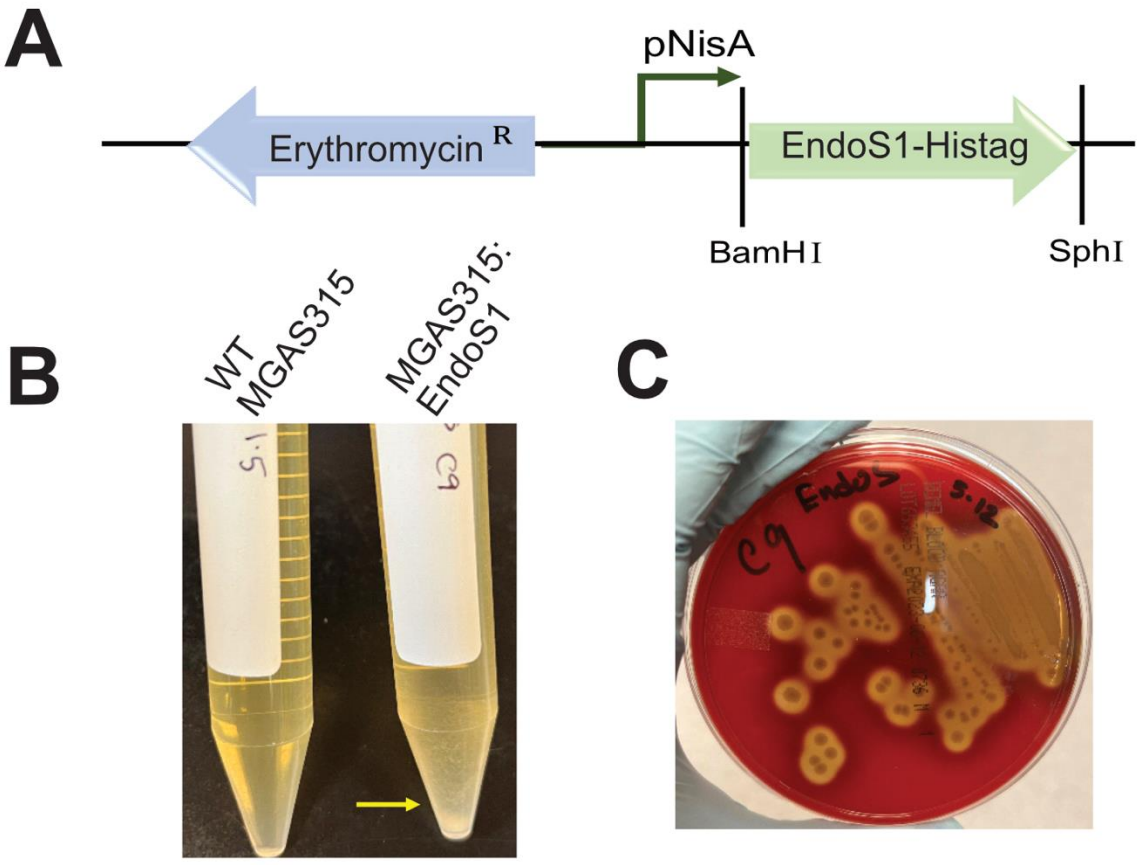


Figure 12: Over expression of EndoS1 in MGAS315. (A) Cloning strategy of EndoS1::3535. EndoS1 with His_6-tag was cloned under a nisin inducible promoter. **(B)** MGAS315 transformed with EnodS1:3535 grew in presence of 1.5µg/mL Erythromycin (right) while wild type MGAS315 did not (left). The cloudy appearance exhibited by growth is shown by an arrow. **(C)** EndoS1:3535 transformants were streaked on sheep blood agar and hemolysis by the strain was observed.

4 Discussion

4.1 Characterization of Prx interaction partners

Prx is a phage protein with a novel fold [41]. *prx* acts as a hot spot for prophage DNA recombination and plays direct role in the dissemination of phage DNA among *S. pyogenes* strains[55]. *prx* is always located beside a phage encoded toxin gene, such that the toxin gene and Prx are always inherited as a single genetic cassette. Although not well understood, this genetic linkage between Prx and toxin genes suggests some evolutionary advantage. Prior to this study, Prx was only known to interact with the transcription regulator ComR to inhibit natural competence and quorum sensing (QS) in *S. pyogenes*[38]. However, the strict location of Prx in the genome hinted that it may interact with multiple proteins to participate in several biochemical pathways. This hypothesis is also supported by the fact that phage proteins evolve to perform multiple biological functions to compensate for small genomic size. Prx shares functional similarity to a phage protein found in *Pseudomonas aeruginosa* termed Aqs1, although the proteins have drastically different folds[58] . Aqs1 is an anti-activator of QS in *P. aeruginosa*, and is known to inhibit LasR, the master regulator of QS. In addition to that, Aqs1 also prevents super infection of *P aeruginosa* by related phages[58]. It achieves this by inhibiting PilB, the type IV pilus assembly protein, and prevents infection by phages that require entry through pili. Aqs1 is a remarkable example of how a single phage protein can interact with multiple proteins to influence key biological pathways.

