THE UNIVERSITY OF CHICAGO

FUNCTIONS OF CONSERVED SCCMEC PROTEINS: A MOBILE ELEMENT INVOLVED IN ANTIBIOTIC RESISTANCE SPREAD

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON MICROBIOLOGY

BY

ALEXIS A. THOMAS

CHICAGO, ILLINOIS

AUGUST 2023

TABLE OF CONTENTS:

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

The work presented in this thesis was completed with the assistance and support of many people within and outside of the Rice lab. First and foremost, I would not have been able to complete the following work without the support and mentorship of Dr. Phoebe Rice. Thank you for all the support, advice, and flexibility over the years. I would like to thank members of the Rice lab: Ying Pigli, Dr. Heewhan Shin, Dr. Ignacio Mir Sanchis, and Dr. Alexsandra Bebel. I have enjoyed my time in the Rice lab the last 7 years and grown a lot as a person and scientist. I would like to thank all the Rice lab members past and present for all their support.

I would like to thank my thesis committee, Dr. Dominique Missiakas, Dr. Aretha Fiebig, and Dr Sampriti Mukherjee. Their advice and guidance has been instrumental in completion of my thesis work especially towards the end. I would especially like to thank Aretha, as she has been tremendously helpful with RNAseq analysis and experimental design.

I would like to thank all my friends that I have made in graduate school. It's been a long a crazy road especially with a pandemic in the middle but we made it through and I have appreciated all the support. I would like to thank my roommate of 5 years Christina Filipaki for all of her support. She has always listened to all my venting over the years and let me in the apartment when I have forgotten or lost my keys.

I would like to thank my family for all their support throughout not just graduate school, but my entire life. While they may not have understood what I was talking about they were still always available to listen and offer advice or serve as a distraction. When things got hard it was always reassuring to know that everyone was just a phone call or short car ride away. I would like to thank my mother especially for always believing in me and being there every step of the way.

I would not have been able to complete the work contained in this thesis without the support of my advisor, labmates, committee member, friends, and family. Everyone above thought I could finish even when I thought I couldn't and for that I will always be thankful.

Abstract

Methicillin resistant *Staphylococcus aureus* (MRSA) is a global public health concern. The mobile genetic element responsible for the transition from methicillin sensitive to methicillin resistant *S. aureus* is Staphylococcal Cassette Chromosome *mec* or SCC*mec*. Little is known about the mechanisms employed by SCC elements to be maintained within the bacteria and mobilize to new cells. While SCC elements carry a variety of virulence factors as accessory proteins, to further understanding of how SCC elements are maintained and transferred, I have focused on the conserved proteins of SCC elements. The goal of this work has been to gain a better understanding of the functions of the SCC element conserved proteins. By understanding the function of these conserved proteins, we can then extrapolate how they may be utilized in maintenance and transfer.

Each SCC element encodes 7 conserved proteins grouped into 3 adjacent operons. The first operon of SCC elements contains proteins with functions related to DNA replication. I developed a plasmid - based assay to test whether or not these proteins could facilitate extrachromosomal DNA replication. Although they did not, the optimized assay will be useful in the study of other mobile genetic elements. The second operon encodes the site-specific recombinases that facilitate excision and integration into the chromosome. Lastly little was known about the 3rd conserved operon except that one of the encoded proteins, SaUGI, inhibits uracil DNA glycosylase. I showed that SaUGI has additional activities: overexpression of it led to a decrease in biofilm production, cell wall defects, and many changes in gene expression. Additionally, I identified Ccg as a uracil DNA glycosylase. Finally, I discuss potential roles for the conserved proteins in the horizontal transfer and how they fit within a new model of transfer via natural competence.

Chapter 1: Introduction

Staphylococcus aureus is a common skin bacterium that can cause infections ranging from minor skin and soft tissue infections to bacteremia [1,2]. *S. aureus* colonizes the skin and nasal passages of an estimated 32.4% of people in the United States[3]. With many people being colonized by *S. aureus* it has become important to understand the factors that influence infection severity and impact treatment. *S. aureus* is an opportunistic pathogen as many people can be colonized and never get an infection. There are multiple factors that tip the balance between colonization and serious infection, including the health of the colonized individual and the route of infection[2]. In relation to the infected individual, conditions that impact the body's immune system make an individual more susceptible to infection. As the bacteria colonizes the skin, abrasions that then allow the bacteria to then enter the body can lead to infections. The bacteria can also form a biofilm on medical equipment that is inserted into the body leading to infections as well. For these reasons severe infections frequently occur in a hospital setting[1]. Another important factor the balance between colonization and serious infection is the virulence of the bacteria.

The virulence of *S. aureus* varies due to mutations and acquisition of mobile genetic elements(MGE)[4]. These mobile elements range from transposons, plasmids, *Staphylococcus aureus* Pathogenicity Islands (SaPIs), and Staphylococcal Cassette chromosome (SCC) elements[4]. These various elements can obtain, toxins, resistance genes and genes important for their own movement[4]. These MGEs can be obtained from other *Staphylococcal species* or other bacteria through the environment or during the colonization of people/animals[4]. For example there are multiple methods by which *S. aureus* can become resistant to vancomycin, but one method is through obtaining an MGE from *Enterococcus faecalis*[5]. To combat infections and

prevent the spread of antibiotic resistance and virulence factors it is important to understand how these mobile elements are transferred and maintained within the bacteria, in the environment, and during infections.

Infections reported to be caused by *S. aureus* are frequently described based on their antibiotic resistance. For example, vancomycin resistant *S. aureus* (VRSA), methicillin sensitive (MSSA), and methicillin resistant *Staphylococcus aureus* (MRSA) are common designations given to clinical strains. Strains that have been found to have more mobile DNA incorporated into their genome have been shown to be more pathogenic in general and can infect those with a more robust immune system versus strains that have fewer mobile elements, which tend to infect those in a hospital setting who are more vulnerable[6–9]. The more virulent MRSA strains tend to be community-acquired (CA-MRSA) while healthcare-associated/hospital acquired (HA-MRSA) tend to be weaker[8].

The rise of MRSA

Along with the introduction of antibiotics came the occurrence of antibiotic resistant bacteria. The first MRSA strain was detected in the UK in 1961[10]. MRSA strains were soon discovered in hospital settings across the world. By the 1970s the growing number of HA-MRSA strains began to cause many problems. These strains were overall less pathogenic and caused problems due to the abundance of immune compromised individuals in hospital settings. Over the last 20 years there has been a growing number of strains isolated from people that are not hospitalized, CA-MRSA[11,12]. These CA-MRSA strains can infect affect those who are generally healthy and can be multi-drug resistant. In addition to the growing number of multidrug resistance *S. aureus* (MDR) infections, antibiotic resistant *S. epidermidis* has been increasingly isolated from infected wounds[13,14]. This is important as *S. aur*eus and S. *epidermidis* both are

part of the normal flora and can share mobile DNA. Over time the line between the different MRSA designations has become blurred as strains continue to share mobile DNA, the location where people have acquired their MRSA infections is undetermined, and as CA-MRSA strains spread through hospital settings[15,16]. The thing that all these MRSA and many pathogenic *S. epidermidis* strains have in common is SCC*mec*.

What are SCC elements?

SCC elements are a family of large variable mobile elements present within many *Staphylococci species*. SCC elements were first identified in 1999, as it was found that the region that differed between MRSA and MSSA was a genomic island[17,18]. This genomic island was identified as Staphylococcal Cassette Chromosome or SCC*mec*. While SCC elements were described in 1999, *mecA* had been previously described, and was known to the effector of methicillin resistance[13,19]. PBP2a, encoded by *mecA,* is a penicillin-binding protein that is not inhibited by beta-lactam antibiotics[13,19]. Penicillin-binding proteins are involved in cell wall synthesis by creating crosslinks[20]. PBP2a presence allows for a functional cell wall to be formed even in the presence of beta-lactam antibiotics[19,21,22].

Despite their variability it was noted early on that SCC*mec* elements always integrate into *rlmH* (orfX) and that they encode site specific serine recombinases with specificity for that locus[18]. As the main factor that led to their discovery was the *mecA* locus which encodes methicillin resistance, SCC*mec* elements are typed based on the recombinase genes and the *mecA* locus that is present[23]. As more SCC containing strains were isolated and sequenced the number of types grew as did the variety of virulence factors found to be encoded within the elements. SCC elements were found to not only contain the *mecA* locus but also to be a hotspot for IS element, transposon, and plasmid insertion[24]. As more strains of pathogenic and non-pathogenic

coagulase negative Staphylococci (CoNS) were sequenced it was observed that not all SCC elements contained the *mecA* locus[24,25]. These new SCC elements were then annotated throughout the literature based on what was integrated in the element - for example SCC*mer* contains a mercury resistance cassette[26]. These accessory proteins are important to study for their impacts on infection outcomes and bacterial survival in different environments, however, little is known about the features of SCC elements that regulate their maintenance or mobility. To better understand the mechanisms involved in movement and maintenance we sought to characterize the conserved genes encoded in all SCC elements.

How might SCC elements function and move?

With any mobile element, especially one known for carrying antibiotic resistance genes, a key question is: how do they move?. This is important to understand as stopping the mobilization of the mobile element would stop the spread of said antibiotic resistance. In this respect very little is known about SCC elements. Sequencing of clinical isolates has revealed that the elements are mobile as identical elements have been found in very different strains[24,27]. However, despite their abundance and intense study, very little is known about the mechanism of transfer and what triggers it. A major problem that has plagued study of SCC elements is that until recently researchers had not been able to observe movement in the lab. To further explore how SCC elements move we have looked at the mobilization mechanisms of other mobile elements. The main mechanisms utilize by mobile elements in *S. aureus* is through conjugation, phage transduction, and transformation[4,28].

Plasmids frequently utilize conjugation as a method of transfer and can either encode all the genes needed for transfer or can become mobilizable[29]. The proteins needed for conjugative transfer are the *tra* gene or transfer[29,30]. These proteins are involved in transferring the plasmid

DNA from 1 cell to another. In order for the DNA to be transferred a pore is created in the donor and recipient cell. A relaxase protein binds the origin of transfer or *oriT*. At the *oriT* the DNA is cleaved, and the single strand is covalently bound to the relaxase protein and brought to the transfer pore. The protein DNA complex is then transferred to the new cell. The relaxase is involved in replicating the plasmid in the donor cell and is involved in recircularization and replication of the second strand in the recipient cell[29,30]. Not all plasmids that transfer via conjugation encode all the proteins needed for transfer. There are plasmids that are mobilizable and transfer using the transfer proteins of conjugative plasmids[29,31]. These mobilizable plasmids contain a an *oriT* and may or may not contain a relaxase protein, Mob. While the relaxase protein and other transfer proteins are not required for conjugation to occur, *oriT* is indispensable. Additionally for a mobile element to be mobilizable, a corresponding conjugative plasmid is needed. Conjugation is a prominent method of MGE transfer that is utilized by not only conjugative plasmids, but mobilizable plasmids, transposons and insertions sequences (IS elements) that integrate within these plasmids, and integrative conjugative element or ICE elements. Similar to conjugative plasmids ICE elements encode their own transfer machinery[32].

SCCs carry no conserved motifs for conjugation, which as stated above, would require an *oriT* and transfer proteins. Some SCC elements have mobilizable plasmids inserted within them, but they are not conserved among all SCCs and thus do not provide a universal explanation for SCC mobility. One study found that they could achieve transfer via conjugation. However, extensive selection was required to obtain a conjugative plasmid donor carrying an SCC element[33]. This was done by utilizing the fact that SCC elements frequently contain IS431. IS431 is an insertion sequence element that has been observed in multiple *S. aureus* plasmids and in the chromosome[34]. In SCC elements IS431 has been observed flanking the integrated plasmids

indicating that it likely facilitated the integration of these plasmids into SCC*mec*. Ray et al achieved movement of SCCmec by utilizing a conjugative plasmid that also contained IS431 and expressing the SCC recombinase proteins from a plasmid. The recombinases excised the SCC element from the chromosome while homologous recombination between copies of IS431 inserted it into the conjugative plasmid, which was then transferred via conjugation to a new cell. While this was interesting and could happen in nature theoretically, the frequency would be very low. This is because expression of the recombinase genes occurs only sporadically (in 1-3% of cells)[35–37]. This sporadic expression would need to align with the presence of a conjugative plasmid containing IS431 and an SCC element that also contains IS431. Furthermore, there are no reports of plasmid-borne SCC elements. Therefore, I believe that is likely not a common mechanism of transfer.

The second mode of MGE transfer is through the use of helper phage. For phage transduction to occur, a bacterium is infected with the bacteriophage. The phage DNA is inserted into the new cell and phage proteins are made, or in the case of lysogenic phages, the phage DNA is integrated into the chromosome. Fragments of phage, chromosomal, and plasmid DNA are then packaged into the phage capsid and inserted into a new cell. This is frequently used in the lab to move mutations between strains via generalized transduction. In these cases MGEs are moved serendipitously, however there are some mobile elements that parasitize the phages and are preferentially transferred via phage. These are phage inducible chromosomal islands or PICIs.

One family of PICIs, that are similar to SCC elements, are SaPIs. SaPIs are genomic islands that are induced and transferred via a phage[38,39]. The expression of SaPI genes are regulated by the SaPI encoded repressor that is derepressed by a phage encoded protein[40]. PICIs

like SaPIs, are difficult to study because one must find the correct helper phage[39]. SCC elements may be a type of PICI and require a helper phage. A former post-doc in the Rice lab, Ignacio Mir-Sanchis, worked to screen many different phages to determine if they were able to transfer SCC*mec*, but was not successful. Because SCC elements are found in different *Staphylococcal species* and can be very large, a helper phage would need to have a broad host range and have a large capsid to accommodate the larger elements[25]. There have been papers that have shown small SCC*mec* elements or simply methicillin resistance transfer via phage[41–44]. This is likely through generalized transduction and did not require recombinase expression.

The final method of MGE transfer is transformation. This would be through natural competence. Although *S. aureus* has the genes to become naturally competent, natural competence has been difficult to observe occurring and was believed to not play a major role in MGE transfer in *S. aureus*. However, studies have shown that natural competence is controlled by the sigma factor sig H[45–47]. A recent study showed that by growing *S. aureus* in a biofilm, this activated natural competence[48]. Activation of natural competence in the biofilm, then facilitated the transfer of an SCC element to a new cell in a recombinase dependent manner[48]. The necessity of the recombinases is what highlighted this as the likely the major pathway for SCC element transfer. The work shown throughout this thesis was done prior to this study being published, however this new development provides a new lens to interpret the results, explains some of the issues faced, and provides new direction for future experiments.

Due to the issues faced with getting SCC elements to move in the lab, the focus of my thesis work has been to understand how the SCC conserved proteins function within the cell. We hypothesized that by understanding how these proteins function we could then extrapolate what the mechanism of transfer may be.

SCC conserved genes:

As stated previously SCC elements contain site-specific serine recombinases (CcrAB or CcrC) that catalyze integration and excision of the element at the *rlmH* locus. Further investigation into the genes surrounding these recombinases led to the discovery of 7 conserved SCC genes[49]. Analysis of the many sequences allowed for the simplification of the many SCC types to classifying them into 2 patterns (Figure 1)[49]. Further work was then done based on the assumption that anything that would be important for SCC maintenance and movement would be in this conserved gene region.

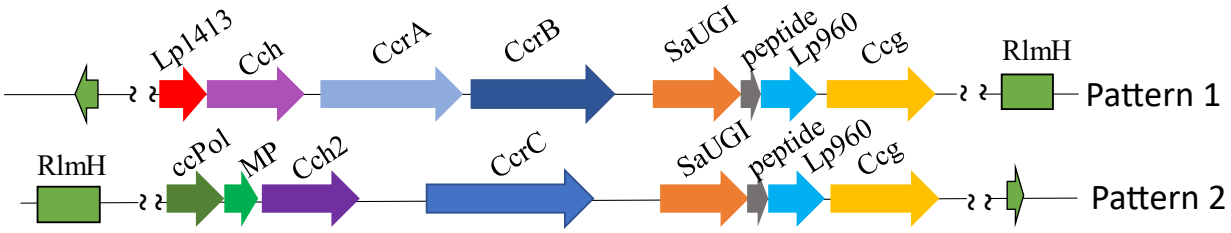


Figure 1.1: Schematic of conserved SCC proteins.

The conserved genes of both patterns are grouped into 3 operons: the first encodes proteins likely to be involved in DNA replication, the second encodes the site-specific recombinases responsible for integration and excision of the element, and the third contains encodes a set of proteins of previously unexplored function.

Pattern 1 SCC elements encode a single-stranded (ss) DNA binding protein (Lp1413) in the same operon as a helicase, Cch[49,50]. The next set of genes are the recombinases, CcrAB. The last operon in the conserved gene region is found in both patterns of SCC elements and contains 4 ORFs: SaUGI (a uracil DNA glycosylase inhibitor), a peptide, Lp960, and Ccg (a DNA

glycosylase)[51]. Pattern 2 elements differ in that the first operon contains a primase (ccpol-MP) and a different helicase, Cch2. Pattern 2 SCC elements only have one recombinase, CcrC.

The first operon varies between the two patterns but they both contain a helicase. Work done by other lab members has shown that the both of these helicases actively unwind DNA. [49,52]. The presence of these helicases was important as structurally similar helicases from SaPIs have been shown to facilitate replication of the element[49,52,53]. Although sequence-specific DNA binding is rare among helicases, the SaPI helicases bind to their own origin of replication, and Cch2 preferentially binds the direct repeats in the intergenic region between the first two operons of pattern 1 SCC elements. A single stranded DNA binding protein, Lp1413, is co-expressed with Cch, and might be expected to protect the single-stranded DNA created by helicase action. Cch2 is instead co-expressed with Ccpol and MP which together act as a primase[49,50,52]. The presence of an active helicase, primase, and / or ssDNA binding protein lead us to believe that SCC elements can replicate. This is especially true for pattern 2 SCC elements as they contain a helicase, primase and direct repeats that the helicase will bind to and thus may mark an origin of replication. This architecture had been previously shown to be sufficient to replicate SaPIs[53].

SaPIs replicate using a replicative helicase, Rep, and a co-expressed primase, Pri. When expression is activated the Rep protein binds the repeats downstream of the rep gene and opens the replication bubble. Pri binds SaPI Rep and makes RNA primers using single-stranded DNA as a template (Heewhan Shin and Phoebe Rice, unpublished data). It is believed that Rep recruits a host DNA polymerase which will then extend those primers to replicate the SaPI DNA. The specifics of SaPI replication are still being worked out, however the presence of long concatemers of SaPI DNA that are cleaved and packaged into phage capsids indicates that it utilizes rolling-circle replication[53,54]. In this case SaPI Rep protein acts like a multi-functional

replication initiator that recognizes the origin, opens up the replication bubble and unwinds the DNA. It has been shown that SaPI Rep-Pri and the repeats that are the replication origin are sufficient to maintain a plasmid in *S.aureus*[53,55]. We hypothesized that SCC elements function in a similar manner to the SaPIs when they are excised from the chromosome. Cch2 would bind the direct repeats and open a replication bubble, then Ccpol-MP would make primers and the host DNA polymerase would then replicate the excised SCC element. As pattern 2 elements have the primase and specific binding of the helicase to direct repeats in the intergenic region, a pattern 2 element was the focus for replication experiments described in chapter 3. I was unable to detect replication and we think that SCC elements may replicate at very low levels or not at all. Now that transfer has been observed via natural competence, we believe that the putative replication proteins may not be involved in replication in the donor cell but instead is involved in producing the second strand in the recipient cell by recruiting the cellular polymerase and in the case of pattern 1 SCC elements a primase.

Downstream of the helicases are the recombinase genes. These proteins act together as a heterotetramer (CcrAB) or a homotetramer (CcrC) to integrate and excise SCC elements from the chromosome[56,57]. These site-specific large serine recombinases bind specific DNA sites termed attSCC and attB (in *rlmH*) to facilitate integration into the chromosome[56,57]. Conversely they recognize and bind integration product sites, attL and attR, leading to excision from the chromosome[56,57]. It has been observed that overexpressing the recombinase genes from a plasmid leads to loss of the element. Therefore simply expressing the recombinase proteins is all that is needed for excision and integration to occur[58]. Most large serine recombinases utilize a directionality factor to cause either integration or excision, however for the SCC elements it appears as though none is needed as excision and integration have been observed without

one[56,59]. As there appeared to be no directionality factor and expression seemed to be the driving force for recombinase function, I explored the regulation of *ccrAB* expression (chapter 2).

Chapter 4 examines the function and purpose of the proteins in the last operon. This operon is the same between the 2 patterns of SCC elements. Little is known about this operon. The first protein in this operon, SaUGI, has been previously crystallized and was shown to bind the *S. aureus* native uracil DNA-glycosylase, UDG[51]. UDG is involved in base excision repair and helps prevent mutations by removing uracil that results from spontaneous deamination of cytosine or is mis-incorporated by DNA polymerase[60]. SaUGI binds the DNA binding region of UDG[51] preventing this activity. The second gene encodes a previously undescribed peptide and the third encodes a protein with domain of unknown function 960 that we termed Lp960. There is no predicted function for Lp960. The last protein in the operon was annotated as containing domain of unknown function 1643. I found that it has uracil DNA glycosylase activity and therefore named it Ccg (Cassette Chromosome Glycosylase). Due to so little being known about the proteins in this operon in relation to how they are benefitting SCC elements or their hosts we examined their effect on *S. aureus* by overexpressing them. This led to many gene expression changes most notably upon SaUGI overexpression.

A lot is still unknown when it comes to the function of SCC elements and their conserved proteins. The following study aims to gain a better understanding of the functions of these conserved proteins. I was able to isolate mutants within the promoter of *lp1413/cch* which leads us to believe the promoter is weak. While Cch and Cch2 are homologous to known replicative helicases, I was unable to observe SCC replication[49,52]. Lastly we show that Ccg is a DNA glycosylase and that overexpression of SaUGI leads to many changes in gene expression and a loss of biofilm. These results taken with revelation that SCC elements can be transferred in a

biofilm due to natural competence leads to a new hypothesis for how the element is functioning and the role that the conserved proteins are playing in this transfer and maintenance into a new host cell. This will be discussed in chapter 5.

SCC elements have been found increasingly in strains that have been isolated from infected patients and in those found in the environment. Their abundance and importance in antibiotic resistance spread makes understanding their function important for the fight against antibiotic resistance spread. Additionally, from an intellectual standpoint, these elements don't appear to follow the paradigms established for other mobile genomic elements. This makes them both interesting and difficult to study.

Chapter 2: Regulation of Conserved SCC genes

2.1 Introduction

SCC conserved genes are not expressed under standard lab conditions[35,37,61]. The lack of gene expression and movement in the lab has made genetic analysis very difficult. It is for these reasons that my goal was to determine what regulates gene expression with the idea that by determining what naturally turns these genes on/off, we could then manipulate the system to turn the genes on and then observe excision and hopefully movement.

Previous gene expression work has focused on the recombinase genes, *ccrABC*. In many strains *ccrABC* have "stochastic" expression: they are expressed in about 3% of cells in early log phase and less than that in stationary phase[61]. One group found that disruption of an inverted repeat that is upstream of *ccrAB*, led to an increased expression of a beta-galactosidase reporter[35]. I was unable to repeat these results and we hypothesize that the inverted repeat may act as a transcriptional terminator for *cch* and that their results were possibly due to transcriptional read-through. Other studies have argued that LexA, SarS, and SigB regulate *ccrAB* expression and facilitate SCC*mec* mobility, however these findings were not consistent across all strains and signals were low[35,37,62]. None of these studies went beyond a slight correlation to describe mechanism and no studies have examined the other conserved genes. The issue that has plagued former studies is the use of highly variable clinical isolates. These strains contain a variety of other mobile elements and insertions within the SCC elements that are likely confounding their results and leading to the observed inconsistency. This has led me to move away from variable clinical isolates that have been used in other studies and use a more controlled lab adapted strain that lacks most mobile elements. Previous studies have used

reporter plasmids like I will use, but none have done an unbiased genetic screen to examine regulation of *ccrAB*.

To determine how the conserved SCC genes were regulated I used mutagenesis to isolate mutants that allowed for the expression of a reporter fused to either the *lp1413/cch* or *ccrAB* promoters as well as screening conditions and chemicals that would trigger expression. As the *S. aureus* strains were mutated after transformation of the reporter plasmids, I was able to isolate mutations in the lp1413/cch promoter even though there were no chromosomal mutants. In addition to the mutagenesis experiments, I fused the SCC gene promoters to a beta-galactosidase reporter. This allowed me to screen for different growth conditions that would lead to expression of the SCC genes.