To find additional Prx binding partners, a pull-down was performed using MGAS315 lysates with Prx as a bait. To capture proteins that might only be expressed by MGAS315 under different conditions, the lysate used for the assays were treated with either XIP (to induce the quorum sensing pathway) or mitomycin C (to induce the phage). Multiple streptococcal and phage proteins were found to be present in the pull-down sample. Presence of multiple proteins immediately indicated that Prx has multiple interaction partners, and hence may be involved in multiple biochemical pathways. For this study, two proteins from the pull-down assay were selected for structural and biochemical characterization. These proteins included SpyM3_0890 (EndoS1) and SpyM3_1246 (P1246). EndoS1 had the highest hit or E value in the mass-spectrometry analysis and closely resembled the specificity (S) subunit of a type I restriction modification (RM) enzymes. Type I RM use the S subunit to bind to target DNA recognition sites. P1246 is a putative phage protein and bioinformatic searches were unable to predict any known biological function for this protein aside from DNA binding.

Because mass-spectrometry could not confirm binding between Prx and these proteins, the first step was to validate this interaction. The genes for each protein were amplified from the MGAS315 genome and cloned in suitable expression vectors for protein purification. Both proteins were purified using affinity chromatography using the affinity tags that were fused to them. Purification of EndoS1 was achieved easily, and a large yield was obtained. P1246 however required some optimization. The protein seemed unstable when cloned with a 6His-tag and had to be re-cloned with an N-terminal GST (glutathione-S-transferase) tag. Even after these modifications, the yields were significantly lower compared to EndoS1. Due to low yields a full biochemical characterization could not be pursued with P1246. For future purification attempts, the protein can be screened with different buffers to identify a composition that can help stabilize it. Construct

modification of P1246 can also help in its stability. For example, instead of constructing the full-length protein, trimming off regions that are predicted to be disordered may also increase stability.

After successful purification of EndoS1 and P1246, they were used in binding assays. Prx and each protein was incubated and loaded over size exclusion chromatography (SEC) columns to test for complex formation. Both EndoS1 and P1246 bound Prx directly, which are indicated by a clear shift in the elution volumes of the individual complexes. This was further confirmed by observing the complexes by Coomassie stained SDS-PAGE gels which showed co-elution of Prx with the proteins. EndoS1 retained its 6His-tag in all binding assays, however P1246 fused to either 6His-tag or GST tag failed to bind Prx. Complex formation with Prx was observed only when P1246 was relieved from its cloning artifacts, and this suggests some directionality needed for Prx binding.

To test the Prx residues that are important for these interactions, binding assays were performed using different Prx point variants. It was seen that Prx residue E6 was critical for interaction with EndoS1 and residue D32 was critical for binding to both EndoS1 and P1246. Prx residues E6, E9 and D32 are also known to be critical for interacting with the DNA binding domain of ComR. Our crystal structures and this overlap in the key residues indicate that Prx is using a similar interface to interact with its binding partners. Isothermal titration calorimetry (ITC) ITC was performed to calculate the binding constant of EndoS1:Prx interaction. Prx binds EndoS1 with a K_d of ~18μM, at a stoichiometry of 1:1. The binding constant of Prx interacting with ComR-DBD is 50 nM, so the Prx binds EndoS1 with a much lower affinity that it binds ComR-DBD. The negative enthalpy (ΔH) and unfavorable entropy (ΔS) of the reaction indicates that the interaction is mostly electrostatic, and both proteins were utilizing their polar/charged interfaces during complex formation.

EndoS1 eluted earlier than expected during purification using size exclusion chromatography (SEC), appearing larger than its predicted size. This indicated formation of higher order oligomers by the protein. Since oligomerization cannot be detected on an SDS-PAGE gel, dynamic light scattering (DLS) and size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) were performed to investigate the molecular weight of EndoS1. Both assays confirmed that EndoS1 existed as a dimer in working buffer conditions, with a molecular weight of 46 kDa. As mentioned above, a monomeric EndoS1 resembled a partially assembled S-subunit. The dimeric EndoS1 resembled a completely assembled S-subunit.

We pursued an X-ray crystal structure of EndoS1 on its own and in complex with Prx. Supporting DLS and SEC-MALS, EndoS1 crystallized as a dimer. Each monomer of EndoS1 consisted of a globular head and helical arm, which were arranged antiparallel to each other. The helical arms are critical for the dimerization as well as the overall solubility of the protein, as the protein was highly insoluble when the globular heads were expressed on their own. A cocrystal structure of an EndoS1:Prx complex revealed that two Prx bind two EndoS1 directly at the globular heads. Surface analysis of the binding interface revealed that EndoS1 uses an electropositive surface to interact with a highly electronegative interface of Prx, which explains the unfavorable entropy observed in ITC before. Conserved Prx residues E6 and D32 are seen at the binding interface, forming essential hydrogen bonds with residues in the EndoS1 head.