2.2 Methods:

Strains and culturing:

In this study I used RN4220 grown in tryptic soy broth/agar. Cells were grown at 37. The reporter plasmids were made starting with the shuttle vector pCN34[63]. The transcriptional terminator region "TT" was cloned from pCN50. The beta-galactosidase reporter gene, *bgaB*, was cloned from the pMAD plasmid and the chloramphenicol resistance reporter (cat194) was cloned from pCN50[63,64]. The promoter regions for *ccrAB* and *lp1413/cch* from USA300 FPR3757 were cloned upstream of the reporter genes. As a control I also used the cadmium inducible promoter, Pcad, from pCN51. The final reporter plasmids are shown below in Table 2.1. Plasmids were maintained with 50ug/ml kanamycin. Induction of the cadmium inducible promoter was done with 1uM $CdCl_2$.

Plasmid Name	promoter	reporter	resistance
pAT1	*lp1413/cch*	*bgaB*	kanamycin
pAT2	cadmium inducible	*bgaB*	kanamycin
pAT3	*lp1413/cch*	*cat194*	kanamycin
pAT4	*ccrAB*	*bgaB*	kanamycin
pAT5	*ccrAB*	*cat194*	kanamycin
pAT6	cadmium inducible	*cat194*	kanamycin

Table 2.1 reporter plasmids used in this chapter.

Chemical Mutagenesis via hydrogen peroxide:

The plasmids containing the chloramphenicol resistance reporter fused to the SCC promoters (pAT3 and 5) were transformed into the cells prior to mutagenesis. An overnight liquid culture of TSB was diluted 1:100 to a final volume of 1.5 ml. cells were then grown to an OD600 of 0.5. 80ul of 3% hydrogen peroxide was then added to 1ml of cells. (Note: this must be done in a culture tube that is not fully sealed). Cells are then incubated at 37 shaking for 24 hours, washed with 1ml of TSB and plated on 150 mm plates of TSA + 5ug/ml chloramphenicol. Resulting colonies were then restruck on plates with 10ug/ml chloramphenicol. If actually resistant then I tried to determine if there was a mutation in the chromosome or in the plasmid that led to the observed chloramphenicol resistance. Plasmids isolated from mutant strains were subjected to Sanger sequencing. The overall process is outlined in Figure 2.1.

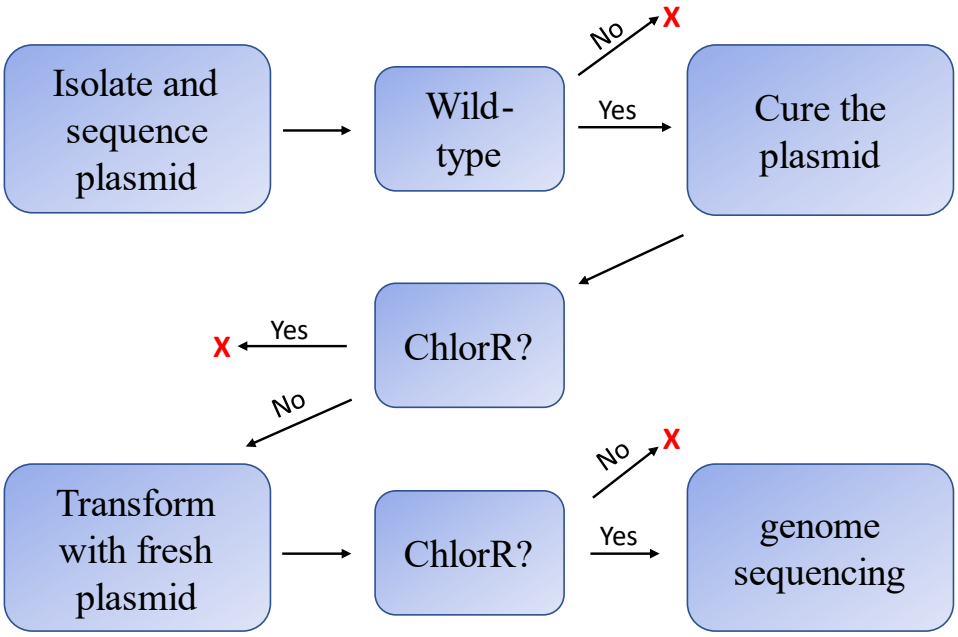


Figure 2.1: Schematic for testing if the chloramphenicol resistant colonies contained mutations in the chromosome or on the plasmid.

Screening for expression from the Beta-galactosidase reporter.

I screened various compounds and culture conditions using the beta-galactosidase reporter plasmids. An overnight culture in TSB was diluted 1:100 into a 2ml culture. Cells were grown for 3 hours shaking at 37ºC. 25ul of culture was then added to a 6 well tissue culture plates with 1ml of TSA+ 50ug/ml of X-gal. The cells were evenly spread, and the plates incubated for 3 hours at 37. Whatman paper was cut into roughly 5mm squares and autoclaved. These pieces were then dipped into various solutions and spotted onto the plates of bacteria. For temperature testing, cells were simply streaked onto TSA plates containing X-gal and incubated at the appropriate temperature.

2.3 Results:

Isolation of mutants:

Using the chloramphenicol resistance selection, I was able to isolate an initial total of 205 chlor^R strains. However, as described below, the only genetic changes that we could confidently identify and describe as enhancing transcription were in the *lp1413/cch* promoter itself.

The characteristics of the initial mutant strains are shown below in Table 2.2, and Figure 2.2 shows an example of 1 mutant strain that has a morphological change - a yellow hue compared to wildtype. (Note some mutants were discarded due to lack of growth upon restreaking and inability to isolate the plasmid for sequencing). I was able to isolate mutations within the plasmid for the *lp1413/cch* reporter plasmid, but not with the ccrAB reporter plasmid. These mutations are shown in more detail in figure 2.2. These mutations were isolated multiple times, 2-5 from different experiments and appear to make a stronger promoter. The mutations indicate that the promoter for *lp1413/cch* is weak and may require an activator. I was unable to isolate any *ccrAB* promoter mutations. The 2 mutations that do not occur within the promoter region are large truncations of the plasmid. These truncations allow for expression of the chloramphenicol resistance gene from other promoters in the plasmid.

For the mutants that did not have plasmid mutations, the plasmid proved difficult to remove. In the process of passaging the cultures to cure the plasmid, the cultures began to grow faster and mutants with colony morphology changes lost these changes. I was able to cure the plasmid from 4 mutants, 2 from each construct. When I transformed them with fresh plasmid they were not chloramphenicol resistant. This indicated that there was a problem with revertants.

promoter	Wildtype plasmid	Mutant plasmid	Yellow colonies	Slow growth	Chor^{R} without plasmid
ccrAB	80	0	0	8	1
lp1413/cch	74	13	1	5	2

Table 2.2: Description of analyzed mutants.
These are only the mutants for which I was able to sequence the plasmid. The rest were removed from the table.

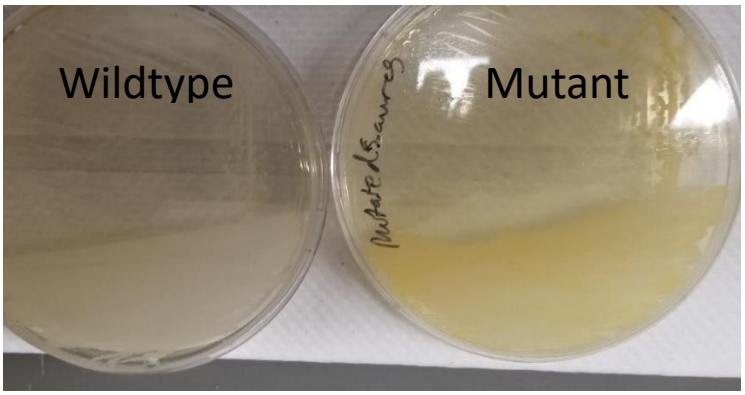


Figure 2.2: Mutant *S. aureus*
A yellowish mutant of RN4220 (right) isolated in the chloramphenicol resistance screen, vs wildtype (left).

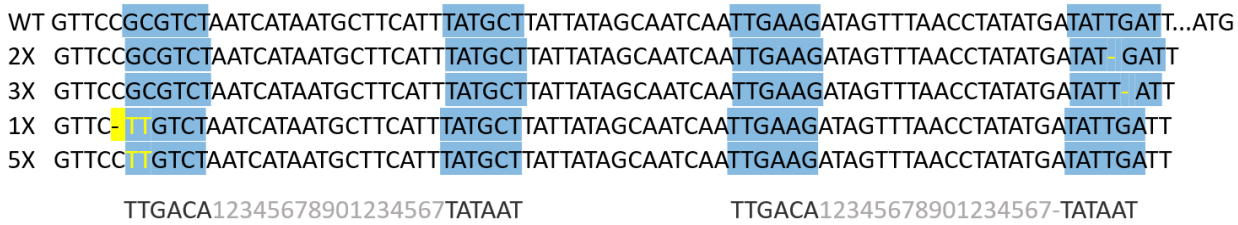


Figure 2.3: Alignment of promoter mutations
Alignment of the 5 promoter mutants and the number of times they were isolated. Below the sequence is an example of a constitutive *S. aureus* promoter with optimal spacing [65].

Testing for compounds or temperatures that induce expression:

I began by testing for expression under standard growth conditions with the Beta-galactosidase reporters. Figure 2.4A shows that the reporters work. As with the literature, there was weak expression from the *ccrAB* reporter, leading to colonies that appear light grey. I observed no expression from the *lp1413/cch* reporter. I tested conditions that the bacteria would

be exposed to in the environment or during infection, such as changing temperature, various antibiotics, different pH, and high salt. Additionally, I just used solutions that were laying around the lab. As we thought that SCC elements may be phage inducible I also tested phage. To determine if expression was temperature dependent I incubated plates at 30º,37º, and 42º. There was no expression at any of the temperatures. I then screened various compounds for expression inductions using $CdCl_2$ induction of the Pcad reporter plasmid as a control. Figure 2.4B shows the results from some of the solutions used. In addition to the solutions shown in figure 2.4B I tested: 1M NaCl, 1M ammonium acetate, 10% glycerol, phi11 phage lysate, 1ug/ml erythromycin, 1ug/ml of chloramphenicol, and .5M sucrose (data not shown). None of these led to any expression, even after extended incubation for 4 days.

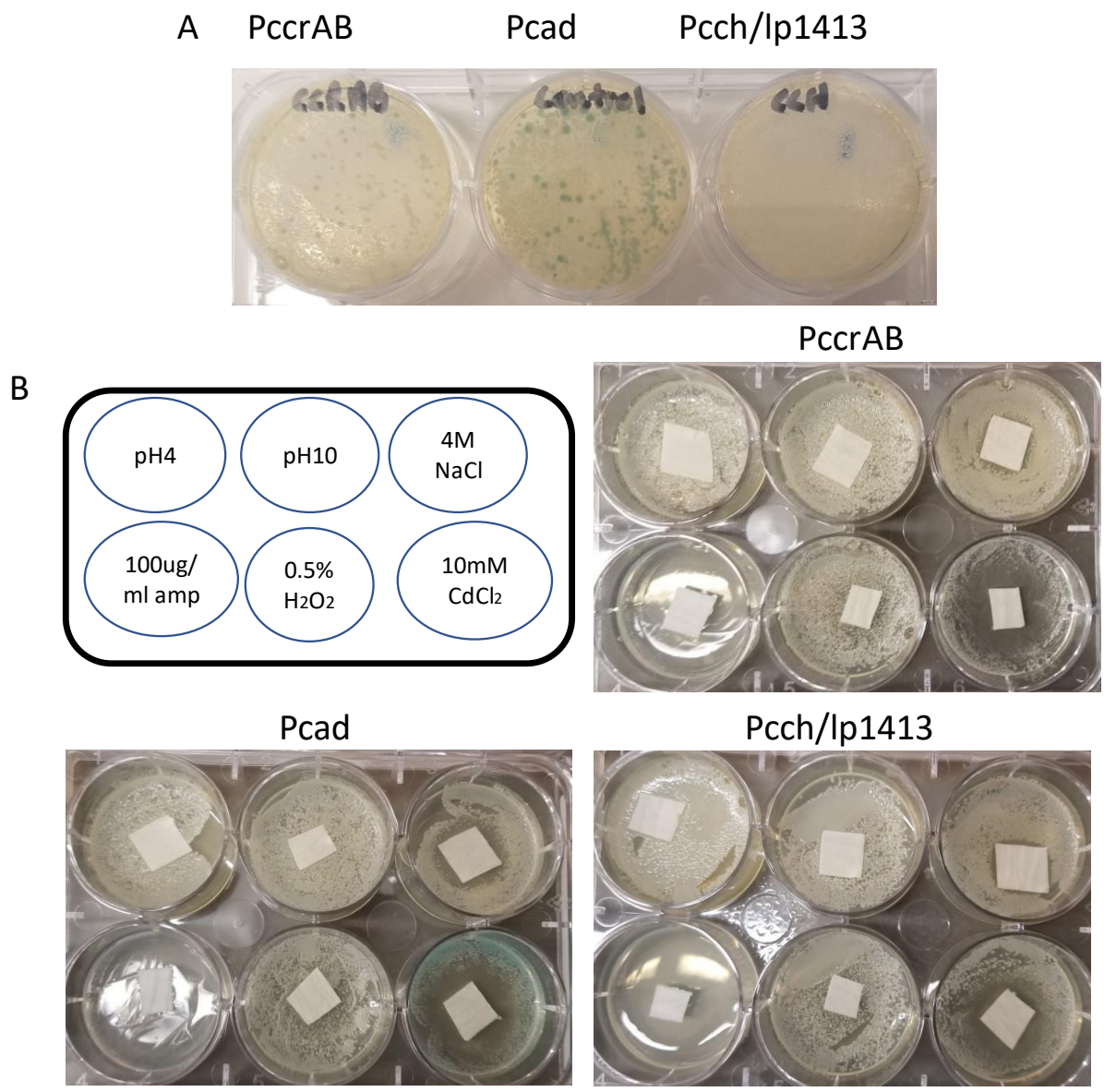


Figure 2.4 Screening for expression of SCC conserved genes
A: Expression test of beta-galactosidase reporters.
B: Screening solutions for induction of expression of beta-galactosidase. Note: pH 4 and pH 10 were the pH meter solution standards.

2.4 Discussion:

The goal of the work described in this chapter was to gain an understanding of how the conserved SCC genes were regulated as it had long been shown that they were not expressed[35,37,65]. In that respect I was able to identify mutants of the lp1413/cch promoter that

would allow for expression, leading us to conclude that natural expression likely requires an unknown activating factor. While it was unclear if that activator is in the SCC element itself from the data shown above, RNA-seq data described in chapter 4 indicates that the SCC conserved genes do not activate each other.

Furthermore, I completed these experiments unable to isolate mutants that express the SCC genes due to chromosomal mutations. The experiments were stopped due to technical issues. The plasmid reporter design led to many revertants, and the hydrogen peroxide mutagenesis was not robust enough. Future mutagenesis experiments would either need to use a chromosomal reporter or at least a reporter plasmid that has a temperature sensitive replicon so that it could be removed from the mutant strains more easily.

Another method that I used to try mutagenesis was using MuA transpososomes. This had been previously shown to create a mutant library in *S.aureus* and we had an abundance of protein available, including hyperactive mutants of MuA[66,67]. The transpososomes contained a fragment of DNA with the transposon ends with the reporter gene and SCC promoter or Pcad control in between. The DNA fragments were combined with purified MuA protein and the transpososomes were electroporated into the cells. Unfortunately, I was unable to transform large amounts of transpososomes into the cell.

As stated previously, SCC elements are not just within *S. aureus* strains. The regulation of expression could be different in other staphylococci sp. Future directions would be to test for expression of the conserved genes in CoNS.

Lastly, with it being shown that SCC elements can move via natural competence[48], the question of how the genes are regulated and expressed can be looked at through a different lens. The work described in this chapter was looking at how expression can be triggered if the element

was already in the cell and in the chromosome. As the DNA containing SCC elements enters the cell via natural competence it would be as a single strand. Future work will be aimed at identifying if the genes are expressed from a single stranded promoter upon translocation of the element into the cell. While this is currently our leading theory on how SCC genes are regulated as they can move via natural competence, this movement still occurs at a low frequency and SCC elements are highly abundant. We cannot rule out that there is another method of transfer and gene regulation. One hypothesis is that the elements are moved via a phage that has yet to be identified and/or tested. Said phage could encode an activator that would trigger SCC gene expression leading to movement, similar to SaPIs and other phage inducible mobile elements[38,68].

Chapter 3: Replication of SCC elements

3.1 Introduction

Replication is an important mechanism for the maintenance and spread of mobile genetic elements. Replication allows for a mobile element to remain in the donor and recipient strains and/or transfer to multiple recipient cells at a time. While many mobile elements replicate, they do so in a variety of ways. Plasmids commonly replicate using theta or rolling circle mechanisms[28,31,69,70]. Some mobile elements have been shown to switch between the two[71]. It is known that SCC elements can excise as a circle from the chromosome[18,56], however the question of what happens after that excision has yet to be explored.

To explore if SCC elements will replicate, we looked at how similar elements replicate, with the closest relative being SaPIs. SaPIs are a group of pathogenicity islands found in *Staphylococci*. These elements are integrated into the chromosome and through bacterial infection with a helper phage, expression of the SaPI genes is activated[40,55,72]. The element is excised from the chromosome and replicated to high numbers, packaged into phage, and dispersed between neighboring cells via phage transduction[54,73].

To facilitate replication, SaPIs encode a replicative helicase and a primase, termed rep and pri. This helicase binds 10-12 iterons (varies based on which SaPI) directly downstream of the helicase this is the SaPI origin of replication or ori[53,55]. The helicase opens a replication bubble in the excised SaPI at the ori and the primase, Pri, makes primers initiating replication[53]. The cellular polymerase then comes in and replicates the SaPI[53]. When we looked at pattern 2 SCC elements we see a similar structure to the SaPI replicon (Figure 3.1). Pattern 2 SCC elements have a helicase, primase, and after the helicase gene there are 3 direct repeats that the

helicase binds to.[52] In addition to having a similar genomic structure, Cch from pattern 1 and Cch2 are structurally homologous to the helicases from SaPIbov1 and SaPI5 respectively[49,52]. For these reasons we hypothesized that, similar to what has been shown for SaPIs, the rep protein, primase, and direct repeats from pattern 2 SCC elements could support replication of a plasmid[54]. To test this I cloned the putative SCC replicon and the SaPIbov1 replicon into a temperature sensitive plasmid. If the SCC proteins can facilitate replication, then there would be detectable plasmid at the restrictive temperature.

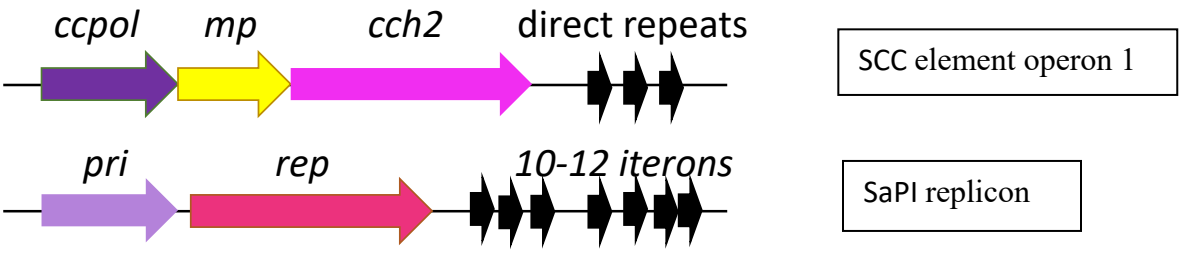


Figure 3.1: Comparing putative SCC replicon with SaPI replicon

3.2 Methods:

Culturing and strains

All experiments were done in *S. aureus* RN4220. Bacteria were grown on TSA supplemented with 5ug/ml Erythromycin and 1uM $CdCl_2$ (to induce protein expression). The SCC genes were cloned from the SCC*mec* found in *S. aureus* strain TSGH-17[74].

Plasmids for the initial experiments were based on pCN51. The replicon in pCN51 was swapped with the pE194-ts replicon from plasmid pBursa [75] . The SCC genes, SaPI control, and BgaB were cloned downstream of the cadmium inducible promoter. Functional mutations were made in Ccpol, Cch2 or both as controls. In addition to the mutants, a control without the putative origin, the intergenic region containing the direct repeats, was cloned. The cloning of these plasmids labeled (pAB…) was completed by Aleksandra Bebel. The BgaB expression

plasmid was used as a control for the effects of simply overexpressing a protein. The SaPI used as a control was the bovine derived SaPIbov1.

Growth assays for testing for replication were done by taking colonies from a fresh transformation and growing for 2 days at 30ºC. The cells were then normalized in 1ml to an OD600 of 0.1 and serial dilutions were made. 10ul from each dilution were then spotted on plates described above. The plates were then incubated at 28º and 37º for 2 days and 1 day respectively. Colonies were counted and cfu was calculated at each temperature.

Plasmids used for the expression of the 3rd operon proteins were cloned by taking the plasmid pIMAY. This plasmid is temperature sensitive and has a tet inducible promoter. I cloned the 3rd operon downstream of the tet inducible promoter. The replicons were cloned from the previous replication plasmids.

Plasmids	**Genes expressed**	**inducer**	**resistance**	**Plasmid backbone**
pAB086	Full SCC putative replicon	cadmium	erythromycin	pCN51
pAB091	Mutated Cch2	cadmium	erythromycin	pCN51
AB093	Mutated ccPol	cadmium	erythromycin	pCN51
pAB095	Mutated Cch2, Mutated ccPol	cadmium	erythromycin	pCN51
pAB097	SaPI pri-rep	cadmium	erythromycin	pCN51
pAT39	Full SCC putative replicon	cadmium, tetracycline	chloramphenicol	pIMAY.
pAT41	Mutated Cch2	cadmium, tetracycline	chloramphenicol	pIMAY.
pAT42	Mutated ccPol	cadmium, tetracycline	chloramphenicol	pIMAY.
pAt43	Mutated Cch2, Mutated ccPol	cadmium, tetracycline	chloramphenicol	pIMAY.
pAT44	SaPI pri-rep	cadmium, tetracycline	chloramphenicol	pIMAY.
pBgaB	BgaB	cadmium	erythromycin	pCN51

Table 3.1: Plasmids used in this chapter

Southern blotting for replicated plasmid

Digoxigenin (DIG) labeled probes were created using the PCR DIG Labeling Mix ^{PLUS} (Roche). To label plasmid DNA, I used a probe that matched Pcad the cadmium inducible promoter region and to label genomic DNA, I used a probe that bound to *rlmH*. The DIG PCR labeling mix was added (in place of DNTPs) to a standard PCR using Vent exo- DNA polymerase (NEB).

A colony from a fresh transformation was taken and grown overnight at 28°C for 2 days. Cultures were normalized to an OD of 0.01 and grown for 4,5,6, or 24 hours. Time points varied by experiments. At the timepoints 1 ml of cells was removed and pelleted. The pellet was resuspended in TSM+ lysostaphin (2ug/ml) and incubated for 1hour at 37ºC. Plasmid and gDNA were then isolated using the PureLink ^{TM} Genomic DNA Mini Kit. 1ug of DNA was digested with BamHI. This was then run on a 1% agarose gel with no DNA dye. The gel was then washed for 30 minutes in denaturing solution (0.5MNaOH 1.5MNaCl) with rocking. The denaturing solution was removed, and the gel was washed 2 times for 15 minutes in neutralizing solution (1MTris-HCl 1.5M NaCl (pH 7.5)). The gel was then given a quick rinse with water and added top down to a glass plate with Whatman paper that had been pre-soaked in 20X SSC buffer(0.3M sodium citrate in 3M NaCl). The Whatman paper is used as a wick into the reservoir of 20X SSC buffer. The gel is outlined in parafilm. A positively charged nylon membrane was presoaked in 10X SSC buffer then added on top of the gel making sure there are no bubbles while also avoiding touching the membrane. The membrane is then topped with 3 sheets of Whatman paper soaked in 10x SSC buffer. A large stack of paper towels and a second glass plate are added. Lastly a weight is added to the top. This is incubated overnight and the next day the membrane is gently removed. The membrane was crosslinked using a UV cross linker, with the auto crosslink

setting. The membrane was then incubated in prewarmed hybridization buffer rocking for 30 minutes at 41ºC. The probe was added to the hybridization buffer at a concentration of 25 ng/ml. The pre-hybridization buffer was removed and the hybridization buffer with probe was added. This was then incubated at 41ºC 6 hours-overnight. This time is dependent on efficiency of labeling. For a fresh high yield probe, I was able to incubate at the lower end of 6 hours, but for an old probe I would incubate overnight.

The hybridization buffer was the DIG easy hyb buffer and was part of the DIG High Prime DNA Labeling and Detection starter Kit II (Sigma-Aldrich). I followed the kit instructions minus the labeling steps. The labeling of the probe was done as described above. For experiments where I had to strip the membrane and re-probe with the rlmH probe, I followed the kit instructions.

3.3 Results:

Testing replication of pattern 2 SCC elements:

To determine if SCC elements can replicate, I took the putative replication locus, ccpol-mp-cch2-direct repeats (figure 3.1) and cloned that into a shuttle vector with a temperature sensitive *S. aureus* replicon. If the SCC genes could facilitate replication of this synthetic plasmid, then the cells would grow at the restrictive temperature (42ºC). As a positive control, I used SaPI bov1 pri-rep. To further explore what was necessary for replication, I utilized ccpol and cch2 mutants. These mutants have no helicase or primase activity[52]. Figure 3.2 shows that we did not observe maintenance of the plasmid when cells were grown at the restrictive temperature and had a complete SCC putative replicon.

An issue that arose with this assay is that there did appear to be some toxicity of SaPI pri-rep and some of the SCC proteins when grown at 28º. This made counting difficult and altered the number of colonies. I tried using less inducer (0.1uM cadmium, data not shown) but even in the no cadmium controls the leaky expression was enough to cause minor growth issues. Figure 3.2 C shows that there were indeed growth defects with SaPI *pri-rep* overexpression and mutated SCC genes. While there was growth issues at the permissive temperature, the lack of any growth at the restrictive temperature with the full wildtype SCC replicon (WT SCC) seems to indicate that there is likely no replication occurring. If there was replication I would expect to see a difference in growth at the restrictive temperature with the WT SCC vs the cch2 or ccpol mutants.

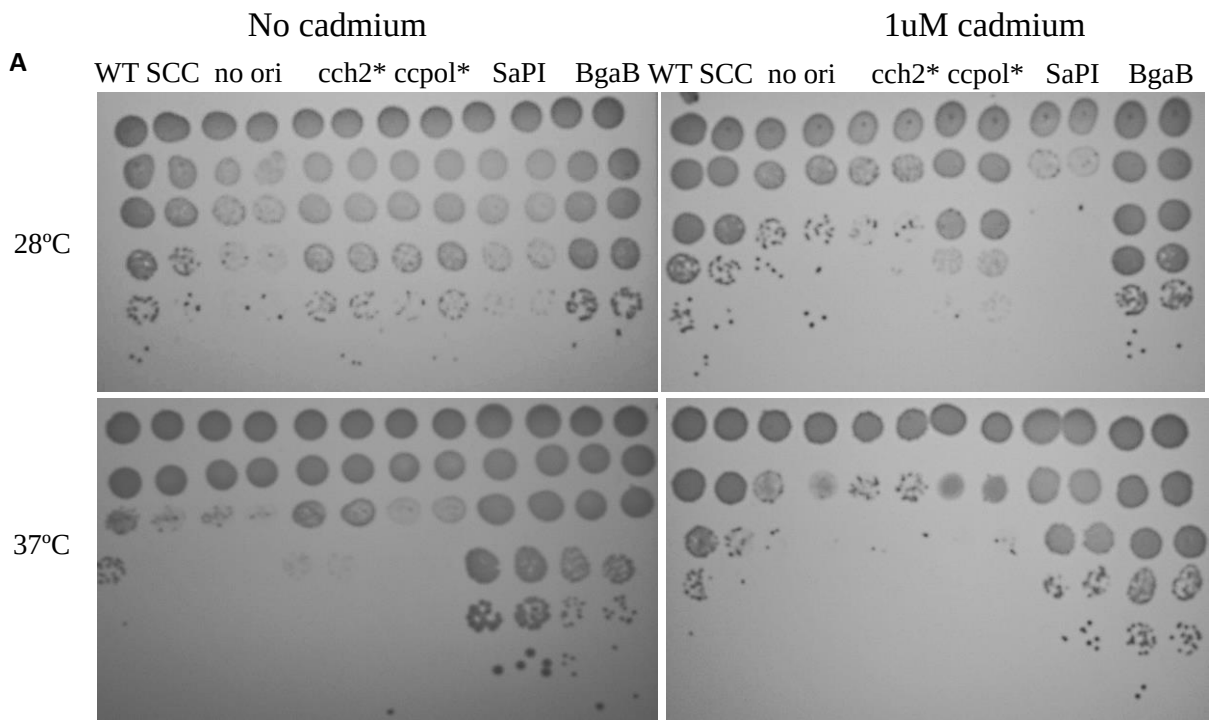


Figure 3.2: Testing for replication via growth at the restrictive temperature

A,B : Growth of *S. aureus* containing replication plasmids at 28°C and 37°C. WT SCC denotes the full wildtype SCC putative replication locus. cch2* and ccPol* denotes functional mutations. SaPI refers to the SaPIbovI pri-rep + iteron complex. Panels on the right are with 1uM cadmium in the plates. Panels on top were of plates incubated at 28° and panels on the bottom are of plates incubated at 37°C. In A the control was the BgaB overexpression plasmid, however in B the double mutant was used as a negative control. Two columns of each strain are 2 replicates from 2 different colonies post transformation with the expression plasmids.

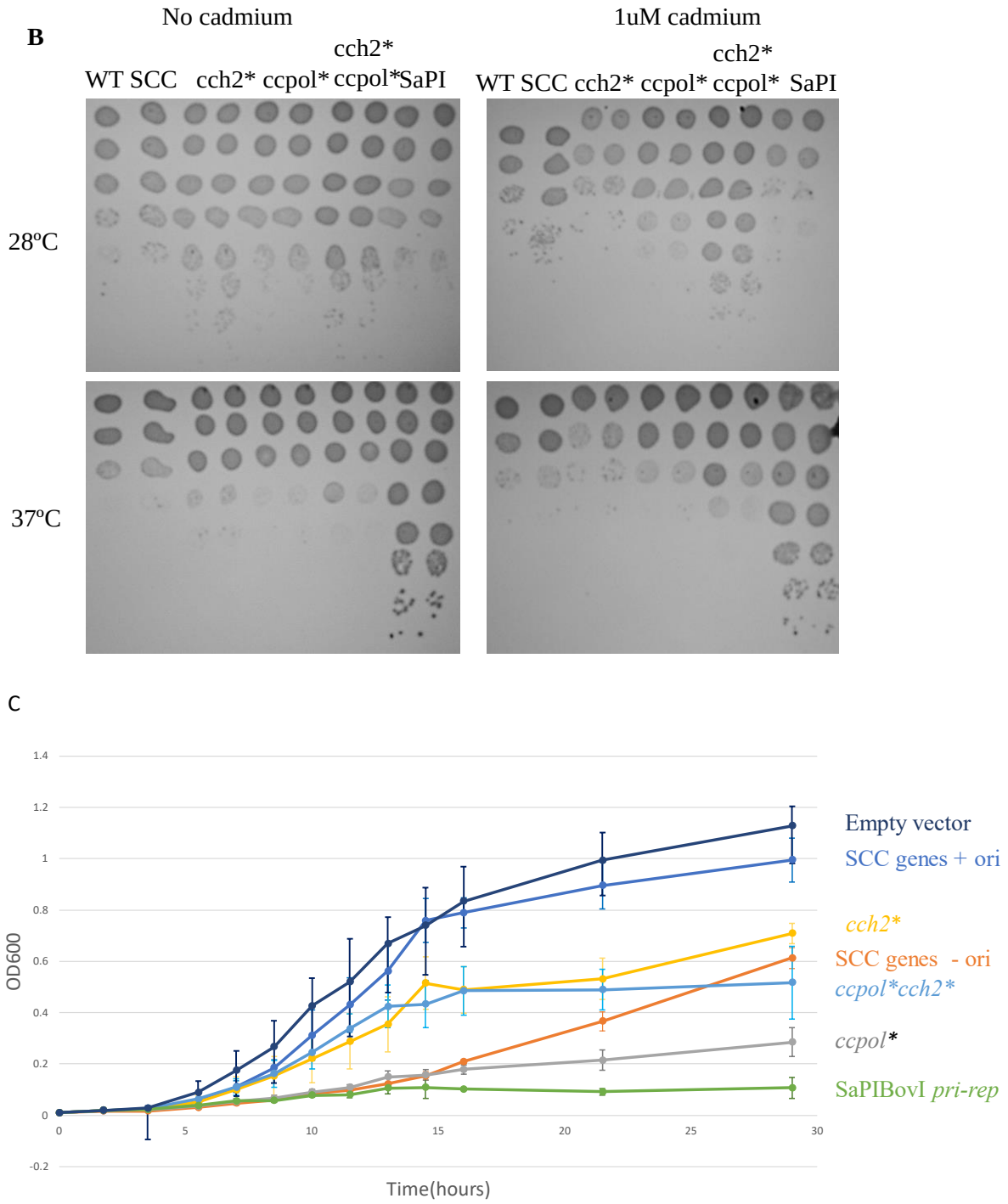


Figure 3.2 continued

C: growth curve of strains overexpressing the replication proteins. Cells were grown at 28°C. n=3

Due to the growth issues I decided to test for replication a different way, using a southern blot to test for the plasmid, as opposed to the growth assay. This utilized the temperature sensitive plasmids but worked at earlier timepoints. Also, this tested for the presence of the plasmid rather than a full colony which would allow me to be able to detect weaker replication. Figure 3.3 shows that I was unable to detect any plasmid with the SCC putative replicon at 6 or 24 hours. I was able to detect SaPI replication at 6 and 24 hours. The amount of SaPI observed does decrease overtime, which is expected as SaPI replication leads to long concatemers that are not cleaved into individual plasmids in this system[54]. By 6 hours the plasmid was gone in the empty vector controls meaning that the temperature sensitive replicon was easily lost with no background. This was repeated twice with the same results. There is no time 0 for the experiment below. The plasmid was selected for prior to initiating the experiment.

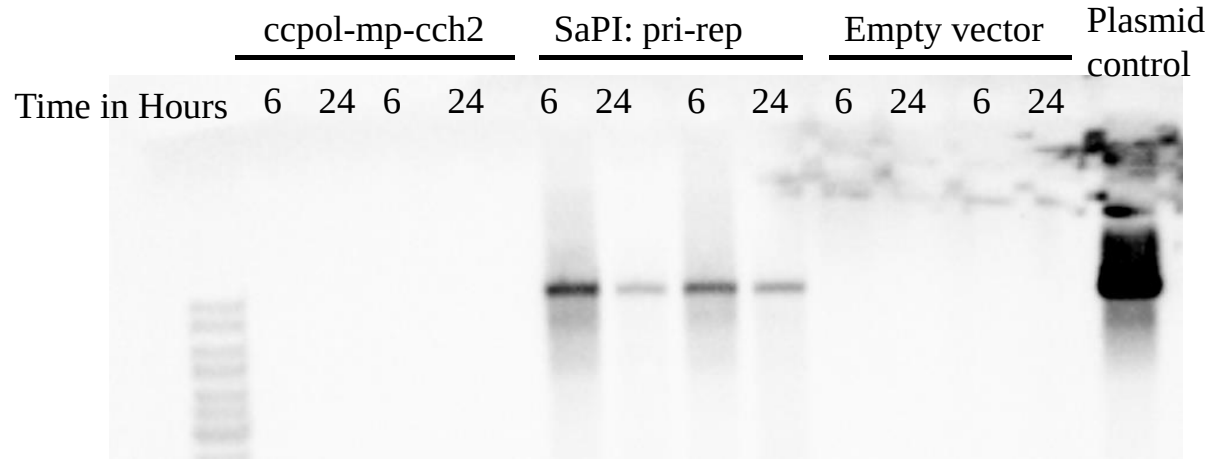


Figure 3. 3: Southern blot for replicated plasmid
Two replicates from 2 different colonies post transformation with the expression plasmids.

To follow-up we wanted to test if the proteins in the last operon of the conserved SCC locus assisted in replication. This operon is present in both patterns of SCC elements and other mobile elements. This operon will be described more in chapter 4, however the first protein

encoded in this operon is a uracil DNA glycosylase inhibitor, SaUGI[51]. A phage encoded uracil DNA glycosylase inhibitor has previously been shown to improve replication[76]. It is hypothesized that this is due to the blocking of base excision repair that would destroy the ssDNA intermediate. For this reason, I cloned the 3rd SCC conserved operon into a new set of replication plasmids that also contain the SCC putative replicon or the control SaPI replicon.

Figure 3.4A shows that with expression of the 3rd operon, there was still no replication from the SCC genes. Interestingly there was a decrease in the amount of SaPI replicated plasmid that was detected. This assay was repeated with the SaPI replication plasmid at an earlier timepoint and with a DNA amount control (Figure 3.4B).

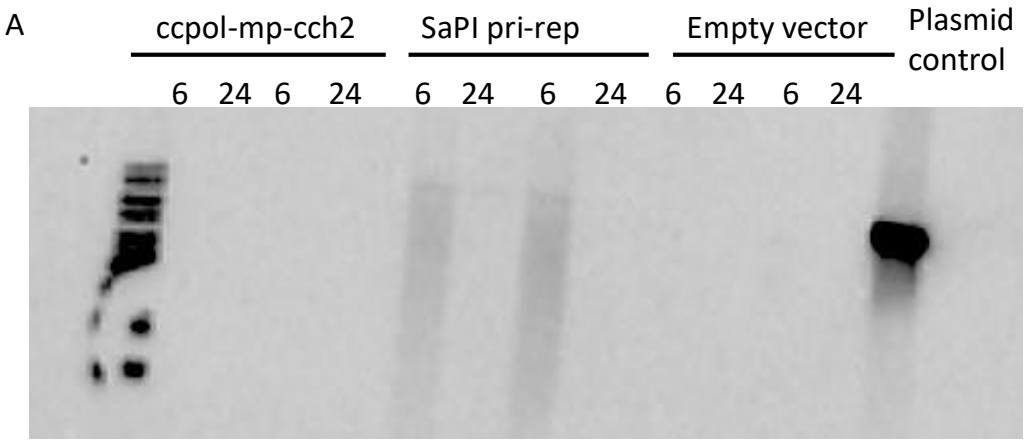


Figure 3.4: Replication of SCC and SaPIs with the 3rd conserved SCC operon induced

A: Southern blot for plasmid with Pcad probe at 6 and 24 hours of growth at the restrictive temperature.

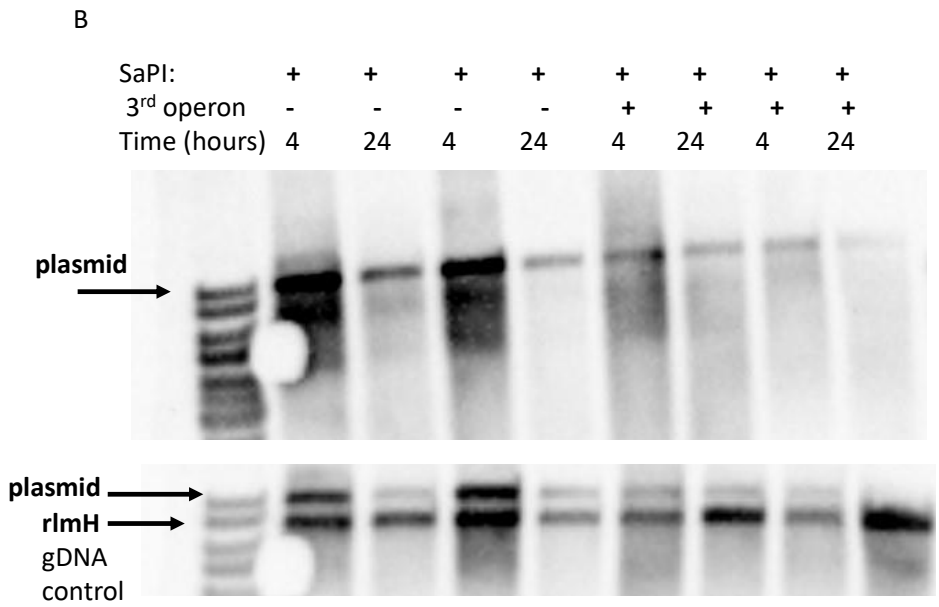


Figure 3.4 continued
B: Southern blot for plasmid with Pcad probe. The blot was then stripped and reprobed for *rlmH*

3.4 Discussion:

While we hypothesized that SCC elements would replicate, I was unable to detect replicated plasmid. This does not mean that the elements do not replicate as there could be weak replication or replication under unknown conditions. With the new knowledge of SCC transfer via natural competence, this leads us to hypothesize that the putative replication proteins are not replicating an extrachromosomal SCC element but may be adding the second strand to an imported single stranded SCC. This will be discussed further in chapter 5.

While I was unable to find evidence of SCC replication, the assay has proved useful in understanding SaPI replication. The decrease in SaPI replication with the 3rd operon expressed as well as the testing of mutant rep proteins are currently being studied. One thing that was observed was the decrease of SaPI replicated plasmid overtime, due to large concatemers. For this reason, Sin resolvase and its cognate resH DNA site were cloned into the replication

plasmid[77,78]. This should resolve the long concatemers into individual plasmids, allowing for efficient separation into the two daughter cells upon division. Figure 3.5A shows the amount of plasmid that was detected at 24hrs was more when Sin and the binding site resH was added as opposed to no binding site indicating that the addition of Sin did lead to plasmid stability at 24 hours. In figure 3.5B I was able to detect supercoiled plasmid indicating that Sin was leading to individual plasmids and not the long concatemers that SaPI replication is known to produce. The improvements to the replication plasmids through the addition of Sin will improve the ability to test for changes in SaPI replication.

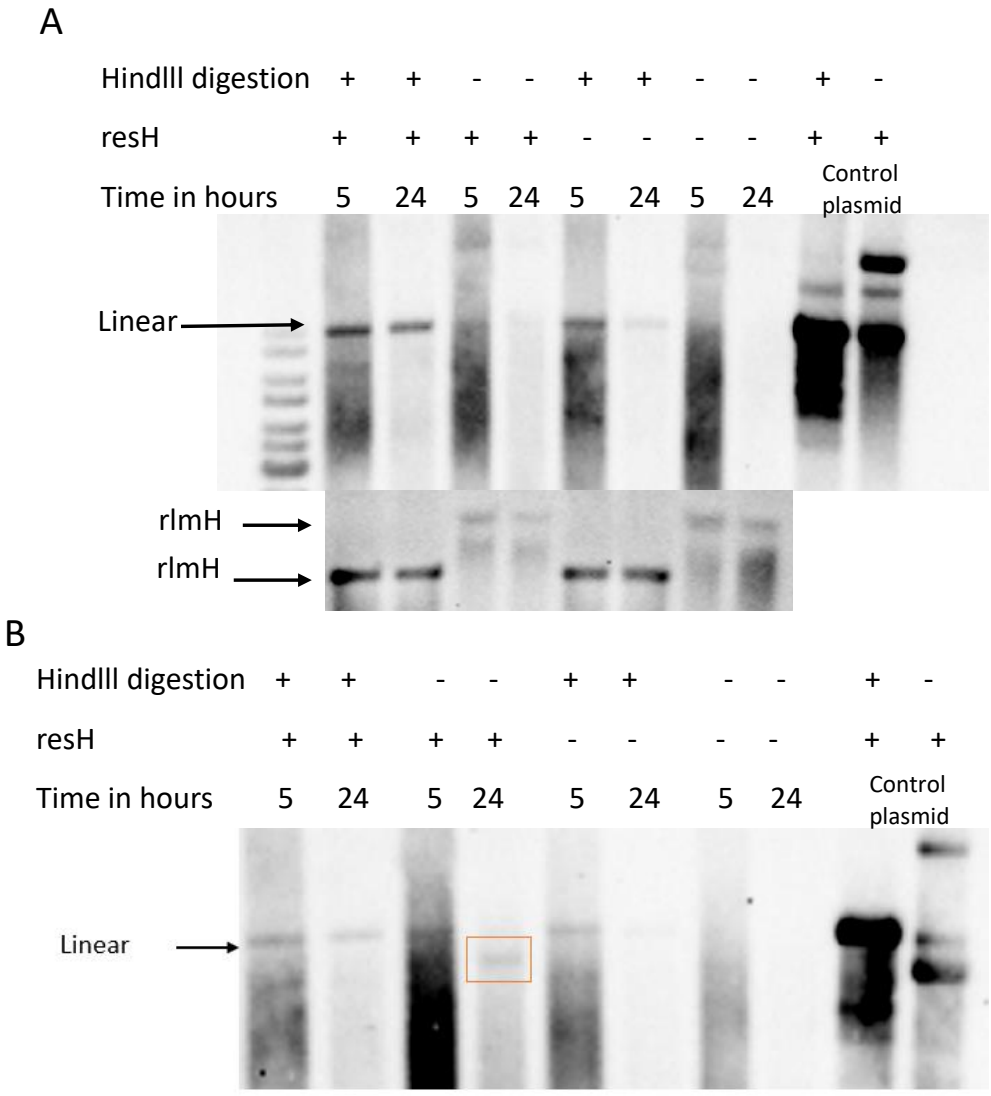


Figure 3.5: Replication of SaPI plasmids with Sin resolvase

A: Southern blot for the plasmid with Pcad probe at 5 and 24 hours. Plasmids were + and – ResH the binding site for Sin. HindIII is a single cutter in the plasmid. Blot was stripped and reprobed with the gDNA marker rlmH.

B: Repeat of the experiment in A. Orange box denotes supercoiled plasmid

Chapter 4: Functional analysis of a highly conserved operon carried by Staphylococcal SCC elements.

The following chapter is a paper to be submitted for publication and to biorxiv. As part of this paper, I designed and completed the phenotypic assays. All the bacterial work was done by myself. Proteins were purified by Ying Pigli. I carried out the initial assays demonstrating UDGase activity for CcG. Later UDGase assays were completed by Ying Pigli. RNAseq analysis was completed with the assistance of Aretha Fiebig.

Functional analysis of a highly conserved operon carried by Staphylococcal SCC elements.

Alexis Thomas, Aretha Fiebig, Ying Z. Pigli and Phoebe A. Rice

4.1 Introduction:

Staphylococcus aureus is a bacterial pathogen that can cause mild to severe infections. The severity and treatability of infections vary in part due to the acquisition of mobile genetic elements[79]. These mobile elements carry a variety of virulence factors including antibiotic resistance genes and toxins. One group of these mobile elements, Staphylococcal Cassette Chromosomes, or SCC elements, are important because acquisition of SCC*mec* leads to the transition from methicillin sensitive (MSSA) to methicillin resistant *S. aureus* or MRSA[24]. In addition to the *mecA* locus, SCC elements can carry a large variety of cargo including other mobile elements such as transposons, IS elements, and mobilizable plasmids[24]. This work aims to further understand how SCC elements function by examining a conserved operon.

SCC elements are genomic islands that are present in many *S. aureus* strains as well as other non-pathogenic and pathogenic Staphylococcal *species*[24,25,80]. SCC elements are large and highly variable but contain a conserved gene locus and always integrate into the same chromosomal location, *rlmH*[24]. While SCC elements have been studied extensively, little is

known about the mechanisms utilized to maintain the elements in the chromosome and to mobilize them. It has been shown that SCC elements utilize site-specific serine recombinases to excise the element from the chromosome as a circle and to integrate the element into the *rlmH* gene[24,56,57]. A recent study also found that SCC elements can transfer to new hosts in a recombinase-dependent manner within a biofilm due to natural competence[81]. To further understand SCC elements, we focused on the conserved gene locus, with the idea that anything that is required for the element to function would be conserved.

The focus of this study is a conserved 4-ORF operon found in all SCC elements. SCC*mec* variants are typically typed by the *mecA* locus and the encoded recombinase[23,82]. While there are many types of SCC*mec* elements, by examining the genes that flank the encoded recombinases, a previous study found that there are 7 conserved SCC genes, and that they follow 2 patterns (Figure 1A)[49]. The operon studied here is found in both patterns immediate downstream of the recombinase gene(s) in both patterns. Little work had been done previously to understand the function of this operon. While SCC elements vary even in the conserved locus this operon is present in both patterns of conserved genes. In addition to being present in all SCC elements, this operon is present in other mobile elements, most notably ACME (Arginine catabolic mobile element)[83] (Figure 1B), which is found downstream of SCC*mec* in some community acquired or CA-MRSA strains[84]. As the operon is highly conserved, we wanted to explore what benefit it conferred to the host bacteria and/or the element itself.

The 4 ORFS encoded in this operon are: *S. aureus* uracil DNA glycosylase inhibitor (SaUGI), a previously undescribed peptide, Lp960 (which contains DUF960), and Ccg (described in this study). SaUGI is a small protein that mimics the shape and charge of DNA and binds tightly to the DNA binding site of *S. aureus* uracil DNA glycosylase (UDG), thus

inhibiting excision of uracils from DNA[85]. However, whether or not it also occludes other DNA-binding proteins had not been previously addressed. Proteins like SaUGI have been shown to be encoded by certain phages and are important for their replication[76], something that has not been observed to occur with SCC elements. The next protein, Lp960, is present in all SCC elements but has no predicted function. The 3rd protein, here named Ccg, contains DUF1643. This operon sometimes contains a 5th ORF encoding a protein annotated as RadC, but that ORF is often truncated (most commonly by insertion of Tn554).