X-ray crystal structures of both EndoS1 alone and EndoS1:Prx complex point at the possibility of EndoS1 as a DNA binding protein in several ways. First, the globular heads of EndoS1 neatly aligned with the DNA binding terminal recognition domains (TRD) of the actual S subunit of type I RM of MGAS315, indicating that the globular heads of EndoS1 are putative DNA binding domains. Next, Prx interacts with the globular domains of EndoS1, and Prx is also known to bind

the DNA binding domain of ComR. In addition to that, EndoS1 interacts with Prx using a helical motif that contains residues conserved in both EndoS1 and ComR-DBD. These specific residues in ComR are known to be critical for interacting with DNA[38]. Lastly, the globular heads use an electropositive region to interact with Prx, and DNA binding domains are characterized by the presence of positive charges to counteract the negative charges imparted by the phosphate backbones of DNA. In a normal S subunit, the TRDs have high sequence identity. They recognize non-palindromic bipartite DNA, that is group of nucleotides separated by a stretch of few non-specific nucleotides[78]. The TRDs are interchangeable and a C-terminal truncated HsdS with one TRD is still able to perform like a functional S subunit. In such case, the intact amino terminal half substitutes for the carboxyl half. These truncated versions recognize an interrupted palindrome of the intact N-terminal site. This would be the case with EndoS1 if it were to bind DNA, since both the globular heads are identical.

Overlaying the X-ray crystal structure of EndoS1:Prx complex with that of a ComR-DBD:Prx complex identified a helical motif in both EndoS1 and ComR that was found at the Prx binding interface. Three residues were found conserved at the same locations in EndoS1 and ComR. In EndoS1 these included R65, Q66, R69 and in ComR these were R33, Q34, R37. In ComR these residues form salt bridges with Prx residues and are critical for stabilizing complex with Prx. To test if these residues in EndoS1 are also important for binding, point mutants were created for Q66 and R69. When Q66 was substituted with alanine, EndoS1 still formed a complex with Prx. To see if this mutation had caused any change in binding affinity, a more quantitative assay like isothermal titration calorimetry would be required. When R69 was mutated, binding was abrogated, proving that this residue is critical for binding Prx. As demonstrated by the X-ray crystal structure, R69 forms salt bridges with Prx residues D32 and E38. These are critical for the overall orientation of

EndoS1 that exposes the binding interface to key Prx residues. It was noticed that the helix motif was observed in P1246 as well as SpyM3_0964, another potential Prx binding protein identified in pull-down assay. However, conservation of the key DNA binding residues was not observed in either protein. Moreover, the helix in P1246 was not as electropositive as it was in EndoS1. Overall, this suggests that it is the helical conformation rather an exact amino acid composition that Prx utilizes to interact with its binding partners.

In order to understand the biological role of EndoS1, we tried to investigate if it resembled S subunit functionally as much as it resembled an S subunit structurally. The S subunit of type I RM interacts with the M subunit to form a methyltransferase. A binding assay was performed to see if EndoS1 interacts with the M subunit (HsdM) of MGAS315. Mixing EndoS1 and HsdM *in vitro* lead to precipitation so instead a SEC binding assay was performed with HsdM mixed with an EndoS1:Prx complex. EndoS1 and HsdM were also co-expressed to provide any unknown cellular cofactors that may be necessary for interaction. Despite all attempts, EndoS1:HsdM complex formation was not observed. The conserved helical coils (CRs) of S subunits serve as the major site of interaction with the M subunits[92]. The CRs and the helices from M subunits form a four-bundle helix that stabilize this interaction. Together they form a clamp that opens up to allow access for DNA[65]. Comparing the AlphaFold model of the HsdS of MGAS315 with an X-ray crystal structure of EndoS1 showed a difference in conformation between the helical regions of the two proteins. This difference in conformation may disrupt the four-bundle helix arrangement that is typically seen in 1S-2M methyltransferase structure. Inability of EndoS1 to bind HsdM also indicates that EndoS1 may not share the same function as HsdS in a typical type I RM.

The structural similarity between EndoS1 and HsdS also prompted the investigation of whether HsdS binds Prx. Unfortunately, the experiment could not be performed due to the high insolubility

of HsdS when expressed on its own. Overlaying the crystal structure of EndoS1 and HsdS revealed that HsdS did not contain the helical motif that was observed in all three Prx binding proteins. This led us to conclude that HsdS is not a Prx binding protein. In addition to that, HsdS was soluble only when co-expressed with HsdM. EndoS1 however was stably expressed on its own, further indicating that it may not share functional similarities with HsdS.