To further understand the proteins encoded in this operon we examined their function *in vitro* as well as observed changes that occur when the proteins are overexpressed. We know that the conserved genes are not expressed under standard lab conditions, apart from sporadic recombinase expression in a low percentage of cells[11–13]. To better understand the function of the conserved proteins we examined their effect on the bacteria, thus providing clarity into how they could be assisting the element in excision, maintenance, or transfer. The most intriguing *in vivo* effects found were due to SaUGI - overexpression of it led to many changes in gene expression and to a loss of biofilm. Our RNA-seq experiments rule out the hypothesis that some of the conserved SCC genes act as transcriptional activators of the others. We also show that Ccg is predicted adopt a glycosylase fold and has uracil DNA glycosylase activity but does not bind to SaUGI. Finally, our *in silico* analysis of the peptide encoded between SaUGI and Lp960 predicts a conserved amphipathic helical structure, similar to that of many phenol soluble molecules (PSM).

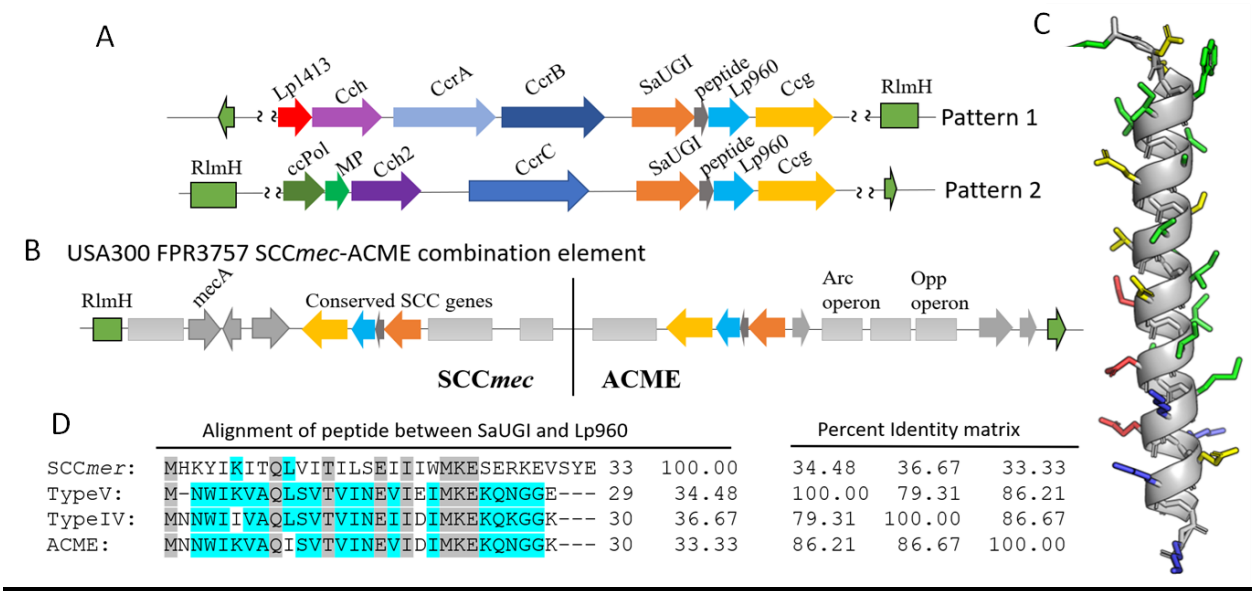


Figure 4.1: Conserved features of SCC elements

A: Graphic depiction of pattern 1 and pattern 2 SCC elements. Both are inserted into the 3' end of *rlmH*.

B: Depiction of the compound SCC*mec* and ACME element found in the community acquired MRSA strain USA300 FPR3757. Relevant genes are annotated. Grey blocks represent a multiple additional genes and operons not discussed here.

C. Alphafold2 prediction of structure of the peptide from the SCCmec type IV (a pattern 1 SCC element) found in USA300 FPR3757. N is on the top and C on the bottom; red = negatively charged, blue = positively charged, yellow = neutral polar (N/Q/S/T) and green = hydrophobic.

D: Multiple sequence alignment of peptide amino acid sequence from SCC*mer* (strain TSGH-17), a type V SCCmec/pattern 2 (strain TSGH-17), Type IV SCCmec/pattern 1 SCC (strain USA300 FPR3757), and from ACME (strain USA300 FPR3757). The percent identity matrix was determined using the clustal 12.1.

4.2 Methods:

Bacterial strains/Culturing

The *S. aureus* strains used in this work were grown at 37ºC in either tryptic soy broth/agar (TSB, MP Biomedicals/Difco) or Brain heart infusion broth/agar (BHI, Millipore Sigma/Research products international). *E. coli* strains were grown at 37ºC or 20ºC in LB, Miller broth/agar (Fischer). Antibiotics needed for plasmid maintenance and to test for transposon

transfer were used at concentrations of: 10ug/ml chloramphenicol, 100ug/ml ampicillin, and 10ug/ml erythromycin. Congo red agar (CRA) contains BHI agar with 0.08% Congo red (Fisher chemical) and 5% sucrose.

S. aureus RN4220 transposon mutants were obtained by using phage Phi85 transduction from other strains following the method found in [86]. Transposon insertions in *icaA and sarX* in the USA300 JE2 background were obtained from BEI resources. They were transduced from USA300 JE2 to RN4220. The USA300 JE2 mgrA::erm and RN4220 mgrA::erm mutants were created the same way as above using Newman mgrA::erm transposon mutant (kindly provided by Dominique Missiakas) as a donor[87].

Lysostaphin resistance was assayed similarly to the method of Wiling *et al*[88]. A 4-ml overnight culture in BHI was washed twice with 50 mM Tris-HCl, pH 7.5, and cells were suspended in 4ml of the same buffer + 5ug/ml lysostaphin. The OD 600 was then measured at 5 min intervals while incubating with shaking at 37°C.

Plasmid name	Resistance	purpose
pET21a-SaUGIstrep	amp	Expression of twin strep tagged SaUGI
pET21a-SaUGI	amp	Expression of His_6 tagged SaUGI
pET21a-Ccgstrep	amp	Expression of twin strep tagged SaUGI
pET21a-SaUGI-SarX	amp	Twin strep tagged SaUGI and His_6-tagged SarX
pAT57	cat	Empty vector
pSaUGI	cat	SAUGI induction
pLp960	cat	lp960 induction
pCcg	cat	Ccg induction
pPeptide	cat	peptide induction
pSarX	cat	SarX overexpression
pSarX+SaUGI	cat	SarX+SaUGI expression

Table 4.1: plasmids used in this chapter

Strain		Source
E. coli		
Dh5α	Used for cloning plasmids	
Rosetta (DE3) plysS	Protein expression and purification	
S. aureus		
RN4220	Restriction mutant (MSSA)	[89]
USA300 JE2	USA300 LAC strain cured of plasmids (MRSA)	[90]
USA300 JE2 *sarX*::erm	Transposon mutant of SarX	[90]
USA300 JE2 *icaA*::erm	Transposon mutant of IcaA	[90]
USA300 JE2*mgrA*::erm	Transposon mutant MgrA	This study
RN4220 JE2 *sarX*::erm	Transposon mutant of SarX	This study
RN4220 *icaA*::erm	Transposon mutant of IcaA	This study
RN4220 *mgrA*::erm	Transposon mutant MgrA	This study

Table 4.2: Strains used in this chapter

Isolation of RNA/RNA sequencing:

An overnight liquid culture was diluted 1:100 and grown in 2 ml of TSB at 37°C for 1hr then cadmium chloride was added to a final concentration of 2 uM and the culture was grown at 37°C for 2 additional hours to reach an OD 600 of between 0.5 and 0.8. The 2 ml cultures were then flash frozen in liquid nitrogen. The 2 ml frozen cultures were mechanically lysed using the SPEX Sample Prep Freezer Mill 6875. Machine settings were as follows: 3-minute pre-cool, 5 min run time, 1 min cool time, for 5 cycles as at a rate of 10 CPS. The lysates were then mixed with 4 mL of TRIzol (Thermo Fischer scientific). 800ul of chloroform was then added to the tubes and the lysates were vortexed for 1 min. The samples were then spun at 12,000g for 15 min at 4°C. The aqueous layer was then combined with an equal amount of 100% ethanol. The RNA was then extracted by adding the lysate plus ethanol to the Zymo RNA Clean and Concentrator-25 kit. The RNA was then sent to the SeqCenter (Pittsburgh, Pennsylvania). Depletion of rRNA and library prep was done by the SeqCenter using Illumina's Stranded Total RNA Prep Ligation Ribozero Plus kit. 50 bp paired end reads were generated using the Illumina NextSeq 2000

platform. The sequencing reads were mapped to the *S.aureus* USA 300_FPR3757 genome, accession number NC_007793 using default mapping parameters in CLC Genomics Workbench (Qiagen).

Biofilm assay:

10 ul of a 2ml overnight liquid culture in TSB was added to a 96-well clear bottomed plate. 190ul of BHI supplemented with 2% glucose, 2uM cadmium chloride and 10ug/ml of chloramphenicol was added. The plates were then incubated at 37° for 24 hrs (without shaking). The media was then gently removed, and cells were fixed by adding 200 ul of methanol for 30 min. The methanol was gently removed, and the cells were stained with 1% crystal violet for 15 minutes. Wells were washed with water 3 times. Plates were dried overnight. The next day 100 ul 33% acetic acid was added to the wells to remove the crystal violet. 10ul was removed and added to a fresh 96-well plate with 90ul of water. The absorbance was read at 590nm using the Synergy Neo HST Plate Reader.

Triton X-100 induced autolysis assay.

This procedure follows the procedure described in Chu *et al*, 2013 [91]. An overnight liquid culture of cells in 1 ml BHI was diluted 1:100 in 5 ml BHI+ 10 ug/ul chloramphenicol plus 10 uM $CdCl_2$. The cells were then grown for 2.5 hours at 37ºC to reach an OD600 of 0.5-0.7. The cells were then harvested and washed twice with cold water, then resuspended in 5 ml of 50 mM Tris-HCl (pH 7.5) and 0.05% Triton X-100. The OD600 was then measured at 30-minute intervals.

Purification of SaUGI and Ccg

Twin strep tag-Ccg: The Ccg coding sequence was cloned into the pET21a expression vector with the twin strep tag (SAWSHPQFEKGGGSGGGSGGSAWSHPQFEKGSGSG) at the N-terminal end. The protein was expressed in *E. coli* Rosetta (DE3) plysS cells. 1-liter cultures were grown at 37°C until OD_{600} reached ~0.7. Expression was then induced by adding 0.5mM IPTG, after which the culture was shifted to 20C and grown overnight. Cell pellets were resuspended in Strep buffer (100 mM Tris, 0.5mM EDTA, 1M NaCl, 5% glycerol, 1mM DTT, pH 8) + protease inhibitor cocktail (1 tablet/liter of cells-cOmplete Mini, EDTA free) + lysozyme (final concentration 200 ug/ml), then sonicated 3 times for 1'. The supernatant was collected after centrifuging at 20,000 rpm in a ss-34 rotor for 1 hour. This was filtered using a 0.22micon then loaded onto a strep column (StrepTrap^{TM} HP GE 5 ml). The column was then washed with strep buffer. The protein was then eluted using strep buffer + 2.5 mM desthiobiotin (Iba sciences). The pooled fractions containing the Ccg peak were then loaded onto a heparin column (HiTrap^{TM} Heparin HP 20 ml GE) for further purification. After washing with 90% heparin buffer A (25mM Hepes, 0.5mM EDTA, 5% glycerol, 1mM DTT, pH7) / 10% B (A + 2M NaCl, pH7), Ccg was eluted by a 90-minute gradient of 10% to 70% B

Twin strep tagged-SaUGI: The entire conserved 4-ORF operon was cloned into a pET21a vector with the twin strep tag (same sequence as above) at the N-terminus of SaUGI as it is the first protein encoded in the operon. The proteins were expressed and purified using the Strep Trap column the same as above, except that instead of loading the Strep Trap eluate onto a heparin column, the SaUGI eluate was added to a QFF column (Hi Prep^{TM} QFF 1610 20 ml GE). The column was washed with 90% QFF buffer A (20 mM Tris, 0.5mM EDTA, 5% glycerol,

1mM DTT, pH8)/10% QFF buffer B (A + 2M NaCl, pH8). The protein was then eluted with a 90-minute gradient of 10% to 70% buffer B.

His tagged SaUGI: SaUGI was cloned into pET21a using the NdeI and XhoI sites and has a C-terminal His_6 tag. Protein expression was done in the Rosetta (DE3) plysS strain. Protein expression was induced by adding 0.5mM IPTG at 37C OD600 of 0.7, followed by 2 hours additional growth. The cell pellets were resuspended in Nickel buffer (50 mM $NaPO_4$, 1M NaCl, 5% glycerol 1 mM DTT, pH 7.5) + protease inhibitor cocktail (1 tablet/liter of cells-cOmplete Mini, EDTA free) and lysozyme (final concentration 200 ug/ml). Cells were lysed via sonication for 1 min 3 times. Debris was then pelleted at 20,000 rpm in an ss-34 rotor for 1 hour, after which the supernatant was filtered using a 0.22-micron filter, then loaded onto the nickel column. The column was washed with nickel buffer, then the protein was eluted with nickel buffer plus a gradient of 0-500 mM imidazole. Peak fractions were rechromatographed on the nickel column to improve purity and then further purified with the QFF column as above.

Testing interaction between SaUGI and Ccg

A column was made using 2 ml (1 ml final column) of Streptactin-XT Sepharose (Iba sciences). This was then equilibrated with low salt strep buffer described above. Purified twin strep tagged Ccg was then incubated sepharose at a final concentration of 0.5 mg/ml. The protein – Sepharose slurry was incubated on the column for 15 minutes with gentle rocking at 4°C. Everything that didn't bind is found in the flowthrough (Figure 2C). An equimolar amount of His tagged SaUGI was then added to the column and again incubated rocking at 4°C for 15 minutes. The column washed 3 times with 4 ml of low salt strep buffer. The protein was then eluted using low salt strep buffer + 2.5 mM Desthiobiotin.

Co-purification of SaUGI and SarX

SarX was cloned into the pET21a vector containing the whole operon downstream of SaUGI, replacing the peptide, Lp960 and Ccg. This created a construct with SaUGI with an N-terminal twin strep tag and SarX with a C-terminal His_6 tag. The proteins were expressed in Rosetta (DE3) plysS cells. Two 1-Liter cultures were grown at 37°C until OD_{600} reached ~0.7. Expression was then induced by adding 0.5mM IPTG, and the culture was shifted to 20C and grown overnight. The cell pellet from 1 liter was resuspended in low salt-strep buffer (100 mM Tris, 0.5 mM EDTA, 250 mM NaCl, 5% glycerol, 1 mM DTT, pH 8) + lysozyme (200 ug/ml) + 1 tablet protease inhibitor cocktail, while the cell pellet from the other liter was resuspended in low salt nickel buffer (50 mM $NaPO_4$, 250mM NaCl, 5% glycerol, 1 mM DTT, pH 7.5) + lysozyme (200 ug/ml) + 1 tablet protease inhibitor cocktail. Cells were lysed via sonication for 1 minute 3 times. Cell debris was then pelleted at 20,000 RPM in an ss-24 rotor for 1 hour then the supernatant was filtered using a 0.22-micron filter and loaded onto the nickel or strep column, then washed with their respective low salt buffers. The proteins were then eluted using the respective strep or nickel column elution buffers discussed above.

Pull-downs/Mass-spectrometry

S. aureus RN4220 cell lysate was made by diluting an overnight 10 ml TSB liquid culture was added to 1 L TSB and growing at 37 to reach an OD of 0.7 The cells were then treated with 2 ug/ml of lysostaphin in TSM (50 mM tris-HCl, 0.5 M sucrose, 10 mM $MgCl_2$, pH 7.5) to remove cell surface proteins. These were “sticky” and if not removed would nonspecifically bind the strep column at later steps. The resulting protoplasts were then lysed by resuspending the protoplasts in strep buffer (100 mM Tris, 0.5 mM EDTA, 250 mM NaCl, 5% glycerol, 1 mM

DTT, pH 8) and then sonicating for 1 minute 3 times. The lysate was aliquoted, and flash frozen in liquid nitrogen. Purified twin strep tagged SaUGI was then combined with 1 ml of step buffer for a final concentration of 0.5 mg/ml. This was then added to 1 ml Streptactin-XT Sepharose (Iba sciences) that had equilibrated in strep buffer in an empty disposable column. The protein – Sepharose slurry was incubated with gentle rocking at 4C for 15 minutes. After draining excess liquid, 10 ml of RN4220 lysate was passed through the column. The column was washed 3 times with 15ml of strep buffer. Lastly the proteins were eluted with 1mL strep buffer plus 2.5 mM Desthiobiotin. The eluted proteins were then precipitated by adding 1 volume of 100% TCA to 4 volumes of protein sample. The samples were then incubated on ice for 1 hour. Samples were then spun at 13,000g for 15 minutes. The pellet was then washed with ice cold acetone 2 times. The pellet was then dried in a vacufuge. The precipitated pellet was then sent to the Taplin mass spectrometry facility at Harvard medical school. Samples were treated according to their standard protocol as follows:

Samples were reconstituted in 5 - 10 µl of HPLC solvent A (2.5% acetonitrile, 0.1% formic acid). A nano-scale reverse-phase HPLC capillary column was created by packing 2.6 µm C18 spherical silica beads into a fused silica capillary (100 µm inner diameter x ~30 cm length) with a flame-drawn tip[92]. After equilibrating the column each sample was loaded via a Famos auto sampler (LC Packings, San Francisco CA) onto the column. A gradient was formed, and peptides were eluted with increasing concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid). As peptides eluted, they were subjected to electrospray ionization and then entered an LTQ Orbitrap Velos Pro ion-trap mass spectrometer (Thermo Fisher Scientific). Peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences (and hence protein identity) were determined by

matching protein databases with the acquired fragmentation pattern by the software program, Sequest (Thermo Fisher Scientific)[93]. All databases include a reversed version of all the sequences and the data was filtered to between a one and two percent peptide false discovery rate.

Five parallel experiments were performed: an empty vector control, and 4 bait proteins (SaUGI, Ccg, SaPI bov1 Rep, SaPI bov1 primase). Peptides were searched against *S. aureus* and *E. coli* reference genomes (*E. coli* hits were removed from Table 4.3). Table 4.3 is ordered by the ratio of the "sum intensity" for each protein found in the SaUGI-as-bait experiment to the mean of the "sum intensity" for the 4 other samples. We reasoned that non-specifically sticky proteins would be more prevalent when a bait protein was present than in the empty vector control, and therefore that the best comparison would be to the mean from all the other samples rather than simply the empty vector (to avoid dividing by zero, entries of "not found" were replaced by the lowest reported value, 1.9E+03).

Glycosylase Assays:

Various concentrations of purified Ccg or hSMUG1(NEB) were incubated with synthetic oligonucleotides at a final concentration of 1uM oligonucleotide. The sequence of the oligonucleotides was: 5-TAGACATTGCCCTCGAGGTA X CATGGATC where X was T, U, 5-methylC (5mdC), 5-hydroxymethyl C (hmdC), 5-hydroxymethyl U (hmdU), or 2-amino dA. The buffer used was: 45 mM HEPES, 0.4 mM EDTA, 2% glycerol and 50 mM KCl. The final reaction volume was 50 ul. The reactions were incubated overnight (roughly12-16 hours) at 37°C. The next day 25 ul of NaOH was added to each sample and they were heated to 90°C for 30 minutes. The reactions were then cleaned using the Zymo: oligo clean and concentrator kit.

The DNA was eluted from the column with 15 ul of nuclease free water. 4 ul of loading dye was then added and 5 ul of each sample was loaded on a 10% TBE gel then stained with Ethidium bromide.

The oligonucleotides used in these assays were purchased from IDT except for hydroxy methyl dU (hmdU) which was kindly synthesized by Nan-Sheng Li and Joseph Piccirilli. The modified phosphoramidite was purchased from Glen Research and the oligonucleotide was prepared using an Expedite 8900 synthesizer. The oligonucleotide was deprotected with 28% ammonium hydroxide at 55 ºC for 17 hrs. After filtration and removal of ammonia by rotary evaporator, the mixture was extracted with dichloromethane and the aqueous layer was washed with dichloromethane (3 x 100 ul). The oligo was precipitated with ethanol and then checked by reverse phase HPLC. It was further purified by electrophoresis on a polyacrylamide gel. The band was excised from the gel after visualization by brief UV-shadowing.

Structural Analysis

RaptorX and later, Alphafold2, were used for protein structure prediction[94–96] The Dali server was used to find experimentally determined homologs for these predicted structures[97]. Snapgene was used to examine the genomes that contain SCC elements, and structural figures were made using PyMOL (www.pymol.org).

4.3 Results:

Ccg has glycosylase activity but does not bind SaUGI.

Ccg is the 3rd and largest protein in the operon. We used Alphafold2 to predict the structure of Ccg and found that it adopts a DNA glycosylase fold. The most closely related

experimentally determined structure was that of BdiUDG (from *Bradyrhizobium diazoefficiens*) (Figure 4.2A) [98]. Based on these results, we set out to determine if Ccg was an uracil DNA glycosylase. Because many glycosylases work efficiently on ssDNA substrates, we used a simple assay based on synthetic oligonucleotides containing a single deoxy uracil. The DNA backbone at the abasic site created by cleavage of the uracil – deoxyribose bond is readily cleaved under basic conditions. We used a commercially available UDGase, hSMUG1, as a positive control. Initial experiments showed weak activity on ssDNA containing dU (Figure 2B). This leads us to believe that Ccg is an active glycosylase as predicted, but that dU is not its ideal substrate. However, no activity was seen on a panel of potential alternative substrates: hydroxymethyl dU, hydroxymethyl dC, 5methyl dC, and 2 amino dA (Supplementary Figure S4.1).

As Ccg has weak uracil DNA glycosylase activity and is co-expressed with a known uracil DNA glycosylase inhibitor, SaUGI, we wanted to determine if SaUGI inhibited Ccg activity by binding the DNA binding pocket like it does with the chromosomally encoded *S. aureus* UDG. We found that Ccg did not co-purify with twin strep-tagged SaUGI (Figure 4.2C). Due to the high concentrations of protein required to test for Ccg glycosylase activity, we were unable to test if SaUGI inhibited the reaction. That SaUGI did not inhibit Ccg is not surprising as other glycosylases with similar predicted structures have been shown to not be inhibited by SaUGI and other glycosylase inhibitors. SaUGI and proteins like it have been shown to predominantly inhibit DNA glycosylase family 1 such as *S. aureus* UNG and the human homolog hUNG[98–101].

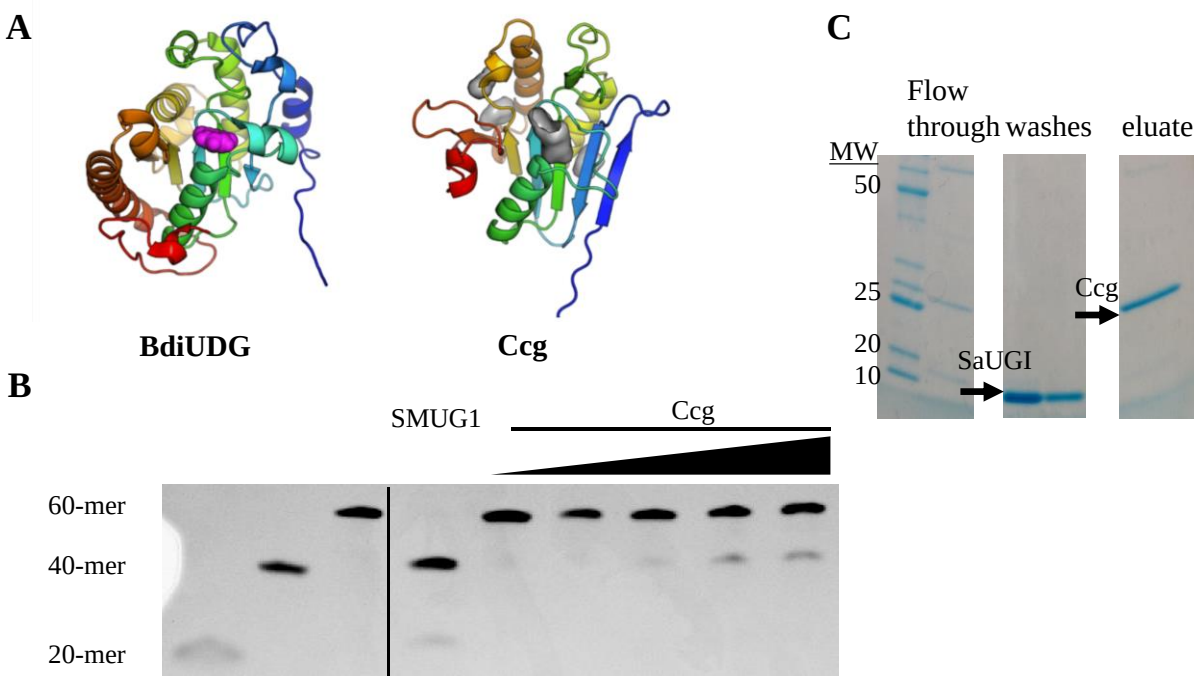


Figure 4.2: Ccg is a Uracil DNA glycosylase that does not interact with SaUGI

A: Ccg is predicted to have a Uracil DNA glycosylase-like fold. Left: Crystal structure of BdiUDG, shaded from blue (N-terminus) to red (C-terminus), with bound uracil shown in magenta. A small C-terminal extension was removed for clarity (PDB id 5gnw [98]). Right: alphafold2 model of Ccg from an SCC element similarly shaded from N to C, with a gray surface highlighting cavities and pockets.