The most effective way of determining whether EndoS1 binds DNA would be to perform chromatin immunoprecipitation (ChIP-seq). Sequencing of isolated DNA bound to EndoS1 from this assay will reveal the DNA binding site of EndoS1. This assay requires expression of a tagged protein as the pull-down of the protein utilizes an antibody against that specific tag. Performing the assay in MGAS315 may be of high importance as EndoS1 may require another protein or unknown cofactors to bind DNA. In addition, if EndoS1 binds *Streptococcus* DNA only, performing the assay within MGAS315 will allow access of EndoS1 to its target DNA. To express a tagged EndoS1 in MGAS315, the protein was cloned in an inducible vector that was transformed in MGAS315. This could also be approached in a different way where tagged EndoS1 is inserted into the genome. However, genetic manipulation of MGAS315 is quite challenging, and transformation of EndoS1 containing plasmid has several advantages. This would allow the over-expression of the protein since it was expressed from an inducible promoter and the vector has an antibiotic cassette for selection. The inducible vector containing EndoS1 was transformed in MGAS315, but expression of EndoS1 was not achieved. The strains grew in presence of the specific antibiotic, which meant they were transformed by the plasmid and were also checked for their purity by their ability to hemolyze blood agar. However, the strains were extremely slow growing, which could be a result of the added antibiotic or toxicity of overexpressed EndoS1. The induction conditions need to be optimized further before proceeding with this experiment. Also, it

is possible that the promoter under which EndoS1 was expressed has been mutated, so a full sequencing of the plasmid isolated from the MGAS transformant needs to be performed.

4.2 Potential role of EndoS1 and P1246 in preserving prophage DNA

Temperate bacteriophages are an abundant resident of *S. pyogenes*[54]. They have evolved to harmlessly integrate into their host genome as quiescent "prophages" and replicate with their host before being passed horizontally. These prophages encode for a number of genes that can influence host physiology along with genes that are required for replication and viral assembly. By playing key roles in mediating virulence and antibiotic resistance, they directly contribute to their host survival and fitness. The temperate phages take this complex bacterial-viral interaction a notch higher, where they are known to encode for a plethora of anti-phage defenses for their own protection[93]. This not only prevents host attack by lytic phages, but also work alongside host anti-phage defense to lower the threat of superinfection by phages that could compete for host resources. This also lowers the fitness cost of the host that would be imposed if the host were to encode for multiple anti-phage defenses. Examples of phage encoded anti-phage systems are widespread in both Gram-positive and Gram-negative bacteria. Examples of phage encoded anti-phage defenses include blocking phage adsorption, as seen in *Pseudomonas* phages, blocking phage genome injection as seen by mycobacteriophage Charlie, and repressor mediated silencing of phage genome[93].

By interacting with multiple proteins, it is clear that Prx performs multiple biological roles to influence *S. pyogenes* biology. It is already known that Prx inhibits natural competence to reduce entry of competing foreign DNA, thereby itself acting as a phage encoded protein that protects prophages from invading genetic elements. It is possible that Prx reinforces its protective function by interacting with EndoS1 and P1246.

EndoS1, the first identified Prx binding partner in this study, shares great structural homology with an S subunit. However, it is unlikely that it acts like an S subunit because type I RMs require other subunits for cleaving and methylating DNA. It was seen that EndoS1 does not interact with the M subunit and interaction with the R subunit was not tested. However, EndoS1 appears to retain a structure that could independently bind DNA. Studies have reported phage encoded restriction enzyme like proteins[94], but that does not seem to be the case with EndoS1, which seems to be of streptococcal origin. Proteins upstream and downstream of EndoS1 are predicted to be transporters, and thus are transmembrane proteins. The protein immediately downstream of EndoS1is only a few base pairs away and maybe in the same operon as EndoS1. It is worth investigating if these channels could serve as entry sites of mobile genetic elements which could be substrates for EndoS1 binding, or if as a likely DNA-binding protein EndoS1is involved in their regulating their expression.

P1246 is an uncharacterized protein with no known biological function. In the MGAS315 genome, *p1246* is part of a 5-kb genome cluster that contains multiple open reading frames. All the proteins encoded in that region are of phage origin, indicating P1246 is part of a prophage. The observed helix-turn-helix fold of P1246 that aligns with ComR-DBD also points at its ability to be a DNA binding protein, like a transcription factor or a repressor.

Example of prophage proteins binding other phage proteins are observed in phage inducible chromosomal islands (PICI). PICI are well studied in *S. aureus* and are known to promote bacterial pathogenicity by disseminating genes that encode for toxins and antibiotic resistance[95,96]. These are phage satellites that maintain a close relationship with temperate helper phages. PICIs can encode for anti-phage defenses that can inhibit infection by other competing phages, giving the helper phage an advantage in its niche. In absence of any induction, the PICIs encode for a

repressor that inhibits its replication and mobilization. This repression can be relieved by helper phage proteins that bind to the repressors and induce PICI cycle.

S. pyogenes strain SF370 is known to carry phage-like chromosomal islands[97], but the presence of such islands is not documented in MGAS315. These islands are usually observed in the MMR operon of *S. pyogenes* that contains genes like *mutL* and *mutS* which regulate the DNA mismatch repair. Genome analysis of MGAS315 revealed that *p1246* is not located in the MMR operon and is likely not in a PICI. However, genes surrounding *p1246* are predicted to encode for proteins that can be involved in DNA repair and phage packaging, indicating P1246 could be involved with DNA integration/excision. Given the knowledge that Prx prevents ComR from binding DNA, it is quite rational to assume that it prevents P1246 from binding DNA as well. It will be of interest to study if Prx acts as a helper phage protein and interacts with P1246 to induce or suppress other phage genes. A quantitative analysis of RNA transcription in *p1246* deleted strain could determine if P1246 is regulating the expression of genes in its operon.