B: Glycosylase assay with Ccg and SMUG1 as a control. Lanes1-3 20-med, 40-mer, and 60-mer oligos as size markers. Lane 4 is SMUG1 1:1 ratio of DNA to protein and lanes 5-9 are Ccg+ 1uM dU oligo. Concentrations go from 1uM, 5uM, 10uM, 16uM, and 22uM respectively.

C. SaUGI and Ccg do not interact. Twin Strep tagged Ccg bound to a strep column lane 1 contains molecular weight marker, lane 2 is he flowthrough of residual non-specific binding. Lane 3 shows His tagged SaUGI that is coming off in the washes. His tagged SaUGI MW= Twin strep tagged Ccg MW=. Empty lanes were removed from the gel image.

Proteins that interact with SaUGI

In part because there are only 7 conserved SCC proteins, we anticipated that SCC elements would need to utilize host proteins to assist in performing necessary functions. We therefore used pull-down assays to ask what host proteins SaUGI and Ccg might bind. Purified twin strep tagged Ccg and SaUGI proteins were used as bait to retrieve binding partners from *S.*

aureus lysate (from strain RN4220). Table 4.3 shows the results of a pull-down using twin strep tagged SaUGI as bait. The pull-down with tagged Ccg retrieved no clear binding partners. The results for the pull-down with SaUGI partially overlapped with those of a previously done pull-down using His-tagged SaUGI: both experiments retrieved the native UDG (*ung*) and LexA[85]. The fact that UDG was the top hit in our experiment (see Methods for an explanation of our ranking scheme) lends confidence to our results. In addition to UDG and LexA our top hits also included SarX, a transcription factor, and genes for a few proteins of unknown function. The most interesting of these interactions was the interaction with SarX as it is a transcription factor that will be discussed later. Note that UDG, LexA and SarX would also be within the top 10 hits if the list were ranked simply by the sum of intensity for each peak in the SaUGI pulldown, without accounting for non-specific binding as is down in the last column of Table 4.3.

	Gene Symbol	MWT (kDa)	Unique Peptides SaUGI	Total Peptides SaUGI	Sum Intensity SaUGI	Rank if sorted by Sum Intensity SaUGI	Sum Intensity SaUGI / ave (Sum Intensity other 4)
1	ung	24.9	13	106	9.40E+08	2	82275.71
2	SAOUHSC_02747	23.8	7	23	1.10E+07	13	5789.47
3	SAOUHSC_02649	64.5	11	29	8.30E+06	16	4368.42
4	lexA	23.3	10	74	1.30E+08	3	2527.95
5	sarX	14.2	7	24	1.80E+07	8	2424.24
6	SAOUHSC_00244	15.8	4	10	3.30E+06	31	250.00
7	SAOUHSC_02581	17.5	6	17	3.90E+07	6	212.01
8	femX	48.5	1	3	1.60E+05	141	41.29
9	rpmC	8.1	1	2	3.70E+05	111	31.03
10	SAOUHSC_01535	43.5	1	1	3.90E+04	190	20.53

Table 4.3: Proteins found to co-elute with SaUGI.

Effects of Overexpression of the SaUGI, Lp960, and Ccg

Like the other SCC conserved proteins, SaUGI, Lp960, and Ccg are expressed poorly or not at all in *S. aureus* when grown in lab conditions[36,102,103]. To combat this issue, we separately overexpressed each conserved operon from a plasmid in a MRSA strain (USA300) then utilized RNAseq to examine the resulting changes in gene expression. While some changes were observed when the first two conserved operons were overexpressed, overexpression of the 3rd operon led to the most expression changes (supplementary figure S4.2Aand B, Figure 4.3A). To further explore these changes, each ORF from operon 3 was overexpressed individually, except for the peptide. The results showed that the majority of the changes in gene expression were due to the expression of SaUGI rather than Lp960 or Ccg (Figure 4.3 B and C, Supplementary figure 4.2 C and D). Upon investigation of the genes that were differentially expressed, the results are similar to changes that have been observed in knockouts of the transcriptional regulators *mgrA* and *rot* [104–106]. However, the overlap was imperfect. The differences may reflect variations in strains or conditions, or that SaUGI is impacting something in the same regulatory pathway(s) as *rot* or *mgrA* but not directly impacting those proteins.

We further explored the effects of overexpressing SaUGI on *S. aureus* cells by looking at what phenotypic changes we would expect based on the genes that are differentially expressed. *LytN* and the adjacent *FmhC* were two of the most highly upregulated genes in our RNAseq data. LytN is a murein hydrolase thought to act during cell division, and FmhC builds peptidoglycan cross bridges that are resistant to LytN-mediated hydrolysis. It has been previously shown that overexpression of FmhC alleviates effects of overexpressing LytN[88]. FmhC has been shown to increase resistance to the cell wall hydrolase lysostaphin and overexpression of LytN has been shown to cause cell division defects[88,107]. Due to this we predicted an increase in lysostaphin

resistance and a growth defect upon SaUGI overexpression, however no changes in growth curves were observed when each of the 3rd operon ORFs were overexpressed individually and there was no change in lysostaphin resistance (Figure 4.3D and Supplementary figure 4.3S). This is likely due to LytN and FmhC counteracting each other.

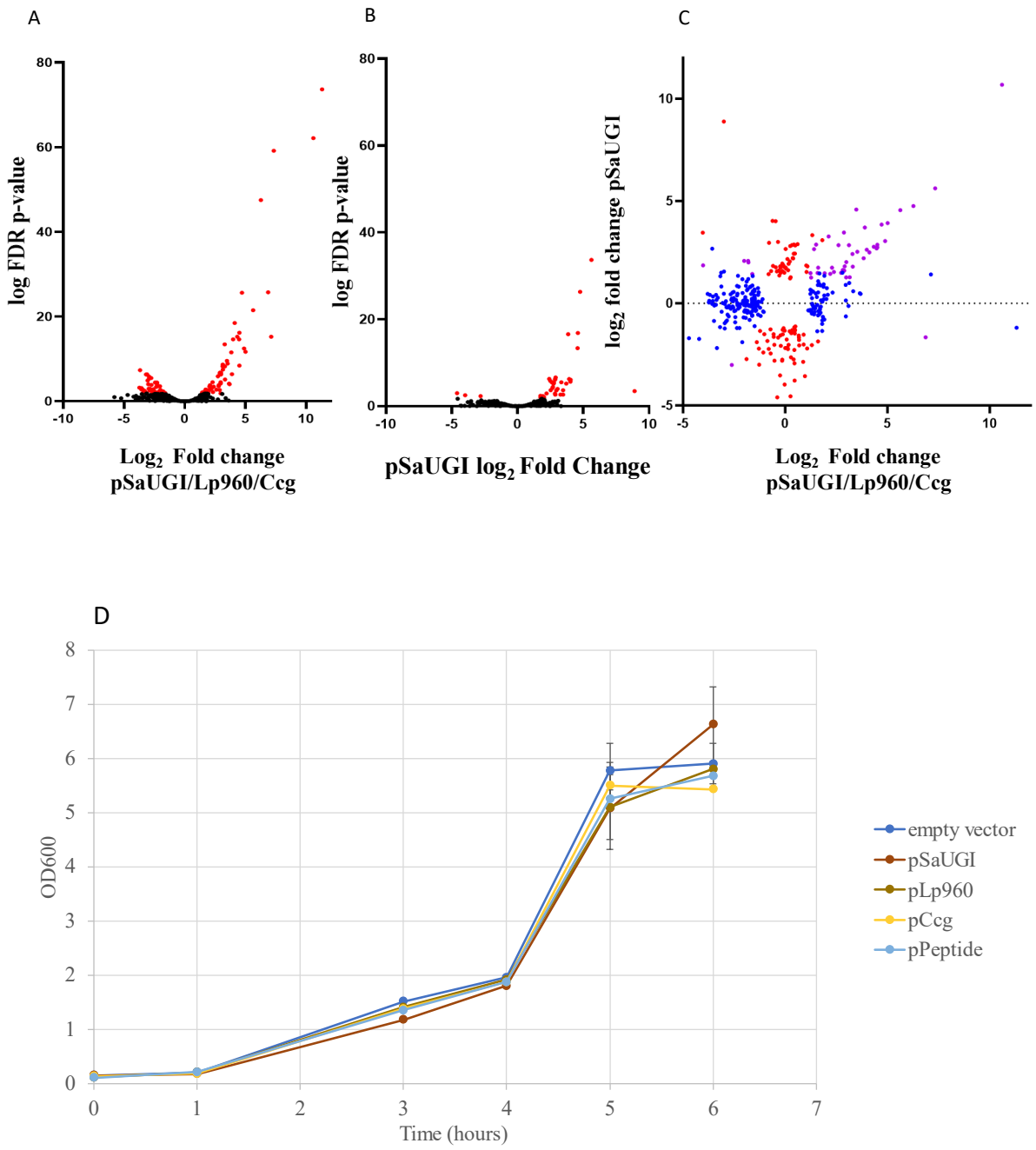


Figure 4.3: Effects of overexpressing operon proteins:

A: Differential gene expression changes comparing the whole operon expressed vs empty vector. Significant data points (red) were chosen with cut off of : Log_2 Fold change >2,<-2 and FDR p-value 0.01

B: Differential gene expression changes comparing the pSaUGI expressed vs empty vector. Significant data points (red) were chosen with cut off of : Log_2 Fold change >2,<-2 and FDR p-value 0 .01

C: Comparison of genes differentially expressed with SaUGI or the whole operon. Genes that were significantly changed with whole operon=blue, genes that were changed with SaUGI overexpression= red both=purple

D: growth curve of RN4220 with operon proteins and peptide overexpressed. N=3

SaUGI leads to a decrease in biofilm in a SarX-independent manner.

To further explore the changes in gene expression and biofilm produced by SaUGI expression we focused on the putative SaUGI-SarX interaction revealed in our pull-down experiments. SarX is a transcription factor that is activated by MgrA and regulates biofilm[108–110]. SarX regulates biofilm by activating expression of the *ica* operon which encodes genes that synthesize polysaccharide intercellular adhesin (PIA)[108,111]. Because SaUGI is a DNA mimic, we hypothesized that it might block the DNA binding surface of SarX. Figure 4.4A shows that SarX and SaUGI do co-purify, providing experimental evidence to support an interaction between these proteins.

We evaluated the amount of biofilm that was produced when SaUGI was overexpressed as compared to conditions known to lead to a decrease in biofilm *(Δica, ΔmgrA, and ΔsarX*). Overexpression of SaUGI led to a decrease in biofilm production. This was found to occur in both a MRSA strain (JE2) as well as a MSSA lab strain (RN4220) figure 4.4B and C. [40]Next, we investigated the impact that co-expressing SarX and SaUGI would have on biofilm production. We found that the decrease in biofilm observed in multiwell plates with SaUGI overexpression is not fully complemented by the overexpression of SarX (figure 4.4C). We also found that co-expressing SarX with SaUGI does not alleviate the effects of expressing SaUGI. These data indicate that although SarX regulates biofilm production and interacts with SaUGI, the decrease in biofilm in these assays is not primarily due to the SaUGI-SarX interaction.

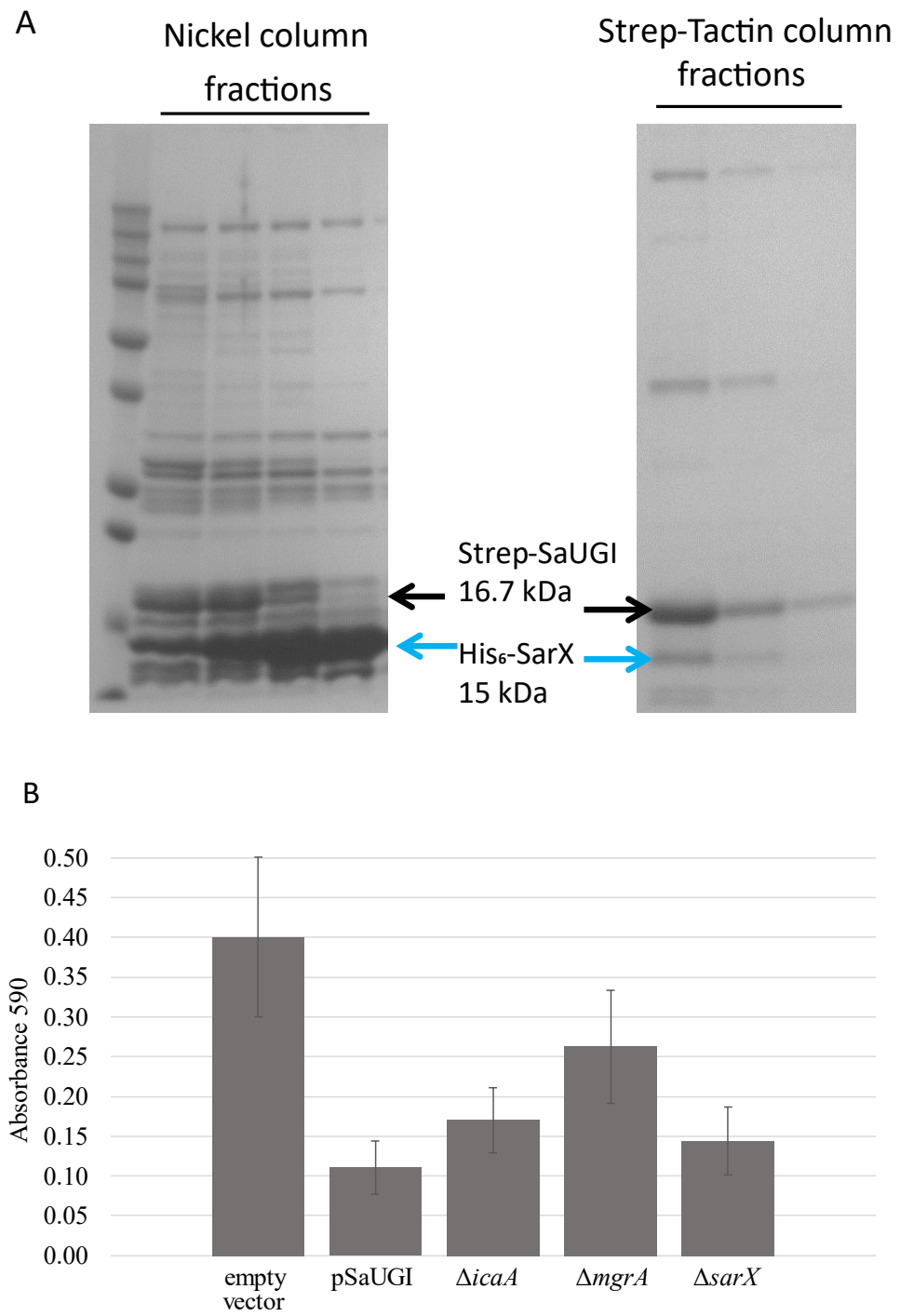


Figure 4.4: Exploring the interaction between SarX and SaUGI and its effects

A: Co-purification of twin strep tagged SaUGI and His-tagged SarX. Fractions eluted from the two different columns were analyzed by SDS PAGE and stained with Coomassie blue.

B: Quantification of biofilm biomass observed with overexpressing SaUGI in MRSA strain JE2. N=5

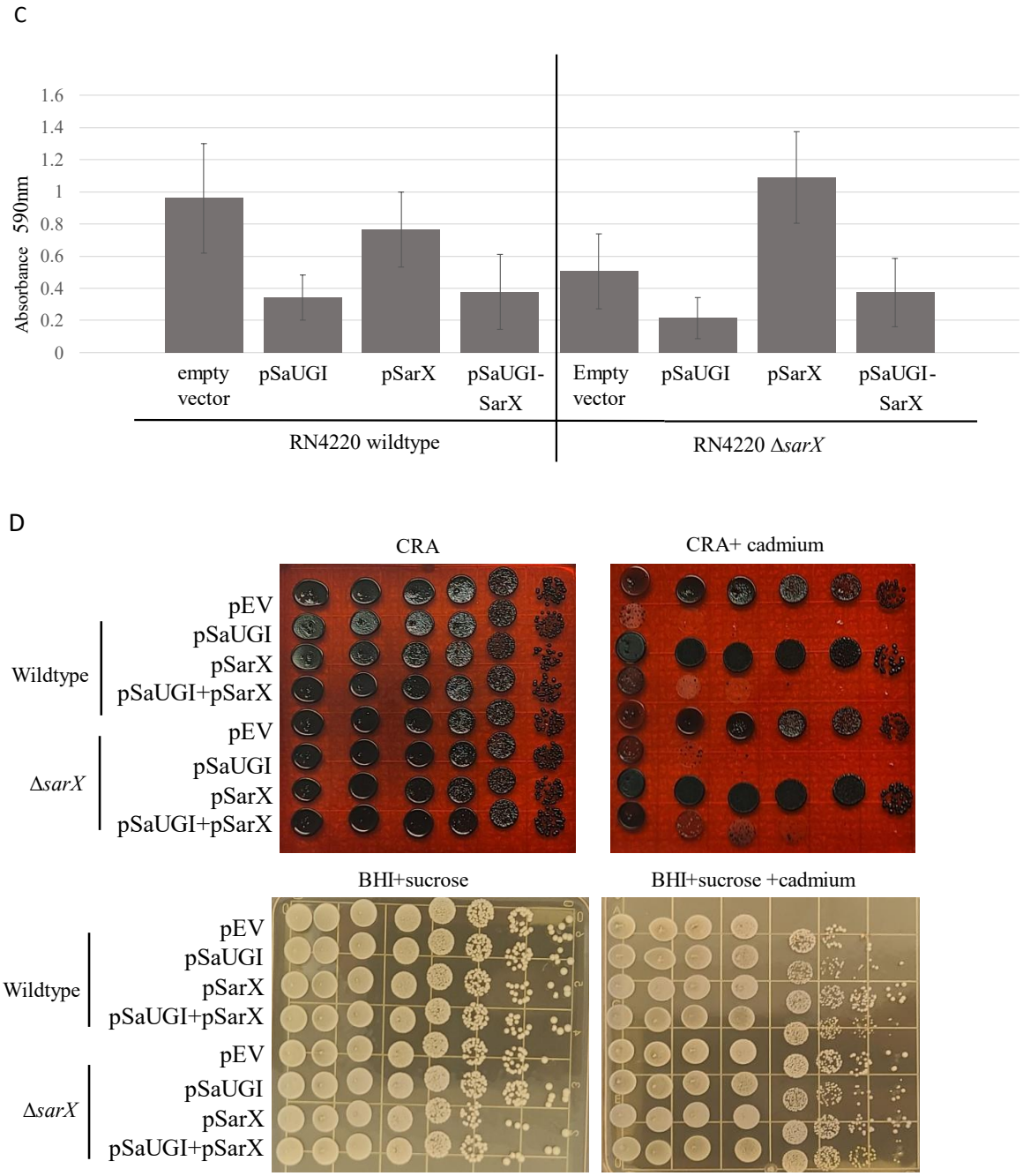
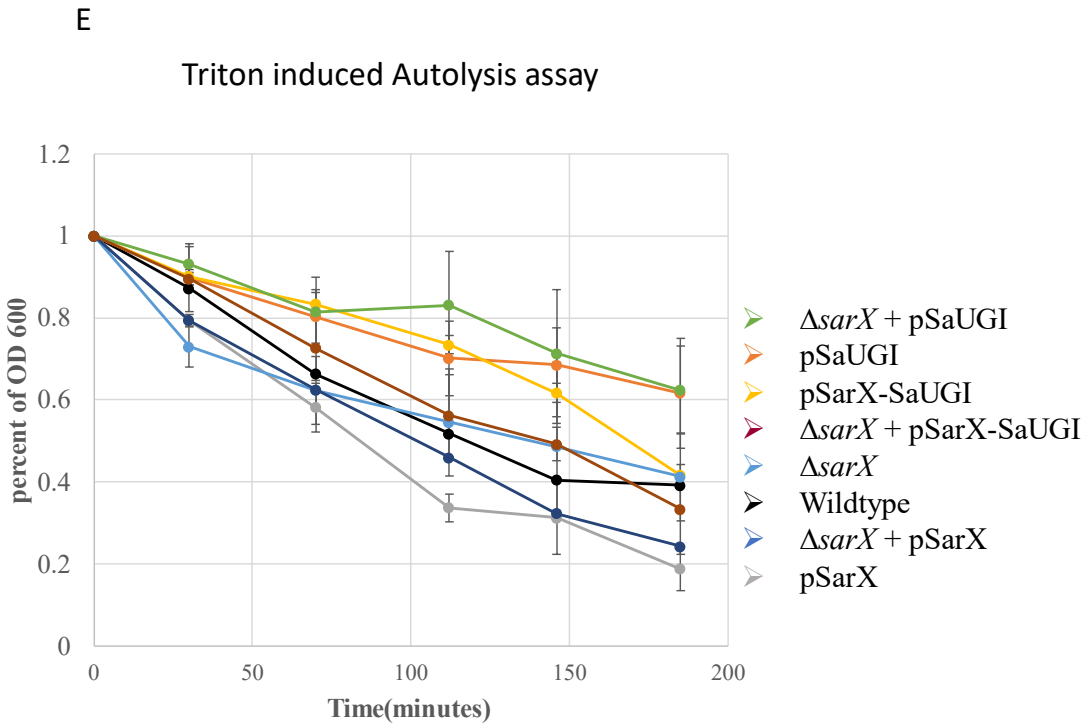


Figure 4.4 continued

C: Quantification of biofilm biomass observed with overexpressing SaUGI and SarX in MSSA strain RN422. N=5

D: Cells were serially diluted onto BHI plates supplemented with Congo red +sucrose (top) or BHI+ sucrose, with and without the addition of $CdCl_2$ to induce protein expression. Strain JE2 (MRSA). Experiment was repeated 3 times with similar results



.

Figure 4.4 continued
E: Triton induced autolysis of wildtype and JE2Δ*sarX* with overexpression plasmids (all samples included $CdCl_2$ to induce protein expression). N=3

SaUGI may block wall teichoic acid synthesis

Next, we tested the effects of SaUGI expression in a different type of assay. Changes in the type of biofilm produced by *S. aureus* and *S. epidermidis* can be visualized on congo red supplemented plates (CRA) [112,113] – the colonies turn black when PIA is produced. Furthermore, *S. aureus* grows poorly on CRA when there are defects in wall teichoic acid[114]. This is due to the fact that congo red inhibits a key enzyme in the synthesis of lipoteichoic acid, which is lethal in strains deficient in wall teichoic acid[115]. We found that wildtype *S. aureus* strain JE2 with the empty vector does lead to black colonies on CRA, and that neither deletion nor overexpression of SarX resulted in obvious phenotypic changes in this assay. However, overexpression of SaUGI

led to a striking growth defect on CRA (but not on BHI; Figure 4.4D). This phenotype indicates that the SaUGI overexpression may block some aspect of wall teichoic acid synthesis.

Overexpression of SaUGI protects against triton-induced autolysis

Another one of the most highly upregulated genes was *atlR*. *AtlR* encodes a MarA family transcriptional regulator that has been shown to repress expression of the major autolysin Atl [116]. We examined autolysis in wildtype and *sarX* knockout strains overexpressing SaUGI and/or SarX. Overexpression of SaUGI leads to a decrease in autolysis. Conversely, overexpression of SarX leads to an increase in autolysis (Figure 4.4E). While the increase in autolysis observed with SarX overexpression is alleviated by SaUGI overexpression, the decrease in autolysis observed with pSaUGI overexpression is similar in a *sarX* knockout to that in a WT background. This indicates that while SaUGI is impacting both biofilm and autolysis, these effects are not primarily due to its interaction with SarX.

Comparing the gene expression changes in strains overexpressing SaUGI and a sarX knockout.

To explore which changes observed in the original RNA-seq experiment were due to SaUGI binding SarX, we compared the changes observed when SaUGI is overexpressed in the wildtype background vs. in a *sarX* mutant. Overall, the gene expression changes observed in the *sarX* knockout were very similar to the changes observed in the wildtype (Figure 5A-C). To determine which gene expression changes are likely due to the SaUGI-SarX interaction, we looked for genes whose expression changed in the *sarX* knockout (relative to the WT background, with empty vector) and also changed with SaUGI expression (in the WT background, relative to empty vector in the WT background), but did not change when SaUGI

was overexpressed in the *sarX* knockout background (relative to empty vector in a SarX knockout) (Figure 4.5E). There were 64 genes that fit these criteria. These are shown in black in (Figure 4.5E) and found in Table 4.5. When the expression changes found in the *sarX* knockout were plotted vs the changes observed when SaUGI is overexpressed, the 64 genes thought to be due to the SaUGI-SarX interaction were on the diagonal indicating that the changes are likely related.

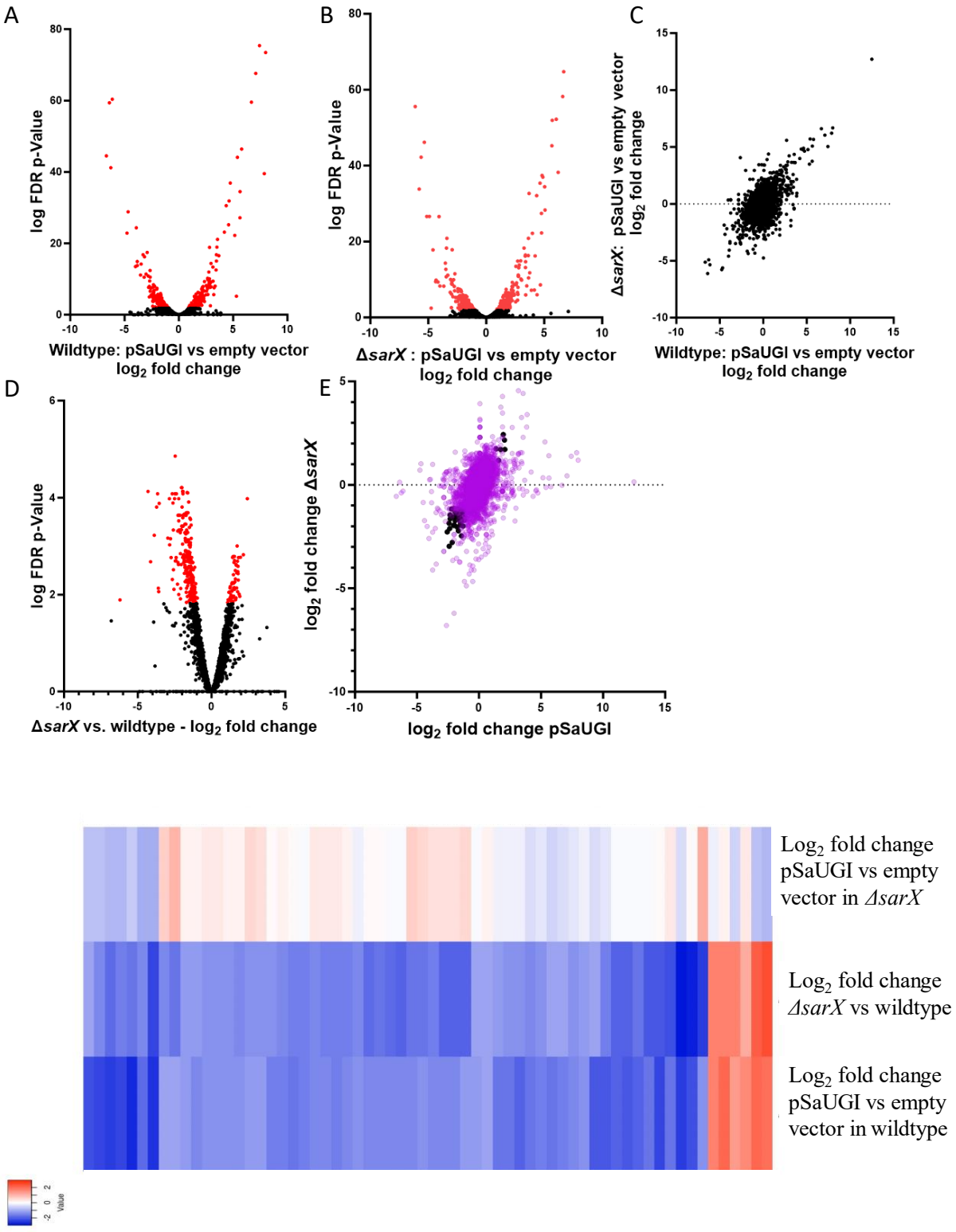


Figure 4.5: Gene expression changes when SaUGI is overexpressed in JE2 *ΔsarX*

A, B Differential gene expression of pSaUGI vs empty vector in wildtype or JE2Δ*sarX*.

C. Comparison of the effects of expressing SaUGI in a WT vs. sarX knockout JE2 strain. The changes in RNA abundance seen for pSaUGI vs. empty vector in the ΔsarX background (vertical axis) were very similar but not completely identical to those seen for pSaUGI vs. empty vector in the WT background (horizontal axis).

D: Differential expression of genes in a *sarX* mutant JE2 compared to wildtype JE2. Significant data points (red) were chosen with cut off of : Log2 Fold change 2,-2 and FDR p-value .01

E: Comparison of log2 fold change SaUGI overexpression in a wildtype strain and log2 fold change Δ*sarX.* Purple data points represent the all genes and black data points represent the genes that were changed in the *sarX* knockout and with SaUGI overexpression, but were not changed when SaUGI was overexpressed in the *sarX* knockout.

F: Heat map showing the genes that were differentially expressed in in the *sarX* knockout and with SaUGI overexpression but were not changed when SaUGI was overexpressed in the *sarX* knockout (black dots in E). Genes are found in Table 4.5

4.4 Discussion:

Previously little was known about this 4-ORF operon. Its ubiquitous nature among SCC elements and its continued maintenance leads us to believe that the proteins it encodes are important for the stability and/or horizontal transfer of the element. The CA-MRSA USA300 strains are some of the more pathogenic MRSA strains that are isolated[8,117]. These strains contain 2 copies of this operon: one in the SCC*mec* typeIV cassette and one in the adjacent ACME cassette[118]. While SaUGI had been crystallized and a function in inhibiting base excision repair of uracil had been shown, it was unknown how that could relate to SCC element function[85]. The functions of other 3 ORFs had not previously been addressed.

Ccg contains DUF1643. Our structural modeling and biochemistry show that this domain of unknown function is indeed a glycosylase. Ccg is encoded in the same operon as SaUGI but did not form a detectable complex with SaUGI, which was not surprising as SaUGI-like glycosylase inhibitors were shown to not inhibit similar glycosylase proteins[98,99,119]. It remains

unclear why SCC elements encode Ccg, and the weak activity of Ccg in removing uracil from DNA implies that its true substrate has yet to be discovered. The variable cargo of SCC elements has recently been shown to include numerous anti-phage functions and it may be that Ccg's true substrate is a modified nucleotide found in certain bacteriophages[120].

We further explored SaUGI activity through overexpression experiments. Uracil DNA glycosylase inhibitors like SaUGI had been previously shown to improve phage replication, specifically phages that encode dU in their DNA[76,121,122]. In that case blocking uracil excision prevented phage DNA damage. Additionally, non dU encoding phage have uracil DNA glycosylase inhibitors if they replicate with a ssDNA intermediate. As replication of SCC elements has not been observed we worked to identify other "jobs" that SaUGI could be doing to assist the element or off-target effects that could be occurring that have a net positive effect for the element. Because the whole operon is maintained in all SCC elements and other mobile elements such as ACME, the genes are likely expressed at some point and benefit the SCC or ACME element. Overexpression of SaUGI led to an increase in expression of the both the native chromosomal and the ACME *arc* operons (arcABDC). The *arc* operon encodes proteins for the arginine deaminase pathway which is important for growth in anaerobic conditions[123]. One study found that by deleting ACME there was a decrease in pathogenicity[124]. The upregulation of the *arc* and *opp* operons on ACME could be why the operon and specifically SaUGI is acquired and maintained in ACME.

As we found that SaUGI bound the transcription factor SarX and SarX regulates biofilm we examined this interaction further[108,110,111]. PIA, protein, and nucleic acid are the major components of *S. aureus* biofilms[105,125–129]. While SaUGI binds SarX it did not lead to a decrease in *ica* operon expression nor was the biofilm decrease observed with SaUGI overexpression

complemented with additional SarX overexpression. This leads us to conclude that the decrease in biofilm observed with SaUGI overexpression is not due to the interaction with SarX and a decrease in PIA. Conversely, cell surface proteins that have been shown to be involved in biofilm production such as SraP, Ebh, and SasG are in fact upregulated with SaUGI overexpression[106,130–132]. These results together with the decrease in autolysis leads us to believe that the decrease in biofilm observed with SaUGI overexpression is likely due to a loss of extracellular nucleic acid. Further work examining the structure of the biofilm is needed to examine this further.

The loss of biofilm observed with SaUGI overexpression may be leading to biofilm dispersal. As stated previously SCC elements have been shown to horizontally transfer in a biofilm via natural competence. Upon acquisition of the SCC element, expression of SaUGI in the recipient cell could trigger biofilm dispersal and dissemination of SCC+ cells. This is supported by the observance of PSMs being upregulated with SaUGI overexpression. PSMs have been previously shown to trigger biofilm dispersal[105]. Expression of SaUGI in a recipient cell could be preventing acquisition of an additional SCC element by dispersing the biofilm preventing transfer. Further work is needed to identify when SaUGI is expressed and how expression impacts transfer.

Lastly we saw a large growth defect when bacteria overexpressing SaUGI were grown on CRA plates. It has been previously shown that bacteria that have defect in wall teichoic acid have growth defects on CRA media[114]. This aligns with the RNAseq data, in which we found a decrease in the amount of TagA which is one of the proteins involved it wall teichoic acid synthesis[133].

While the changes that we have observed in autolysis, biofilm and the vast majority of the gene expression changes were not due to the SarX-SaUGI interaction, some of the changes were. It was interesting as SarX was previously believed to be inactive in log phase (where our RNAseq samples were collected) and has been described as a stationary phase factor[108,109,111]. The changes observed show that there are genes that are being impacted by SarX even at the lower levels of transcription observed in early log phase. Further work can be done to examine the effect of SaUGI-SarX interaction at other growth phases.

Table 4.4: Gene expression changes when SaUGI is overexpressed

Name	Fold change	FDR p-value
saugi	5748.07	9.48E-231
ureB	254.17	3.38E-74
ureA	234.31	2.94E-40
ebh	172.52	3.84E-76
lytN	136.06	2.41E-68
ureC	103.09	2.85E-60
ureE	55.25	3.84E-47
ureF	49.85	3.09E-35
fmtB	49.53	6.79E-28
galE	42.05	7.51E-45
feoA	39.70	6.45E-06
SAUSA300_RS00685	35.49	5.89E-23
fmhC	26.74	1.25E-37
SAUSA300_RS00690	24.68	1.34E-32
ureG	23.99	6.05E-26
SAUSA300_RS12385	20.57	2.71E-31
SAUSA300_RS06585	18.37	7.21E-24
sarV	12.82	2.91E-17
atlR	11.78	7.60E-22
SAUSA300_RS00695	11.38	1.09E-19
asp1	11.04	1.53E-17
sasA	10.98	8.69E-16
cap1B	10.79	2.82E-13
dprA	10.11	9.03E-09

Table 4.4 continued

Name	Fold change	FDR p-value
SAUSA300_RS13140	9.52	2.75E-08
hisG	9.32	2.23E-06
mcsB	8.88	3.92E-10
SAUSA300_RS08525	8.80	4.28E-14
yut	7.85	1.44E-08
SAUSA300_RS14530	7.81	4.14E-15
secY2	7.76	6.16E-10
argG	7.47	3.02E-03
asp2	7.38	2.14E-08
mcsA	7.30	1.52E-09
ddh	7.11	1.27E-19
sasC	7.06	8.47E-11
SAUSA300_RS08530	7.05	9.30E-10
asp3	6.99	5.85E-09
citZ	6.83	2.21E-09
mccB	6.79	3.03E-10
SAUSA300_RS08535	6.76	1.07E-12
SAUSA300_RS06715	6.35	2.10E-06
mntH	6.32	1.34E-10
SAUSA300_RS08610	6.29	6.13E-11
mccA	6.25	1.80E-08
cap1A	6.22	5.12E-08
pxpB	6.16	1.81E-08
SAUSA300_RS13670	5.87	2.28E-07
pxpA	5.76	6.67E-09
hisB	5.61	2.59E-06
SAUSA300_RS13810	5.14	2.02E-04
nfrA	5.12	5.13E-10
recA	5.11	2.21E-07
icd	4.91	5.70E-06
hisD	4.90	2.36E-05
hypR	4.82	4.30E-05
clpC	4.71	1.28E-08
clpB	4.56	9.22E-08
SAUSA300_RS07975	4.55	1.01E-05
ctsR	4.55	2.21E-07
cap1C	4.49	1.93E-08
groES	4.46	1.48E-05
uvrA	4.41	1.97E-06
hisC	4.40	4.44E-06

Table 4.4 continued

Name	Fold change	FDR p-value
uvrB	4.37	6.92E-07
trpF	4.35	2.17E-03
SAUSA300_RS14210	4.26	1.69E-04
hfq	4.26	6.34E-08
SAUSA300_RS06720	4.23	1.13E-04
hisZ	4.18	1.95E-03
sasG	4.16	1.32E-03
dnaJ	4.09	1.96E-06
ebpS	4.05	2.03E-05
hisH	3.99	2.41E-04
dnaK	3.97	1.16E-06
nixA	3.93	4.47E-06
SAUSA300_RS10165	3.89	8.90E-04
fepC	3.88	1.68E-04
efeB	3.85	1.77E-03
iruO	3.84	4.36E-06
grpE	3.79	8.06E-07
hrcA	3.79	2.74E-05
SAUSA300_RS03080	3.58	5.97E-04
cinA	3.52	1.31E-05
hisA	3.47	1.18E-03
frp	3.43	5.24E-07
SAUSA300_RS12805	3.42	9.97E-05
mobB	3.34	4.23E-03
prmA	3.28	1.88E-05
trpC	3.24	7.45E-03
isaB	3.22	4.97E-05
trpD	3.21	8.83E-03
SAUSA300_RS11835	3.18	1.13E-05
umuC	3.11	8.26E-03
menF	3.10	7.06E-04
hslV	3.07	8.48E-03
SAUSA300_RS13260	3.02	7.29E-04
secA	3.02	1.52E-06
cysE	2.97	1.25E-04
fdhD	2.94	1.35E-04
groL	2.90	3.33E-03
SAUSA300_RS13845	2.88	2.66E-04
SAUSA300_RS14115	2.88	7.06E-04
dhaL	2.87	6.24E-03

Table 4.4 continued

Name	Fold change	FDR p-value
dinB	2.86	6.86E-03
mntH	2.85	4.65E-04
clpP	2.85	4.45E-05
opp-4A	2.83	1.17E-02
miaA	2.82	5.97E-04
tarF	2.81	2.85E-04
hmrA	2.78	9.27E-05
SAUSA300_RS11485	2.78	9.81E-04
pepA1	2.77	2.62E-03
hisF	2.77	4.78E-03
hxlA	2.75	2.16E-03
acnA	2.75	9.88E-06
metC	2.74	2.51E-03
moeB	2.71	8.96E-04
trxB	2.68	6.63E-06
gpxA2	2.63	1.71E-03
SAUSA300_RS07600	2.61	7.63E-03
SAUSA300_RS03985	2.60	8.78E-03
menH	2.60	1.11E-04
SAUSA300_RS13995	2.60	9.06E-03
SAUSA300_RS00700	2.56	2.25E-03
SAUSA300_RS14615	2.56	1.01E-02
SAUSA300_RS04635	2.55	2.63E-03
lytM	2.55	2.08E-03
lexA	2.54	5.27E-03
trpB	2.54	6.06E-03
SAUSA300_RS11430	2.49	5.04E-03
SAUSA300_RS01090	2.48	4.77E-03
clfA	2.47	8.77E-03
menD	2.47	5.81E-03
SAUSA300_RS01230	2.46	2.86E-03
SAUSA300_RS13675	2.44	2.84E-03
sufU	2.43	2.47E-03
SAUSA300_RS13735	2.41	1.35E-02
ydbM	2.39	2.24E-03
SAUSA300_RS09235	2.38	2.00E-03
cysS	2.37	9.08E-03
moaB	2.36	3.27E-03
SAUSA300_RS14030	2.34	9.30E-03
ribF	2.33	2.13E-03

Table 4.4 continued

Name	Fold change	FDR p-value
tarI	2.32	1.36E-02
SAUSA300_RS05120	2.31	1.08E-02
SAUSA300_RS04550	2.29	5.92E-03
polX	2.27	5.54E-03
hxlB	2.26	3.43E-03
hchA	2.25	4.61E-04
msaA	2.24	4.66E-03
moaD	2.23	7.45E-03
SAUSA300_RS04865	2.22	3.20E-03
SAUSA300_RS09225	2.18	1.22E-02
bfmBAA	2.18	6.93E-03
SAUSA300_RS03910	2.14	5.00E-03
yjbH	2.12	5.81E-03
SAUSA300_RS02105	2.10	1.17E-02
truB	2.09	1.34E-02
hmp	2.06	8.78E-03
ilvE	2.03	1.06E-02
ccrB	2.00	1.18E-02
tkt	2.00	2.39E-03
ftsH	1.96	8.52E-03
gndA	1.88	8.95E-03
saeS	-1.92	1.42E-02
SAUSA300_RS02930	-1.94	1.01E-02
tcaA	-2.01	7.90E-03
plsC	-2.03	9.23E-03
SAUSA300_RS12235	-2.13	1.17E-02
SAUSA300_RS02905	-2.16	1.16E-02
opp-3B	-2.18	8.83E-03
ldh1	-2.18	2.39E-03
glcC	-2.20	1.36E-02
sarS	-2.24	3.21E-03
SAUSA300_RS10920	-2.24	5.22E-03
sirC	-2.24	8.76E-03
glpQ1	-2.26	1.17E-02
SAUSA300_RS02970	-2.26	4.56E-03
SAUSA300_RS04105	-2.27	2.45E-03
tagA	-2.27	1.08E-02
aapA	-2.29	1.90E-03
SAUSA300_RS12515	-2.30	3.36E-03
SAUSA300_RS12650	-2.30	5.05E-03

Table 4.4 continued

Name	Fold change	FDR p-value
SAUSA300_RS06385	-2.31	1.05E-02
SAUSA300_RS00250	-2.32	2.82E-03
SAUSA300_RS00055	-2.33	3.36E-03
selX	-2.35	5.75E-03
moaC	-2.37	5.00E-03
spn	-2.38	1.31E-02
SAUSA300_RS12875	-2.41	1.46E-02
SAUSA300_RS01875	-2.42	7.13E-03
SAUSA300_RS01560	-2.42	1.28E-02
uppP	-2.44	1.34E-02
SAUSA300_RS12765	-2.44	3.09E-03
lytS	-2.46	3.91E-03
cozEa	-2.48	6.65E-03
manA	-2.48	7.76E-03
cdsA	-2.49	1.39E-03
eap	-2.49	3.36E-03
SAUSA300_RS10520	-2.50	9.04E-03
rnz	-2.52	7.45E-04
mazE	-2.52	1.38E-02
SAUSA300_RS12735	-2.52	1.09E-02
eta	-2.52	5.05E-03
saeQ	-2.53	1.24E-02
SAUSA300_RS12200	-2.53	6.65E-03
SAUSA300_RS14665	-2.54	2.82E-03
galU	-2.54	8.78E-03
SAUSA300_RS12685	-2.55	9.06E-03
SAUSA300_RS06675	-2.56	9.58E-04
SAUSA300_RS02935	-2.57	3.68E-04
SAUSA300_RS10215	-2.58	8.39E-03
ctaA	-2.58	5.77E-03
SAUSA300_RS14325	-2.60	1.90E-03
SAUSA300_RS01060	-2.60	2.93E-03
SAUSA300_RS02000	-2.61	1.71E-03
comC	-2.61	5.58E-03
vwb	-2.62	1.30E-02
SAUSA300_RS13625	-2.62	1.13E-03
tnp	-2.63	6.21E-03
putP	-2.65	2.17E-03
SAUSA300_RS02265	-2.66	3.56E-03
scdA	-2.66	5.81E-03

Table 4.4 continued

Name	Fold change	FDR p-value
sdpB	-2.66	1.03E-03
SAUSA300_RS03990	-2.67	5.54E-03
arsR	-2.67	3.07E-03
SAUSA300_RS15490	-2.68	6.09E-03
SAUSA300_RS11305	-2.69	4.50E-03
SAUSA300_RS13380	-2.69	2.31E-04
tenA	-2.70	9.06E-03
SAUSA300_RS01080	-2.71	2.82E-03
crcB	-2.73	8.52E-03
saeP	-2.74	2.55E-03
sdpC	-2.74	7.30E-04
SAUSA300_RS10510	-2.74	1.59E-04
proP	-2.76	2.97E-04
SAUSA300_RS09900	-2.76	2.17E-04
rnhA	-2.77	1.18E-02
ulaA	-2.77	9.97E-03
cudT	-2.78	8.63E-04
nuc2	-2.79	1.03E-02
SAUSA300_RS09860	-2.81	3.24E-03
SAUSA300_RS04215	-2.81	7.00E-04
bstA	-2.82	8.34E-03
plsY	-2.83	3.86E-03
pcp	-2.83	1.55E-04
SAUSA300_RS10780	-2.84	6.57E-04
ssbB	-2.84	2.82E-04
adsA	-2.85	4.62E-04
cspC	-2.85	8.22E-03
plc	-2.85	2.24E-03
SAUSA300_RS04210	-2.86	8.64E-04
bglR	-2.86	8.96E-04
sasD	-2.87	1.52E-03
comK	-2.88	6.09E-03
SAUSA300_RS05185	-2.88	8.96E-04
SAUSA300_RS03425	-2.89	6.89E-04
SAUSA300_RS12975	-2.90	5.72E-04
ssaA	-2.91	2.70E-04
SAUSA300_RS09630	-2.94	2.01E-03
abc-f	-2.95	2.01E-05
SAUSA300_RS14405	-2.96	2.84E-03
SAUSA300_RS05685	-2.96	8.47E-04

Table 4.4 continued

Name	Fold change	FDR p-value
SAUSA300_RS03895	-2.97	1.52E-03
gpmA	-3.01	8.87E-04
SAUSA300_RS06535	-3.01	8.59E-03
dgk	-3.02	1.76E-04
SAUSA300_RS01620	-3.03	7.30E-04
SAUSA300_RS05990	-3.03	7.92E-04
SAUSA300_RS04405	-3.03	2.22E-03
glpQ	-3.04	5.75E-03
sigS	-3.06	6.89E-04
SAUSA300_RS09535	-3.06	7.44E-04
SAUSA300_RS15095	-3.07	1.55E-04
SAUSA300_RS13920	-3.08	4.30E-04
SAUSA300_RS02005	-3.10	4.30E-03
aroD	-3.12	1.84E-04
nnrD	-3.12	5.86E-05
SAUSA300_RS06570	-3.13	3.91E-04
scpA	-3.15	3.71E-06
SAUSA300_RS01810	-3.15	8.24E-05
SAUSA300_RS13580	-3.16	5.22E-03
SAUSA300_RS04440	-3.17	1.35E-04
ktrB	-3.21	2.19E-06
lctP2	-3.22	1.10E-06
SAUSA300_RS04145	-3.23	5.05E-03
mroQ	-3.25	1.26E-03
hlb-1	-3.29	3.81E-03
SAUSA300_RS01475	-3.31	5.32E-04
gltT	-3.31	1.31E-05
speG	-3.35	5.17E-05
SAUSA300_RS01175	-3.35	1.01E-05
SAUSA300_RS10790	-3.35	1.30E-04
SAUSA300_RS13485	-3.36	7.19E-04
cspB	-3.39	7.01E-05
SAUSA300_RS15710	-3.40	1.50E-04
SAUSA300_RS16000	-3.41	1.86E-03
SAUSA300_RS12545	-3.41	7.05E-05
SAUSA300_RS03995	-3.49	7.96E-05
fosB	-3.51	2.33E-03
lytR	-3.58	1.44E-05
sarZ	-3.58	6.01E-05
sodM	-3.58	1.57E-04

Table 4.4 continued

Name	Fold change	FDR p-value
SAUSA300_RS02190	-3.59	1.11E-04
SAUSA300_RS11540	-3.59	6.81E-05
lukS-PV	-3.61	4.71E-05
SAUSA300_RS09835	-3.63	3.76E-04
SAUSA300_RS05705	-3.65	2.54E-05
SAUSA300_RS15960	-3.65	6.31E-05
lctP1	-3.65	1.61E-04
SAUSA300_RS04395	-3.66	2.43E-04
sarR	-3.69	5.86E-05
SAUSA300_RS05715	-3.75	4.81E-06
isdA	-3.75	9.36E-05
nupC	-3.75	8.47E-04
SAUSA300_RS03490	-3.82	3.86E-05
SAUSA300_RS13160	-3.82	1.97E-06
SAUSA300_RS10180	-3.82	5.80E-05
SAUSA300_RS12965	-3.84	1.50E-04
SAUSA300_RS04705	-3.88	6.12E-04
SAUSA300_RS01470	-3.89	3.39E-05
narK	-3.96	6.55E-04
SAUSA300_RS01615	-3.99	5.00E-05
sarU	-4.00	8.78E-03
SAUSA300_RS04220	-4.05	3.56E-08
SAUSA300_RS13720	-4.08	6.96E-08
SAUSA300_RS13425	-4.09	6.12E-05
SAUSA300_RS09325	-4.14	1.22E-05
SAUSA300_RS03215	-4.16	1.36E-05
nptA	-4.16	1.37E-05
SAUSA300_RS07815	-4.16	1.52E-04
narG	-4.21	3.83E-05
brnQ_1	-4.24	3.32E-05
scrA	-4.29	1.91E-05
hyl	-4.31	5.06E-06
mutS	-4.47	6.07E-06
lrgB	-4.49	6.65E-03
lip1	-4.50	1.00E-04
SAUSA300_RS06580	-4.58	1.02E-02
glvR	-4.69	1.02E-08
adhP	-4.79	1.43E-07
esaA	-4.95	6.87E-07
scn	-5.01	2.28E-07