5 Conclusion and future directions

Prx is a conserved phage ORF that is strictly found adjacent to phage encoded toxin genes. Prx is known to inhibit quorum sensing and natural competence in *S. pyogenes*. The main goal of this study was to investigate if Prx interacts with other proteins to perform additional biological roles. Such roles could also help explain the link of Prx to toxin genes. Two Prx binding proteins were identified in this study, EndoS1and P1246. Interestingly, both proteins share feature with proteins that are known to be involved in binding DNA.

An X-ray crystal structure of EndoS1 revealed that EndoS1 closely resembled the specificity (S) subunit of type I RM enzymes. Since S subunits bind DNA, the DNA binding ability of EndoS1 could be tested by performing ChIP-Seq analysis.

Since P1246 purification yield was lower and insufficient for biochemical characterization, new P1246 constructs can be designed. Disordered regions can be trimmed off or the protein can be co-expressed with Prx to test if that improved its stability. Once the yield of stable P1246 is achieved, other biochemical and biophysical assays like ITC and X-ray crystallography can be pursued with the protein. P1246 also appears to share the DNA binding helix-turn-helix fold with ComR, so ChIP-Seq assay can be performed with this protein as well.

Lastly, since Prx is known to prevent uptake of exogenous DNA as an act of self-preservation, phage infection assays can be performed in Prx deleted strains to check if Prx deleted strains are more susceptible to infection. Overall, this approach will help understand if Prx restricts the entry of foreign DNA by ways other than inhibiting natural competence.

Appendix

Table A1: Primers used for construct generation in this study

Primer ID	Sequence
p140	ACGTCATATGGAAAATAGCATTGAGGCCTATCTAG
p141	CGTACTCGAGACCTAATCCTTTTTGAATTTCATTCTGAATATATTC
p202	GCTAAGGATCCATGAAAAGACACAGACAGTTTAATAAAGATATTAAATAC
p203	CGATTGCGGCCGCTCAGCATCTCTCCACCTCTCTC
p152	AGGTACCATATGACAAAATCTAAACAGCCTCAATATCGC
p206	CGTACTCGAGTTATACGAATAAACGATTGAGTAATGTTTG
p183	TACAGCGGCCGCTTAGTTTCTAAAAGTATCCAATTCGTCTTG
p207	ATCTGAATTCGATGGCAGAGAAAACAACTTCAC
p211	CCTACGATCCATCTGGAAAGTTC
p212	GCGCTCTTGCGTTACCTTTTGGAAG
p213	CAAGAGCTGCCTACGATCCATC
p214	GCTTACCTTTTGGAAGACGGAGATGTC
p223	GCTAGTGGATCCATGGAAAATAGCATTGAGGCCTATCTAGCAG
p224	ACGTCATATGGAAAATAGCATTGAGGCCTATCTAG

References

1. Lancefield, R. C. A serological differentiation of human and other groups of hemolytic *Streptococci*. *J. Exp. Med.* **57**, 571–595 (1933).
2. Lancefield, R. C. The antigenic complex of *Streptococcus hæmolyticus*. *J. Exp. Med.* **47**, 91–103 (1928).
3. McCarty, M. The lysis of group A hemolytic streptococci by extracellular enzymes of *Streptomyces albus*. II. Nature of the cellular substrate attacked by the lytic enzymes. *J. Exp. Med.* **96**, 569–580 (1952).
4. McCarty, M. Variation in the group-specific carbohydrate of group A *Streptococci*. II. Studies on the chemical basis for serological specificity of the carbohydrates. *J. Exp. Med.* **104**, 629–643 (1956).
5. Henningham, A. *et al.* Virulence Role of the GlcNAc Side Chain of the Lancefield Cell Wall Carbohydrate Antigen in Non-M1-Serotype Group A *Streptococcus*. *mBio* **9**, e02294-17 (2018).
6. McMillan, D. J. *et al.* Updated model of group A *Streptococcus* M proteins based on a comprehensive worldwide study. *Clin. Microbiol. Infect.* **19**, E222–E229 (2013).
7. Schmidt, K.-H., Mann, K., Cooney, J. & Köhler, W. Multiple binding of type 3 streptococcal M protein to human fibrinogen, albumin and fibronectin. *FEMS Immunol. Med. Microbiol.* **7**, 135–143 (1993).
8. Horstmann, R. D., Sievertsen, H. J., Knobloch, J. & Fischetti, V. A. Antiphagocytic activity of streptococcal M protein: selective binding of complement control protein factor H. *Proc. Natl. Acad. Sci.* **85**, 1657–1661 (1988).
9. Thern, A., Stenberg, L., Dahlbäck, B. & Lindahl, G. Ig-binding surface proteins of *Streptococcus pyogenes* also bind human C4b-binding protein (C4BP), a regulatory component of the complement system. *J. Immunol.* **154**, 375–386 (1995).

10. Cunningham, M. W. Pathogenesis of Group A Streptococcal Infections. *Clin. Microbiol. Rev.* **13**, 470–511 (2000).
11. Walker, M. J. *et al.* Disease Manifestations and Pathogenic Mechanisms of Group A *Streptococcus*. *Clin. Microbiol. Rev.* **27**, 264–301 (2014).
12. Carapetis, J. R., Steer, A. C., Mulholland, E. K. & Weber, M. The global burden of group A streptococcal diseases. *Lancet Infect. Dis.* **5**, 685–694 (2005).
13. Brigulla, M. & Wackernagel, W. Molecular aspects of gene transfer and foreign DNA acquisition in prokaryotes with regard to safety issues. *Appl. Microbiol. Biotechnol.* **86**, 1027–1041 (2010).
14. Blokesch, M. Natural competence for transformation. *Curr. Biol.* **26**, R1126–R1130 (2016).
15. Pimentel, Z. T. & Zhang, Y. Evolution of the Natural Transformation Protein, ComEC, in Bacteria. *Front. Microbiol.* **9**, 2980 (2018).
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. Llosa, M., Gomis-Ruth, F. X., Coll, M. & Cruz, F. de la. Bacterial conjugation: a two-step mechanism for DNA transport. *Mol. Microbiol.* **45**, 1–8 (2002).
18. Goessweiner-Mohr, N., Arends, K., Keller, W. & Grohmann, E. Conjugation in Gram-Positive Bacteria. in *Plasmids* (eds. Tolmasky, M. E. & Alonso, J. C.) 237–256 (ASM Press, 2015). doi:10.1128/9781555818982.ch14.
19. Smith, J. M., Dowson, C. G. & Spratt, B. G. Localized sex in bacteria. *Nature* **349**, 29–31 (1991).
20. Dowson, C. G. *et al.* Horizontal gene transfer and the evolution of resistance and virulence determinants in *Streptococcus*. *J. Appl. Microbiol.* **83**, 42S-51S (1997).
21. Dowson, C. G. *et al.* Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci.* **86**, 8842–8846 (1989).

22. Dowson, C. G., Coffey, T. J., Kell, C. & Whiley, R. A. Evolution of penicillin resistance in *Streptococcus pneumoniae*; the role of Streptococcus mitis in the formation of a low affinity PBP2B in *S. pneumoniae*. *Mol. Microbiol.* **9**, 635–643 (1993).
23. Maxted, W. R. & Potter, E. V. The Presence of Type 12 M-Protein Antigen in Group G *Streptococci*. *J. Gen. Microbiol.* **49**, 119–125 (1967).
24. McNeilly, C. L. & McMillan, D. J. Horizontal gene transfer and recombination in *Streptococcus dysgalactiae* subsp. *equisimilis*. *Front. Microbiol.* **5**, 676 (2014).
25. Dinkla, K. *et al.* Rheumatic fever-associated *Streptococcus pyogenes* isolates aggregate collagen. *J. Clin. Invest.* **111**, 1905–1912 (2003).
26. Cue, D., Lam, H. & Cleary, P. P. Genetic dissection of the *Streptococcus pyogenes* M1 protein: regions involved in fibronectin binding and intracellular invasion. *Microb. Pathog.* **31**, 231–242 (2001).
27. Magnuson, R., Solomon, J. & Grossman, A. D. Biochemical and genetic characterization of a competence pheromone from B. subtilis. *Cell* **77**, 207–216 (1994).
28. Otto, M., Süßmuth, R., Jung, G. & Götz, F. Structure of the pheromone peptide of the *Staphylococcus epidermidis agr* system. *FEBS Lett.* **424**, 89–94 (1998).
29. Hui, F. M. & Morrison, D. A. Genetic transformation in *Streptococcus pneumoniae*: nucleotide sequence analysis shows *comA*, a gene required for competence induction, to be a member of the bacterial ATP-dependent transport protein family. *J. Bacteriol.* **173**, 372–381 (1991).
30. Zhang, L., Gray, L., Novick, R. P. & Ji, G. Transmembrane Topology of AgrB, the Protein Involved in the Post-translational Modification of AgrD in *Staphylococcus aureus*. *J. Biol. Chem.* **277**, 34736–34742 (2002).
31. Jimenez, J. C. & Federle, M. J. Quorum sensing in group A *Streptococcus*. *Front. Cell. Infect. Microbiol.* **4**, (2014).

32. Atkinson, S. & Williams, P. Quorum sensing and social networking in the microbial world. *J. R. Soc. Interface* **6**, 959–978 (2009).

33. Ng, W.-L. & Bassler, B. L. Bacterial Quorum-Sensing Network Architectures. *Annu. Rev. Genet.* **43**, 197–222 (2009).

34. Rutherford, S. T. & Bassler, B. L. Bacterial Quorum Sensing: Its Role in Virulence and Possibilities for Its Control. *Cold Spring Harb. Perspect. Med.* **2**, a012427–a012427 (2012).