Table 4.4 continued

Name	Fold change	FDR p-value
SAUSA300_RS04780	-5.03	3.71E-06
SAUSA300_RS04180	-5.05	1.69E-04
nrdG	-5.06	3.69E-09
SAUSA300_RS13715	-5.15	2.14E-05
SAUSA300_RS11565	-5.19	1.76E-08
SAUSA300_RS06555	-5.21	1.94E-04
nrdD	-5.22	1.16E-06
SAUSA300_RS12970	-5.24	1.30E-04
glpT	-5.30	4.18E-03
SAUSA300_RS02035	-5.32	3.77E-06
SAUSA300_RS13770	-5.32	9.13E-06
adhE	-5.33	1.81E-06
SAUSA300_RS06500	-5.52	6.63E-06
isaA	-5.92	3.37E-06
pyrR	-6.01	3.87E-08
mntB	-6.04	2.86E-08
mntA	-6.22	2.82E-11
lukG	-6.25	2.65E-07
mntC	-6.25	3.88E-13
ftnA	-6.44	3.54E-10
SAUSA300_RS14595	-6.58	1.12E-08
aur	-6.60	9.88E-06
lip2	-6.89	2.14E-08
SAUSA300_RS01485	-7.02	1.36E-11
norB	-7.54	3.47E-18
citM	-8.11	3.92E-12
sak	-8.94	5.90E-17
sbi	-9.14	3.10E-11
esxA	-9.85	1.53E-17
SAUSA300_RS05050	-11.06	5.98E-15
SAUSA300_RS01130	-11.57	7.42E-12
SAUSA300_RS02185	-14.15	1.53E-14
SAUSA300_RS01635	-14.22	1.17E-15
SAUSA300_RS07245	-15.09	4.19E-25
lukH	-15.94	3.26E-14
tdcB	-25.59	1.39E-29
ald_1	-27.31	1.32E-23
SAUSA300_RS13610	-69.88	4.39E-61
nuc	-76.99	6.08E-42
SAUSA300_RS13605	-84.30	4.01E-60

Table 4.4 continued

Name	Fold change	FDR p-value
chp	-101.84	3.09E-45

Table 4.5: Genes likely chnaged via the SaUGI-SarX interaction:

Gene name	Log2 fold change pSaUGI vs empty vector	Log2 fold change ΔsarX vs wildtype	Log2 fold change ΔsarX (pSaUGI vs empty vector)
nnrD	-1.643	-1.158	-0.368
speG	-1.743	-1.420	-0.190
sirC	-1.166	-1.701	0.959
SAUSA300_RS01175	-1.743	-1.241	-0.208
ldh1	-1.122	-1.069	0.233
glcC	-1.140	-1.491	0.644
lrgB	-2.167	-2.783	0.022
bglR	-1.515	-1.742	0.027
ulaA	-1.469	-1.790	-0.037
glpT	-2.406	-2.965	-0.444
SAUSA300_RS01810	-1.658	-1.522	0.135
SAUSA300_RS02265	-1.410	-1.734	0.561
SAUSA300_RS15095	-1.618	-1.564	0.313
cysE	1.570	1.178	0.282
SAUSA300_RS03425	-1.531	-1.559	0.168
ssaA	-1.933	-1.728	-0.049
uppP	-1.285	-1.993	0.479
SAUSA300_RS04180	-2.336	-1.636	-0.905
cspC	-1.511	-1.833	0.072
SAUSA300_RS04210	-1.514	-1.621	0.342
SAUSA300_RS04240	-1.642	-1.421	0.069
SAUSA300_RS04405	-1.600	-1.598	-0.039
SAUSA300_RS04780	-2.330	-1.456	-0.712
ctaA	-1.368	-1.472	-0.136
SAUSA300_RS05685	-1.566	-1.388	-0.352
SAUSA300_RS06720	2.081	1.713	0.165
plsY	-1.500	-1.918	0.456
rnhA	-1.468	-1.646	0.451
SAUSA300_RS07600	1.384	1.742	-0.504
SAUSA300_RS07815	-2.058	-1.959	-0.078
aapA	-1.196	-1.373	0.401
SAUSA300_RS09835	-1.858	-1.484	-0.976

Table 4.5 continued

Gene name	Log2 fold change pSaUGI vs empty vector	Log2 fold change ΔsarX vs wildtype	Log2 fold change ΔsarX (pSaUGI vs empty vector)
SAUSA300_RS09860	-1.490	-1.942	-0.108
SAUSA300_RS10165	1.959	2.428	-0.919
putP	-1.406	-1.375	0.231
SAUSA300_RS10510	-1.457	-1.232	0.153
SAUSA300_RS10520	-1.319	-1.541	0.236
SAUSA300_RS15490	-1.423	-2.465	1.070
SAUSA300_RS10780	-1.504	-1.941	0.473
mroQ	-1.701	-1.651	0.077
ssbB	-1.508	-1.497	0.316
mobB	1.740	1.707	-0.176
moaC	-1.247	-1.367	0.519
sdpB	-1.414	-1.214	-0.215
SAUSA300_RS12515	-1.199	-1.092	-0.107
SAUSA300_RS12545	-1.770	-1.429	-0.391
glvR	-2.230	-1.170	-0.774
SAUSA300_RS12650	-1.204	-1.250	0.158
SAUSA300_RS12685	-1.350	-1.381	0.315
gltT	-1.726	-1.177	-0.505
narK	-1.986	-1.957	-0.009
bcr	-1.933	-1.311	-0.214
SAUSA300_RS13425	-2.031	-1.542	-0.271
sasG	2.055	2.164	-0.810
SAUSA300_RS13580	-1.661	-2.223	0.272
SAUSA300_RS13625	-1.390	-1.382	0.290
SAUSA300_RS13715	-2.366	-2.109	0.095
SAUSA300_RS13720	-2.029	-1.149	-0.453
ssaA2	-2.412	-1.854	-0.906
isaA	-2.566	-2.273	-1.044
cudT	-1.475	-1.308	-0.241
bstA	-1.496	-1.724	0.668
SAUSA300_RS14595	-2.717	-1.772	-0.664
cspB	-1.760	-2.037	-0.013

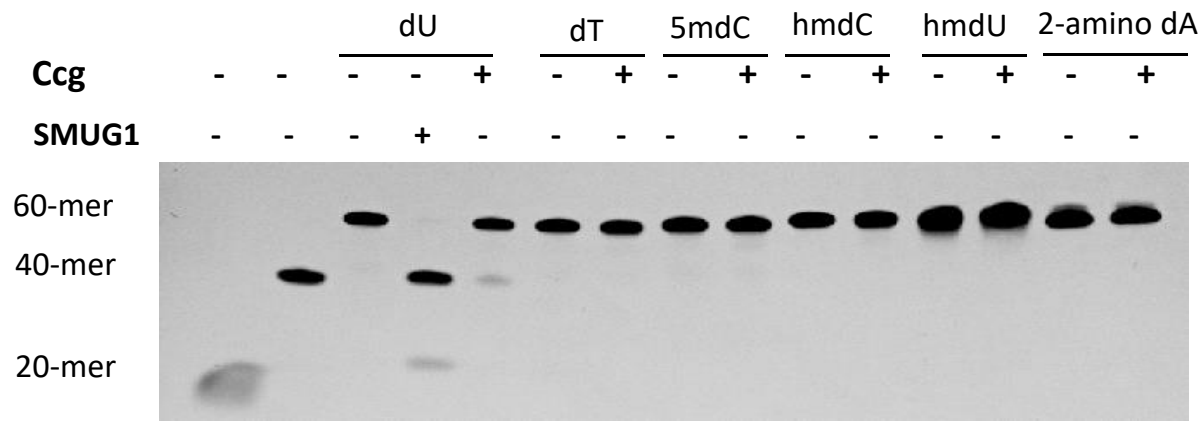


Figure S4.1: Glycosylase assay with different oligo substrates
Oligos with different nucleotide substrates were incubated with either Ccg or SMUG1 as a control. The first 3 lanes are oligo size markers.

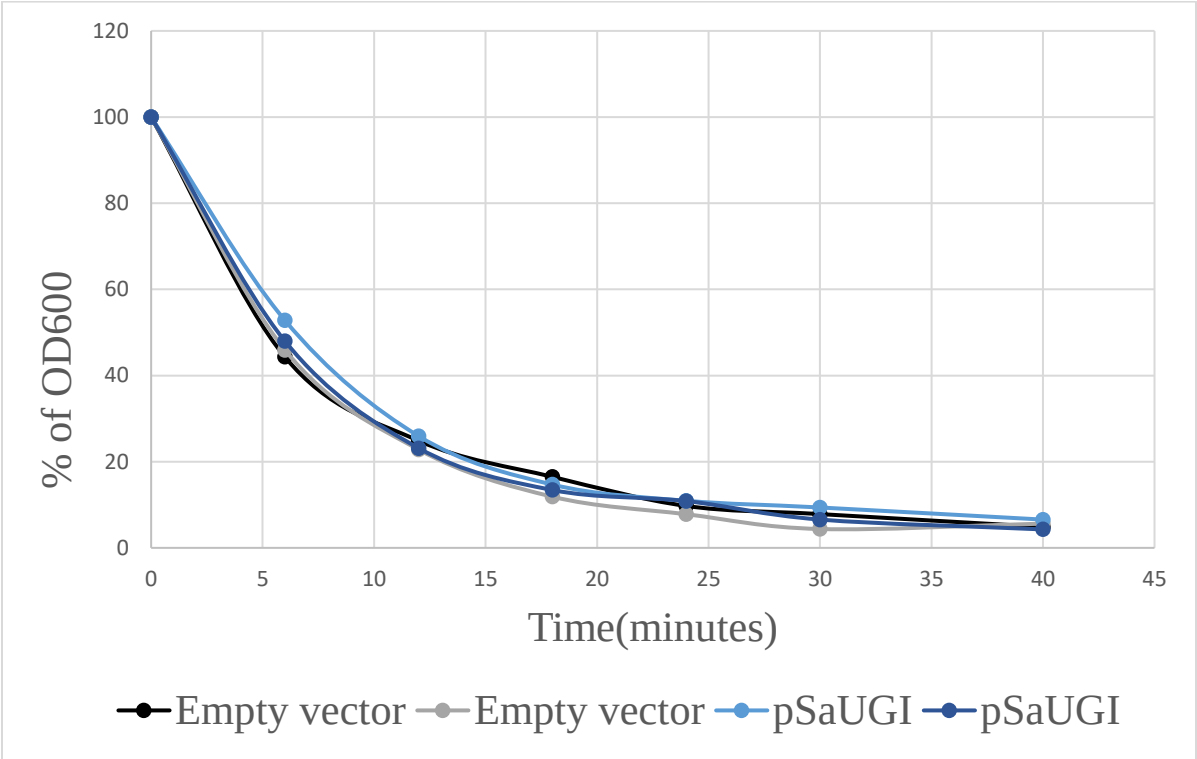


Figure S4.2: Lysostaphin resistance assay

Growth of cells in 5ug/ml of lysostaphin. Each line represents a separate culture.

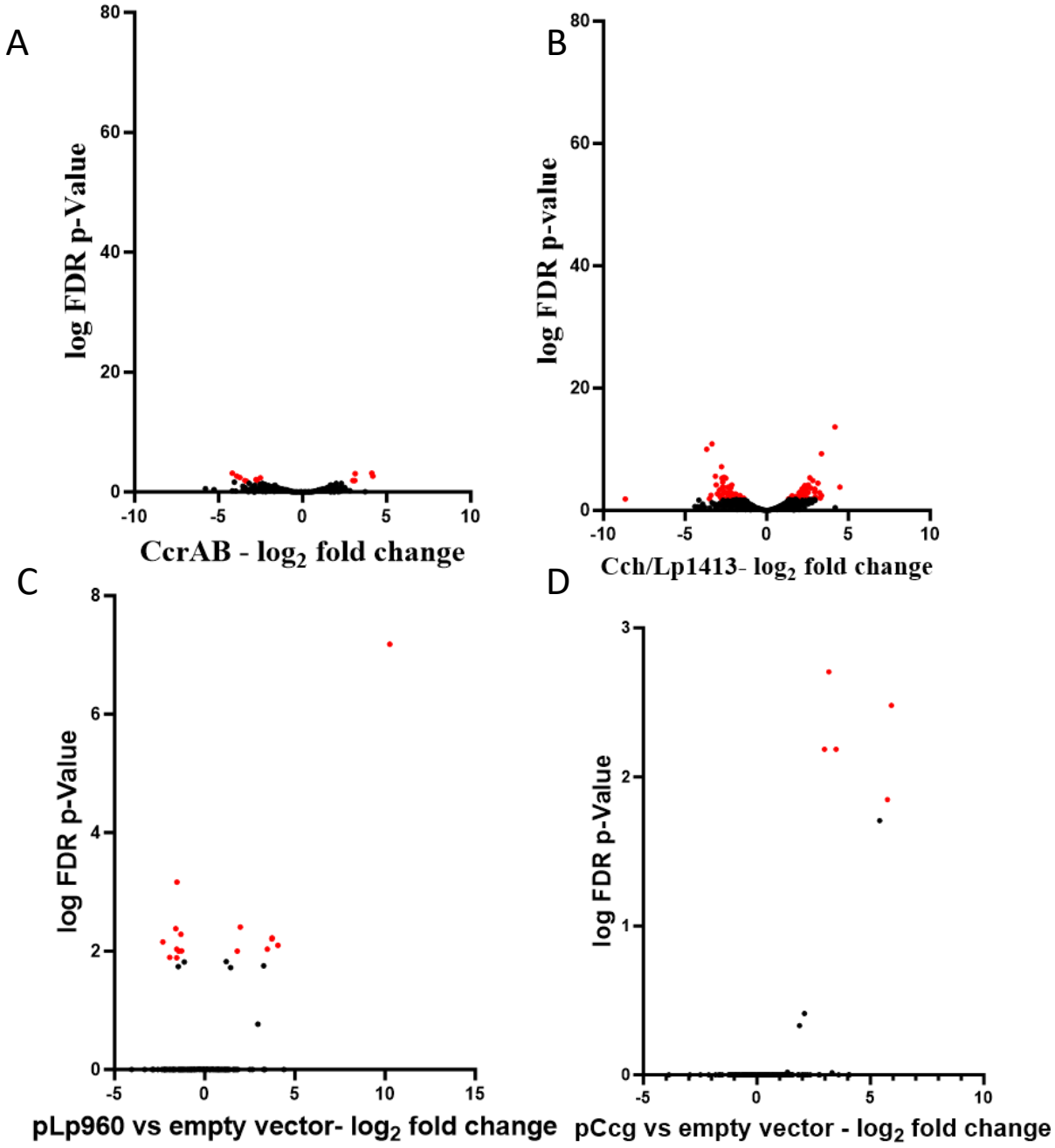


Figure S4.3: Differential gene expression of SCC conserved genes vs empty vector.
Red data points cutoff: FDR p-value <.01 and $\log_2$ fold change >2, <-2.

Chapter 5: Conclusions and Future Directions

Overall, the purpose of the preceding experiments was to better understand the function of the conserved SCC proteins and how they can be involved in the maintenance and mobilization of SCC elements. The conserved proteins are not expressed under lab conditions; however, I was able to determine mutations is the *lp1413/cch* promoter that lead to expression[35,37,61]. While the first operon in both patterns of SCC elements are homologous to mobile element replicative helicases, I was unable to observe replication[49,52,134]. Lastly in chapter 4 I focus on the last operon in the SCC conserved gene locus. There I show that there are many genes that are differentially expressed with overexpression of the conserved genes, and interaction between SaUGI and a transcription factor, a loss of biofilm, and the description of Ccg as a glycosylase. Based on these results and the observation that SCC elements are transferred in a recombinase dependent manner in a biofilm[48]. I propose the mechanism outlined in Figure 5.1. In a biofilm a cell with an integrated SCC element is lysed. The nucleic acid within the cell is then spewed to the extracellular matrix. Within the recipient SCC- cell natural competence is triggered and the cell uptakes DNA fragments from the environment. A DNA fragment containing the SCC element is then imported into the cell as a single strand. We hypothesize that from this single strand the conserved SCC genes are then expressed. In the case of pattern 1 the presence of Lp1413 which is a ssDNA binding protein can protect the single stranded SCC from cellular nucleases. The helicases, cch and cch2 can then work with cellular machinery to copy the second strand. In the case of pattern 2 elements ccpol-mp would act as a primase. The double stranded linear DNA can then be circularized by *ccrABC*. This would lead to the creation of attSCC and would remove flanking genomic DNA. The element would then be integrated into the chromosome. An additional role that we can elucidate is that of SaUGI. From the literature we know that uracil DNA glycosylase inhibitors assist non dU containing phages in

replication through the blocking of base excision repair[76]. It is believed that this is due to the large single strand intermediate. SaUGI could be improving SCC stability in the new cell in the same ways.

Future research exploring this hypothesis will be done first by checking if the conserved proteins are expressed from a single strand promoter. To test this, we have discussed the use of a conjugative plasmid. Similar to the work done in chapter 2 the SCC promoters will be fused to a reporter on a conjugative plasmid as they enter the cell as a single strand. We can then observe reporter signal in the recipient cell upon transfer.

Other future work will be aimed at repeating the natural competence transfer assay. Initial experiments have not been successful, however through troubleshooting the assay once transfer is observed, we can then test which of the SCC conserved genes are needed for transfer and if the expression of absence improves or diminishes said transfer. Specifically looking at SaUGI first. As the biofilm appears to be necessary for SCC transfer and SaUGI overexpression leads to a decrease in biofilm the effect of SaUGI expression in donor and recipient cells will be followed up.

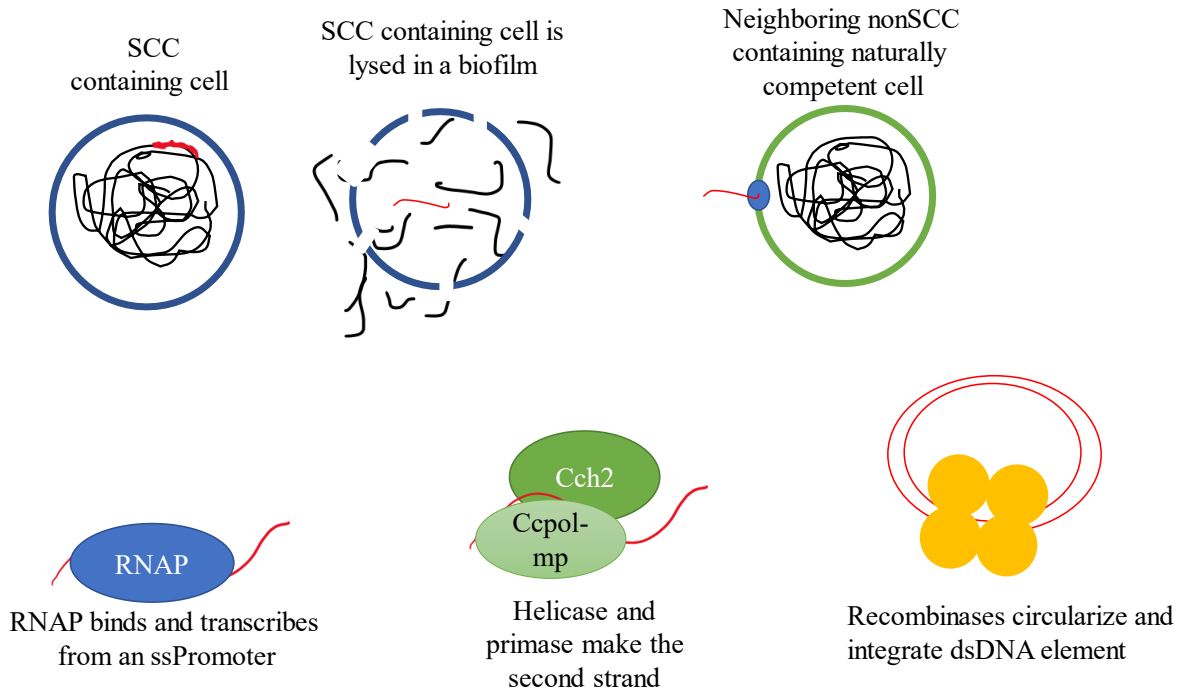


Figure 5.1 Model for SCC transfer and gene expression

References

1. O'Gara, J. P. Into the storm: Chasing the opportunistic pathogen Staphylococcus aureus from skin colonisation to life-threatening infections. *Environ. Microbiol.* **19**, 3823–3833 (2017).
2. Tong, S. Y. C., Davis, J. S., Eichenberger, E., Holland, T. L. & Fowler, V. G. Staphylococcus aureus Infections: Epidemiology, Pathophysiology, Clinical Manifestations, and Management. *Clin. Microbiol. Rev.* **28**, 603–661 (2015).
3. Kuehnert, M. J. *et al.* Prevalence of Staphylococcus aureus Nasal Colonization in the United States, 2001–2002. *J. Infect. Dis.* **193**, 172–179 (2006).
4. Malachowa, N. & DeLeo, F. R. Mobile genetic elements of Staphylococcus aureus. *Cell. Mol. Life Sci.* **67**, 3057–3071 (2010).
5. Gardete, S. & Tomasz, A. Mechanisms of vancomycin resistance in Staphylococcus aureus. *J. Clin. Invest.* **124**, 2836–2840 (2014).
6. Bai, Z. *et al.* Identification of Methicillin-Resistant Staphylococcus Aureus From Methicillin-Sensitive Staphylococcus Aureus and Molecular Characterization in Quanzhou, China. *Front. Cell Dev. Biol.* **9**, (2021).
7. Huang, H. *et al.* Comparisons of Community-Associated Methicillin-Resistant Staphylococcus aureus (MRSA) and Hospital-Associated MSRA Infections in Sacramento, California. *J. Clin. Microbiol.* **44**, 2423–2427 (2006).
8. Li, M. *et al.* Comparative analysis of virulence and toxin expression of global community-associated methicillin-resistant Staphylococcus aureus strains. *J. Infect. Dis.* **202**, 1866–1876 (2010).
9. Lindsay, J. A., Ruzin, A., Ross, H. F., Kurepina, N. & Novick, R. P. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol. Microbiol.* **29**, 527–543 (1998).
10. Jevons, M. P. "Celbenin" - resistant Staphylococci. *Br. Med. J.* **1**, 124–125 (1961).
11. Centers for Disease Control and Prevention (CDC). Four pediatric deaths from community-acquired methicillin-resistant Staphylococcus aureus — Minnesota and North Dakota, 1997-1999. *MMWR Morb. Mortal. Wkly. Rep.* **48**, 707–710 (1999).
12. Saravolatz, L. D., Markowitz, N., Arking, L., Pohlod, D. & Fisher, E. Methicillin-resistant Staphylococcus aureus. Epidemiologic observations during a community-acquired outbreak. *Ann. Intern. Med.* **96**, 11–16 (1982).
13. Beck, W. D., Berger-Bächi, B. & Kayser, F. H. Additional DNA in methicillin-resistant Staphylococcus aureus and molecular cloning of mec-specific DNA. *J. Bacteriol.* **165**, 373–378 (1986).
14. Haque, N. *et al.* Methicillin resistant Staphylococcus epidermidis. *Mymensingh Med. J. MMJ* **20**, 326–331 (2011).
15. Choo, E. J. Community-Associated Methicillin-Resistant Staphylococcus aureus in Nosocomial Infections. *Infect. Chemother.* **49**, 158–159 (2017).
16. Kateete, D. P. *et al.* CA-MRSA and HA-MRSA coexist in community and hospital settings in Uganda. *Antimicrob. Resist. Infect. Control* **8**, 94 (2019).

17. Ito, T., Katayama, Y. & Hiramatsu, K. Cloning and Nucleotide Sequence Determination of the Entire mec DNA of Pre-Methicillin-Resistant Staphylococcus aureus N315. *Antimicrob. Agents Chemother.* **43**, 1449–1458 (1999).

18. Katayama, Y., Ito, T. & Hiramatsu, K. A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in Staphylococcus aureus. *Antimicrob. Agents Chemother.* **44**, 1549–1555 (2000).

19. Hartman, B. & Tomasz, A. Altered penicillin-binding proteins in methicillin-resistant strains of Staphylococcus aureus. *Antimicrob. Agents Chemother.* **19**, 726–735 (1981).

20. Zapun, A., Contreras-Martel, C. & Vernet, T. Penicillin-binding proteins and β-lactam resistance. *FEMS Microbiol. Rev.* **32**, 361–385 (2008).

21. Lim, D. & Strynadka, N. C. J. Structural basis for the beta lactam resistance of PBP2a from methicillin-resistant Staphylococcus aureus. *Nat. Struct. Biol.* **9**, 870–876 (2002).

22. Fishovitz, J., Hermoso, J. A., Chang, M. & Mobashery, S. Penicillin-binding protein 2a of methicillin-resistant Staphylococcus aureus. *IUBMB Life* **66**, 572–577 (2014).

23. International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC). Classification of staphylococcal cassette chromosome mec (SCCmec): guidelines for reporting novel SCCmec elements. *Antimicrob. Agents Chemother.* **53**, 4961–4967 (2009).

24. Hanssen, A.-M. & Ericson Sollid, J. U. SCCmec in staphylococci: genes on the move. *FEMS Immunol. Med. Microbiol.* **46**, 8–20 (2006).

25. Saber, H., Jasni, A. S., Jamaluddin, T. Z. M. T. & Ibrahim, R. A Review of Staphylococcal Cassette Chromosome mec (SCCmec) Types in Coagulase-Negative Staphylococci (CoNS) Species. *Malays. J. Med. Sci. MJMS* **24**, 7–18 (2017).

26. Jena, S., Panda, S., Nayak, K. C. & Singh, D. V. Identification of Major Sequence Types among Multidrug-Resistant Staphylococcus epidermidis Strains Isolated from Infected Eyes and Healthy Conjunctiva. *Front. Microbiol.* **8**, (2017).

27. Bloemendaal, A. L. A., Brouwer, E. C. & Fluit, A. C. Methicillin Resistance Transfer from Staphylocccus epidermidis to Methicillin-Susceptible Staphylococcus aureus in a Patient during Antibiotic Therapy. *PLoS ONE* **5**, e11841 (2010).

28. Firth, N., Jensen, S. O., Kwong, S. M., Skurray, R. A. & Ramsay, J. P. Staphylococcal Plasmids, Transposable and Integrative Elements. *Microbiol. Spectr.* **6**, (2018).

29. Ramsay, J. P. *et al.* An updated view of plasmid conjugation and mobilization in Staphylococcus. *Mob. Genet. Elem.* **6**, e1208317 (2016).

30. Goessweiner-Mohr, N., Arends, K., Keller, W. & Grohmann, E. Conjugation in Gram-Positive Bacteria. *Microbiol. Spectr.* **2**, PLAS-0004-2013 (2014).

31. Fernández-López, C. *et al.* Mobilizable Rolling-Circle Replicating Plasmids from Gram-Positive Bacteria: A Low-Cost Conjugative Transfer. *Microbiol. Spectr.* **2**, (2014).

32. Johnson, C. M. & Grossman, A. D. Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annu. Rev. Genet.* **49**, 577–601 (2015).

33. Ray, M. D., Boundy, S. & Archer, G. L. Transfer of the Methicillin Resistance Genomic Island Among Staphylococci by Conjugation. *Mol. Microbiol.* **100**, 675–685 (2016).

34. Barberis-Maino, L., Berger-Bächi, B., Weber, H., Beck, W. D. & Kayser, F. H. IS431, a staphylococcal insertion sequence-like element related to IS26 from Proteus vulgaris. *Gene* **59**, 107–113 (1987).

35. Zhang, S., Ma, R., Liu, X., Zhang, X. & Sun, B. Modulation of ccrAB Expression and SCCmec Excision by an Inverted Repeat Element and SarS in Methicillin-Resistant Staphylococcus aureus. *Antimicrob. Agents Chemother.* **59**, 6223–6232 (2015).

36. Zhang, S., Ma, R., Liu, X., Zhang, X. & Sun, B. Modulation of ccrAB Expression and SCCmec Excision by an Inverted Repeat Element and SarS in Methicillin-Resistant Staphylococcus aureus. *Antimicrob. Agents Chemother.* **59**, 6223–6232 (2015).

37. Zhang, S., Shu, X. & Sun, B. SigmaB regulates ccrAB expression and SCCmec excision in methicillin-resistant Staphylococcus aureus. *Int. J. Med. Microbiol. IJMM* **306**, 406–414 (2016).

38. Dearborn, A. D. & Dokland, T. Mobilization of pathogenicity islands by Staphylococcus aureus strain Newman bacteriophages. *Bacteriophage* **2**, 70–78 (2012).

39. Fillol-Salom, A. *et al.* Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. *ISME J.* **12**, 2114–2128 (2018).

40. Tormo-Más, M. Á. *et al.* Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature* **465**, 779–782 (2010).

41. Cohen, S. & Sweeney, H. M. Effect of the prophage and penicillinase plasmid of the recipient strain upon the transduction and the stability of methicillin resistance in Staphylococcus aureus. *J. Bacteriol.* **116**, 803–811 (1973).

42. Cohen, S. & Sweeney, H. M. Transduction of Methicillin Resistance in Staphylococcus aureus Dependent on an Unusual Specificity of the Recipient Strain. *J. Bacteriol.* **104**, 1158–1167 (1970).

43. Uchiyama, J. *et al.* Intragenus generalized transduction in Staphylococcus spp. by a novel giant phage. *ISME J.* **8**, 1949–1952 (2014).

44. Scharn, C. R., Tenover, F. C. & Goering, R. V. Transduction of staphylococcal cassette chromosome mec elements between strains of Staphylococcus aureus. *Antimicrob. Agents Chemother.* **57**, 5233–5238 (2013).

45. Fagerlund, A., Granum, P. E. & Håvarstein, L. S. Staphylococcus aureus competence genes: mapping of the SigH, ComK1 and ComK2 regulons by transcriptome sequencing. *Mol. Microbiol.* **94**, 557–579 (2014).

46. Morikawa, K. *et al.* Expression of a cryptic secondary sigma factor gene unveils natural competence for DNA transformation in Staphylococcus aureus. *PLoS Pathog.* **8**, e1003003 (2012).

47. Morikawa, K. *et al.* A new staphylococcal sigma factor in the conserved gene cassette: functional significance and implication for the evolutionary processes. *Genes Cells Devoted Mol. Cell. Mech.* **8**, 699–712 (2003).

48. Maree, M. *et al.* Natural transformation allows transfer of SCCmec-mediated methicillin resistance in Staphylococcus aureus biofilms. *Nat. Commun.* **13**, 2477 (2022).

49. Mir-Sanchis, I. *et al.* Staphylococcal SCCmec elements encode an active MCM-like helicase and thus may be replicative. *Nat. Struct. Mol. Biol.* **23**, 891–898 (2016).

50. Mir-Sanchis, I., Pigli, Y. Z. & Rice, P. A. Crystal Structure of an Unusual Single-Stranded DNA-Binding Protein Encoded by Staphylococcal Cassette Chromosome Elements. *Structure* **26**, 1144-1150.e3 (2018).

51. Wang, H.-C. *et al.* Staphylococcus aureus protein SAUGI acts as a uracil-DNA glycosylase inhibitor. *Nucleic Acids Res.* **42**, 1354–1364 (2014).

52. Bebel, A., Walsh, M. A., Mir-Sanchis, I. & Rice, P. A. A novel DNA primase-helicase pair encoded by SCCmec elements. *eLife* **9**, e55478 (2020).

53. Ubeda, C., Tormo-Más, M. Á., Penadés, J. R. & Novick, R. P. Structure-function analysis of the SaPIbov1 replication origin in Staphylococcus aureus. *Plasmid* **67**, 183–190 (2012).

54. Ubeda, C., Barry, P., Penadés, J. R. & Novick, R. P. A pathogenicity island replicon in Staphylococcus aureus replicates as an unstable plasmid. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14182–14188 (2007).

55. Úbeda, C., Barry, P., Penadés, J. R. & Novick, R. P. A pathogenicity island replicon in Staphylococcus aureus replicates as an unstable plasmid. *Proc. Natl. Acad. Sci.* **104**, 14182–14188 (2007).

56. Misiura, A. *et al.* Roles of two large serine recombinases in mobilizing the methicillin-resistance cassette SCCmec. *Mol. Microbiol.* **88**, 1218–1229 (2013).

57. Wang, L. & Archer, G. L. Roles of CcrA and CcrB in excision and integration of staphylococcal cassette chromosome mec, a Staphylococcus aureus genomic island. *J. Bacteriol.* **192**, 3204–3212 (2010).

58. Noto, M. J. & Archer, G. L. A subset of Staphylococcus aureus strains harboring staphylococcal cassette chromosome mec (SCCmec) type IV is deficient in CcrAB-mediated SCCmec excision. *Antimicrob. Agents Chemother.* **50**, 2782–2788 (2006).

59. Smith, M. C. M. Phage-encoded Serine Integrases and Other Large Serine Recombinases. *Microbiol. Spectr.* **3**, 10.1128/microbiolspec.mdna3-0059–2014 (2015).

60. Schormann, N., Ricciardi, R. & Chattopadhyay, D. Uracil-DNA glycosylases—Structural and functional perspectives on an essential family of DNA repair enzymes. *Protein Sci. Publ. Protein Soc.* **23**, 1667–1685 (2014).

61. Stojanov, M., Sakwinska, O. & Moreillon, P. Expression of SCCmec cassette chromosome recombinases in methicillin-resistant Staphylococcus aureus and Staphylococcus epidermidis. *J. Antimicrob. Chemother.* **68**, 749–757 (2013).

62. Liu, P., Wu, Z., Xue, H. & Zhao, X. Antibiotics trigger initiation of SCCmec transfer by inducing SOS responses. *Nucleic Acids Res.* **45**, 3944–3952 (2017).

63. Charpentier, E. *et al.* Novel cassette-based shuttle vector system for gram-positive bacteria. *Appl. Environ. Microbiol.* **70**, 6076–6085 (2004).

64. Arnaud, M., Chastanet, A. & Débarbouillé, M. New Vector for Efficient Allelic Replacement in Naturally Nontransformable, Low-GC-Content, Gram-Positive Bacteria. *Appl. Environ. Microbiol.* **70**, 6887–6891 (2004).

65. Osmundson, J., Dewell, S. & Darst, S. A. RNA-Seq reveals differential gene expression in Staphylococcus aureus with single-nucleotide resolution. *PloS One* **8**, e76572 (2013).

66. Pajunen, M. I., Pulliainen, A. T., Finne, J. & Savilahti, H. Generation of transposon insertion mutant libraries for Gram-positive bacteria by electroporation of phage Mu DNA transposition complexes. *Microbiol. Read. Engl.* **151**, 1209–1218 (2005).

67. Rasila, T. S. *et al.* Mu transpososome activity-profiling yields hyperactive MuA variants for highly efficient genetic and genome engineering. *Nucleic Acids Res.* **46**, 4649–4661 (2018).

68. Ibarra-Chávez, R., Brady, A., Chen, J., Penadés, J. R. & Haag, A. F. Phage-inducible chromosomal islands promote genetic variability by blocking phage reproduction and protecting transductants from phage lysis. *PLoS Genet.* **18**, e1010146 (2022).

69. Lilly, J. & Camps, M. Mechanisms of Theta Plasmid Replication. *Microbiol. Spectr.* **3**, (2015).

70. Kwong, S. M., Ramsay, J. P., Jensen, S. O. & Firth, N. Replication of Staphylococcal Resistance Plasmids. *Front. Microbiol.* **8**, 2279 (2017).

71. Narajczyk, M., Barańska, S., Wegrzyn, A. & Wegrzyn, G. Switch from theta to sigma replication of bacteriophage lambda DNA: factors involved in the process and a model for its regulation. *Mol. Genet. Genomics MGG* **278**, 65–74 (2007).

72. Szabó, J. E. *et al.* Highly potent dUTPase inhibition by a bacterial repressor protein reveals a novel mechanism for gene expression control. *Nucleic Acids Res.* **42**, 11912–11920 (2014).

73. Maiques, E. *et al.* Role of Staphylococcal Phage and SaPI Integrase in Intra- and Interspecies SaPI Transfer. *J. Bacteriol.* **189**, 5608–5616 (2007).

74. Li, S. *et al.* Novel Types of Staphylococcal Cassette Chromosome mec Elements Identified in Clonal Complex 398 Methicillin-Resistant Staphylococcus aureus Strains ▿ . *Antimicrob. Agents Chemother.* **55**, 3046–3050 (2011).

75. Bae, T. *et al.* Staphylococcus aureus virulence genes identified by bursa aurealis mutagenesis and nematode killing. *Proc. Natl. Acad. Sci.* **101**, 12312–12317 (2004).

76. Serrano-Heras, G., Bravo, A. & Salas, M. Phage phi29 protein p56 prevents viral DNA replication impairment caused by uracil excision activity of uracil-DNA glycosylase. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 19044–19049 (2008).

77. Shearer, J. E. S. *et al.* Major families of multiresistant plasmids from geographically and epidemiologically diverse staphylococci. *G3 Bethesda Md* **1**, 581–591 (2011).

78. Rice, P. A. Serine Resolvases. *Microbiol. Spectr.* **3**, MDNA3-0045–2014 (2015).

79. De Oliveira, D. M. P. *et al.* Antimicrobial Resistance in ESKAPE Pathogens. *Clin. Microbiol. Rev.* **33**, e00181-19 (2020).

80. David, M. Z. & Daum, R. S. Community-associated methicillin-resistant Staphylococcus aureus: epidemiology and clinical consequences of an emerging epidemic. *Clin. Microbiol. Rev.* **23**, 616–687 (2010).

81. Maree, M. *et al.* Natural transformation allows transfer of SCCmec-mediated methicillin resistance in Staphylococcus aureus biofilms. *Nat. Commun.* **13**, 2477 (2022).

82. Ito, T. *et al.* Guidelines for Reporting Novel mecA Gene Homologues. *Antimicrob. Agents Chemother.* **56**, 4997–4999 (2012).

83. Shore, A. C. *et al.* Characterization of a Novel Arginine Catabolic Mobile Element (ACME) and Staphylococcal Chromosomal Cassette mec Composite Island with Significant Homology to Staphylococcus epidermidis ACME Type II in Methicillin-Resistant Staphylococcus aureus Genotype ST22-MRSA-IV ▿ . *Antimicrob. Agents Chemother.* **55**, 1896–1905 (2011).

84. Ellington, M. J., Yearwood, L., Ganner, M., East, C. & Kearns, A. M. Distribution of the ACME-arcA gene among methicillin-resistant Staphylococcus aureus from England and Wales. *J. Antimicrob. Chemother.* **61**, 73–77 (2008).

85. Wang, H.-C. *et al.* Staphylococcus aureus protein SAUGI acts as a uracil-DNA glycosylase inhibitor. *Nucleic Acids Res.* **42**, 1354–1364 (2014).

86. Schneewind, O. & Missiakas, D. Genetic manipulation of Staphylococcus aureus. *Curr. Protoc. Microbiol.* **32**, Unit-9C.3. (2014).

87. An oxidation-sensing mechanism is used by the global regulator MgrA in Staphylococcus aureus | Nature Chemical Biology. https://www.nature.com/articles/nchembio820.

88. Willing, S., Dyer, E., Schneewind, O. & Missiakas, D. FmhA and FmhC of Staphylococcus aureus incorporate serine residues into peptidoglycan cross-bridges. *J. Biol. Chem.* **295**, 13664–13676 (2020).

89. Novick, R. P. *et al.* Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J.* **12**, 3967–3975 (1993).

90. Fey, P. D. *et al.* A genetic resource for rapid and comprehensive phenotype screening of nonessential Staphylococcus aureus genes. *mBio* **4**, e00537-00512 (2013).

91. Chu, X., Xia, R., He, N. & Fang, Y. Role of Rot in bacterial autolysis regulation of Staphylococcus aureus NCTC8325. *Res. Microbiol.* **164**, 695–700 (2013).

92. Peng, J. & Gygi, S. Proteomics: the move to mixtures - PubMed.

93. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).

94. Highly accurate protein structure prediction with AlphaFold | Nature. https://www.nature.com/articles/s41586-021-03819-2.

95. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

96. Källberg, M., Margaryan, G., Wang, S., Ma, J. & Xu, J. RaptorX server: a resource for template-based protein structure modeling. *Methods Mol. Biol. Clifton NJ* **1137**, 17–27 (2014).

97. Holm, L., Laiho, A., Törönen, P. & Salgado, M. DALI shines a light on remote homologs: One hundred discoveries. *Protein Sci. Publ. Protein Soc.* **32**, e4519 (2023).

98. Chembazhi, U. V. *et al.* Uracil DNA glycosylase (UDG) activities in Bradyrhizobium diazoefficiens: characterization of a new class of UDG with broad substrate specificity. *Nucleic Acids Res.* **45**, 5863–5876 (2017).

99. Haushalter, K. A., Stukenberg, P. T., Kirschner, M. W. & Verdine, G. L. Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Curr. Biol.* **9**, 174–185 (1999).

100. Alexeeva, M. *et al.* Excision of uracil from DNA by hSMUG1 includes strand incision and processing. *Nucleic Acids Res.* **47**, 779–793 (2019).

101. Papp-Kádár, V., Balázs, Z., Vékey, K., Ozohanics, O. & Vértessy, B. G. Mass spectrometry-based analysis of macromolecular complexes of Staphylococcus aureus uracil-DNA glycosylase and its inhibitor reveals specific variations due to naturally occurring mutations. *FEBS Open Bio* **9**, 420–427 (2019).

102. Higgins, P. G., Rosato, A. E., Seifert, H., Archer, G. L. & Wisplinghoff, H. Differential expression of ccrA in methicillin-resistant Staphylococcus aureus strains carrying staphylococcal cassette chromosome mec type II and IVa elements. *Antimicrob. Agents Chemother.* **53**, 4556–4558 (2009).

103. Zhang, S., Shu, X. & Sun, B. SigmaB regulates ccrAB expression and SCCmec excision in methicillin-resistant Staphylococcus aureus. *Int. J. Med. Microbiol. IJMM* **306**, 406–414 (2016).

104. Saïd-Salim, B. *et al.* Global Regulation of Staphylococcus aureus Genes by Rot. *J. Bacteriol.* **185**, 610–619 (2003).

105. Jiang, Q., Jin, Z. & Sun, B. MgrA Negatively Regulates Biofilm Formation and Detachment by Repressing the Expression of psm Operons in Staphylococcus aureus. *Appl. Environ. Microbiol.* **84**, e01008-18 (2018).

106. Crosby, H. A. *et al.* The Staphylococcus aureus Global Regulator MgrA Modulates Clumping and Virulence by Controlling Surface Protein Expression. *PLoS Pathog.* **12**, e1005604 (2016).

107. Frankel, M. B., Hendrickx, A. P. A., Missiakas, D. M. & Schneewind, O. LytN, a Murein Hydrolase in the Cross-wall Compartment of Staphylococcus aureus, Is Involved in Proper Bacterial Growth and Envelope Assembly. *J. Biol. Chem.* **286**, 32593–32605 (2011).

108. Hao, Z. *et al.* Deletion of SarX Decreases Biofilm Formation of Staphylococcus aureus in a Polysaccharide Intercellular Adhesin (PIA)-Dependent Manner by Downregulating spa. *Infect. Drug Resist.* **14**, 2241–2250 (2021).

109. Manna, A. C. & Cheung, A. L. Expression of SarX, a negative regulator of agr and exoprotein synthesis, is activated by MgrA in Staphylococcus aureus. *J. Bacteriol.* **188**, 4288–4299 (2006).

110. Rowe, S. E., Mahon, V., Smith, S. G. & O'Gara, J. P. A novel role for SarX in Staphylococcus epidermidis biofilm regulation. *Microbiol. Read. Engl.* **157**, 1042–1049 (2011).

111. Cue, D., Lei, M. G. & Lee, C. Y. Activation of sarX by Rbf Is Required for Biofilm Formation and icaADBC Expression in Staphylococcus aureus. *J. Bacteriol.* **195**, 1515–1524 (2013).

112. Kaiser, T. D. L. *et al.* Modification of the Congo red agar method to detect biofilm production by Staphylococcus epidermidis. *Diagn. Microbiol. Infect. Dis.* **75**, 235–239 (2013).

113. Freeman, D. J., Falkiner, F. R. & Keane, C. T. New method for detecting slime production by coagulase negative staphylococci. *J. Clin. Pathol.* **42**, 872–874 (1989).

114. Suzuki, T. *et al.* Wall teichoic acid protects Staphylococcus aureus from inhibition by Congo red and other dyes. *J. Antimicrob. Chemother.* **67**, 2143–2151 (2012).

115. Vickery, C. R., Wood, B. M., Morris, H. G., Losick, R. & Walker, S. Reconstitution of Staphylococcus aureus Lipoteichoic Acid Synthase Activity Identifies Congo Red as a Selective Inhibitor. *J. Am. Chem. Soc.* **140**, 876–879 (2018).

116. Houston, P., Rowe, S. E., Pozzi, C., Waters, E. M. & O'Gara, J. P. Essential Role for the Major Autolysin in the Fibronectin-Binding Protein-Mediated Staphylococcus aureus Biofilm Phenotype. *Infect. Immun.* **79**, 1153–1165 (2011).

117. Kobayashi, S. D. *et al.* Comparative Analysis of USA300 Virulence Determinants in a Rabbit Model of Skin and Soft Tissue Infection. *J. Infect. Dis.* **204**, 937–941 (2011).

118. Diep, B. A. *et al.* Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant Staphylococcus aureus. *Lancet Lond. Engl.* **367**, 731–739 (2006).

119. Structure and Specificity of the Vertebrate Anti-Mutator Uracil-DNA Glycosylase SMUG1 - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S1097276503002351#aep-section-id22.

120. Hossain, M., Aslan, B. & Hatoum-Aslan, A. Tandem mobilization of anti-phage defenses alongside SCCmec cassettes. *bioRxiv* 2023.03.17.533233 (2023) doi:10.1101/2023.03.17.533233.

121. Wang, Z. & Mosbaugh, D. W. Uracil-DNA glycosylase inhibitor of bacteriophage PBS2: cloning and effects of expression of the inhibitor gene in Escherichia coli. *J. Bacteriol.* **170**, 1082–1091 (1988).

122. Z, W. & Dw, M. Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. *J. Biol. Chem.* **264**, (1989).

123. Makhlin, J. *et al.* Staphylococcus aureus ArcR Controls Expression of the Arginine Deiminase Operon. *J. Bacteriol.* **189**, 5976–5986 (2007).

124. Diep, B. A. *et al.* The Arginine Catabolic Mobile Element and Staphylococcal Chromosomal Cassette mec Linkage: Convergence of Virulence and Resistance in the

USA300 Clone of Methicillin-Resistant Staphylococcus aureus. *J. Infect. Dis.* **197**, 1523–1530 (2008).

125. Archer, N. K. *et al.* Staphylococcus aureus biofilms. *Virulence* **2**, 445–459 (2011).

126. Campoccia, D., Montanaro, L. & Arciola, C. R. Extracellular DNA (eDNA). A Major Ubiquitous Element of the Bacterial Biofilm Architecture. *Int. J. Mol. Sci.* **22**, 9100 (2021).

127. Chiba, A. *et al.* Staphylococcus aureus utilizes environmental RNA as a building material in specific polysaccharide-dependent biofilms. *Npj Biofilms Microbiomes* **8**, 1–10 (2022).

128. Nguyen, H. T. T., Nguyen, T. H. & Otto, M. The staphylococcal exopolysaccharide PIA – Biosynthesis and role in biofilm formation, colonization, and infection. *Comput. Struct. Biotechnol. J.* **18**, 3324–3334 (2020).

129. Speziale, P., Pietrocola, G., Foster, T. J. & Geoghegan, J. A. Protein-based biofilm matrices in Staphylococci. *Front. Cell. Infect. Microbiol.* **4**, 171 (2014).

130. Corrigan, R. M., Rigby, D., Handley, P. & Foster, T. J. The role of Staphylococcus aureus surface protein SasG in adherence and biofilm formation. *Microbiol. Read. Engl.* **153**, 2435–2446 (2007).

131. Sanchez, C. J. *et al.* The pneumococcal serine-rich repeat protein is an intra-species bacterial adhesin that promotes bacterial aggregation in vivo and in biofilms. *PLoS Pathog.* **6**, e1001044 (2010).

132. Lei, M. G., Gupta, R. Kr. & Lee, C. Y. Proteomics of Staphylococcus aureus biofilm matrix in a rat model of orthopedic implant-associated infection. *PLoS ONE* **12**, e0187981 (2017).

133. Kattke, M. D. *et al.* Structure and mechanism of TagA, a novel membrane-associated glycosyltransferase that produces wall teichoic acids in pathogenic bacteria. *PLOS Pathog.* **15**, e1007723 (2019).

134. Ubeda, C., Tormo-Más, M. Á., Penadés, J. R. & Novick, R. P. Structure-function analysis of the SaPIbov1 replication origin in Staphylococcus aureus. *Plasmid* **67**, 183–190 (2012).