35. Fontaine, L. *et al.* Quorum-Sensing Regulation of the Production of Blp Bacteriocins in *Streptococcus thermophilus*. *J. Bacteriol.* **189**, 7195–7205 (2007).

36. Johnsborg, O., Eldholm, V. & Håvarstein, L. S. Natural genetic transformation: prevalence, mechanisms and function. *Res. Microbiol.* **158**, 767–778 (2007).

37. 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).

38. 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.* **297**, 100992 (2021).

39. 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).

40. Shanker, E. *et al.* Pheromone Recognition and Selectivity by ComR Proteins among *Streptococcus* Species. *PLoS Pathog.* **12**, e1005979 (2016).

41. 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**, 16535 (2018).

42. Rutbeek, N. Investigating the role of the phage protein paratox in streptococcal natural competence and phage biology. (University of Manitoba, 2022).
43. McShan, W. M. & Nguyen, S. V. The Bacteriophages of *Streptococcus pyogenes*. in *Streptococcus pyogenes : Basic Biology to Clinical Manifestations* (eds. Ferretti, J. J., Stevens, D. L. & Fischetti, V. A.) (University of Oklahoma Health Sciences Center, 2016).
44. Chelala, C. A. & Margolin, P. Evidence that HT mutant strains of bacteriophage P22 retain an altered form of substrate specificity in the formation of transducing particles in *Salmonella typhimurium*. *Genet. Res.* **27**, 315–322 (1976).
45. Schmieger, H. Packaging signals for phage P22 on the chromosome of *Salmonella typhimurium*. *Mol. Gen. Genet. MGG* **187**, 516–518 (1982).
46. Thierauf, A., Perez, G. & Maloy, and S. Generalized Transduction. in *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions* (eds. Clokie, M. R. J. & Kropinski, A. M.) 267–286 (Humana Press, 2009). doi:10.1007/978-1-60327-164-6_23.
47. Chiang, Y. N., Penadés, J. R. & Chen, J. Genetic transduction by phages and chromosomal islands: The new and noncanonical. *PLoS Pathog.* **15**, e1007878 (2019).
48. Hynes, W. L., Hancock, L. & Ferretti, J. J. Analysis of a second bacteriophage hyaluronidase gene from *Streptococcus pyogenes*: evidence for a third hyaluronidase involved in extracellular enzymatic activity. *Infect. Immun.* **63**, 3015–3020 (1995).
49. McShan, W. M., Tang, Y. F. & Ferretti, J. J. Bacteriophage T12 of *Streptococcus pyogenes* integrates into the gene encoding a serine tRNA. *Mol. Microbiol.* **23**, 719–728 (1997).
50. Barnett, T. C. *et al.* Streptococcal toxins: role in pathogenesis and disease: Streptococcal toxins role in pathogenesis and disease. *Cell. Microbiol.* **17**, 1721–1741 (2015).
51. Remmington, A. & Turner, C. E. The DNases of pathogenic Lancefield streptococci. *Microbiol. Read. Engl.* **164**, 242–250 (2018).

52. Oda, M. *et al. Streptococcus pyogenes* Phospholipase A2 Induces the Expression of Adhesion Molecules on Human Umbilical Vein Endothelial Cells and Aorta of Mice. *Front. Cell. Infect. Microbiol.* **7**, (2017).

53. Sitkiewicz, I., Stockbauer, K. E. & Musser, J. M. Secreted bacterial phospholipase A2 enzymes: better living through phospholipolysis. *Trends Microbiol.* **15**, 63–69 (2007).

54. Aziz, R. K. & Kotb, M. Rise and Persistence of Global M1T1 Clone of *Streptococcus pyogenes*. *Emerg. Infect. Dis.* **14**, 1511–1517 (2008).

55. 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).

56. Talagas, A. *et al.* Structural Insights into Streptococcal Competence Regulation by the Cell-to-Cell Communication System ComRS. *PLOS Pathog.* **12**, e1005980 (2016).

57. León-Félix, J. & Villicaña, C. The Impact of Quorum Sensing on the Modulation of Phage-Host Interactions. *J. Bacteriol.* **203**, (2021).

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

59. Meselson, M. & Yuan, R. DNA Restriction Enzyme from *E. coli*. *Nature* **217**, 1110–1114 (1968).

60. Dy, R. L., Richter, C., Salmond, G. P. C. & Fineran, P. C. Remarkable Mechanisms in Microbes to Resist Phage Infections. *Annu. Rev. Virol.* **1**, 307–331 (2014).

61. Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).

62. 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).

63. Wilson, G. G. & Murray, N. E. Restriction and Modification Systems. *Annu. Rev. Genet.* **25**, 585–627 (1991).

64. Gao, Y. *et al.* Structural insights into assembly, operation and inhibition of a type I restriction–modification system. *Nat. Microbiol.* **5**, 1107–1118 (2020).

65. Kennaway, C. K. *et al.* The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Res.* **37**, 762–770 (2009).

66. Kennaway, C. K. *et al.* Structure and operation of the DNA-translocating type I DNA restriction enzymes. *Genes Dev.* **26**, 92–104 (2012).

67. Morgan, L., Jessen, K. R. & Mirsky, R. The effects of cAMP on differentiation of cultured Schwann cells: progression from an early phenotype (04+) to a myelin phenotype (P0+, GFAP-, N-CAM-, NGF-receptor-) depends on growth inhibition. *J. Cell Biol.* **112**, 457–467 (1991).

68. Bickle, T. A., Brack, C. & Yuan, R. ATP-induced conformational changes in the restriction endonuclease from *Escherichia coli* K-12. *Proc. Natl. Acad. Sci.* **75**, 3099–3103 (1978).

69. Yuan, R., Hamilton, D. L. & Burckhardt, J. DNA translocation by the restriction enzyme from *E. coli* K. *Cell* **20**, 237–244 (1980).

70. Davies, G. P. *et al.* On the structure and operation of type I DNA restriction enzymes. *J. Mol. Biol.* **290**, 565–579 (1999).

71. Grinkevich, P. *et al.* Crystal structure of a novel domain of the motor subunit of the Type I restriction enzyme EcoR124 involved in complex assembly and DNA binding. *J. Biol. Chem.* **293**, 15043–15054 (2018).

72. Studier, F. W. & Bandyopadhyay, P. K. Model for how type I restriction enzymes select cleavage sites in DNA. *Proc. Natl. Acad. Sci.* **85**, 4677–4681 (1988).

73. Nye, T. M. *et al.* DNA methylation from a Type I restriction modification system influences gene expression and virulence in *Streptococcus pyogenes*. *PLOS Pathog.* **15**, e1007841 (2019).

74. Kneale, G. G. A Symmetrical Model for the Domain Structure of Type I DNA Methyltransferases. *J. Mol. Biol.* **243**, 1–5 (1994).

75. Gubler, M., Braguglia, D., Meyer, J., Piekarowicz, A. & Bickle, T. A. Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. *EMBO J.* **11**, 233–240 (1992).

76. Bullas, L. R., Colson, C. & Van Pel, A. DNA Restriction and Modification Systems in Salmonella. SQ, a New System Derived by Recombination Between the SB System of *Salmonella typhimurium* and the SP System of *Salmonella potsdam*. *J. Gen. Microbiol.* **95**, 166–172 (1976).

77. Nagaraja, V., Shepherd, J. C. W., Pripfl, T. & Bickle, T. A. Two type I restriction enzymes from *Salmonella* species. *J. Mol. Biol.* **182**, 579–587 (1985).

78. Kim, J.-S. *et al.* Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proc. Natl. Acad. Sci.* **102**, 3248–3253 (2005).

79. Gubler, M. & Bickle, T. A. Increased protein flexibility leads to promiscuous protein-DNA interactions in type IC restriction-modification systems. *EMBO J.* **10**, 951–957 (1991).

80. MacWilliams, M. P. & Bickle, T. A. Generation of new DNA binding specificity by truncation of the type IC EcoDXXI *hsdS* gene. *EMBO J.* **15**, 4775–4783 (1996).

81. Nagaraja, V., Shepherd, J. C. W. & Bickle, T. A. A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. *Nature* **316**, 371–372 (1985).

82. Price, C., Lingner, J., Bickle, T. A., Firman, K. & Glover, S. W. Basis for changes in DNA recognition by the Type I DNA restriction and modification enzymes. *J. Mol. Biol.* **205**, 115–125 (1989).

83. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci. Publ. Protein Soc.* **30**, 70–82 (2021).

84. Zhang, X. & Studier, F. W. Multiple Roles of T7 RNA Polymerase and T7 Lysozyme During Bacteriophage T7 Infection. *J. Mol. Biol.* **340**, 707–730 (2004).

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

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

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

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

89. Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363–371 (2009).

90. Lupas, A. N., Bassler, J. & Dunin-Horkawicz, S. The Structure and Topology of α-Helical Coiled Coils. *Fibrous Proteins Struct. Mech.* **82**, 95–129 (2017).

91. Oldfield, C. J. *et al.* Utilization of protein intrinsic disorder knowledge in structural proteomics. *Biochim. Biophys. Acta* **1834**, 487–498 (2013).

92. Liu, Y.-P. *et al.* Structural basis underlying complex assembly and conformational transition of the type I R-M system. *Proc. Natl. Acad. Sci.* **114**, 11151–11156 (2017).

93. Bondy-Denomy, J. *et al.* Prophages mediate defense against phage infection through diverse mechanisms. *ISME J.* **10**, 2854–2866 (2016).

94. Dedrick, R. M. *et al.* Prophage-mediated defence against viral attack and viral counter-defence. *Nat. Microbiol.* **2**, 1–13 (2017).

95. Fillol-Salom, A. *et al.* Bacteriophages benefit from mobilizing pathogenicity islands encoding immune systems against competitors. *Cell* **185**, 3248-3262.e20 (2022).

96. Martínez-Rubio, R. *et al.* Phage-inducible islands in the Gram-positive cocci. *ISME J.* **11**, 1029–1042 (2017).

97. 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).