Divergent Immunity Proteins Protect Against a Type VI Secretion System Effector Family Found in the Human Gut Microbiome

Divergent Immunity Proteins Protect Against a Type VI Secretion System Effector Family Found in the Human Gut Microbiome

By Amirahmad Azhieh

A thesis submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Master of Science

McMaster University MASTER OF SCIENCE (2022) Hamilton, Ontario (Biochemistry and Biomedical Sciences)

Title: Divergent immunity proteins protect against a type VI secretion system effector family found in the human gut microbiome

Author: Amirahmad Azhieh (McMaster University)

Supervisor: Dr. John Whitney

Number of Pages: XI, 66

ABSTRACT

Antagonistic interactions between competing species of bacteria are an important driver of bacterial community composition in the human gut microbiota. Of particular significance is the role of the type six secretion system (T6SS), which many species of Gram-negative bacteria use to kill competitor bacteria in a contact-dependent manner. T6SSs are syringe-like nanomachines that function to deliver antibacterial toxins into susceptible competitors. Many bacteria present in the human gut microbiota possess an extremely potent T6SS that is capable of rapidly eradicating nearby bacteria. Remarkably, however, species of beneficial bacteria that coexist in the gut are often resistant to T6SS attack by their neighbours. This resistance is mediated by bacterial immunity proteins that block the activity of the antibacterial toxins delivered by the T6SS. Intriguingly, past studies have shown that the widespread T6SS-mediated competition in the gut has led to the acquisition of repertoires of immunity genes across different bacterial strains.

By examining available human gut metagenomes, I identified a putative immunity locus, named I2, in a species of gut bacteria. This locus is located downstream of its cognate T6SS toxin-encoding locus, E2, and I show when co-expressed with E2 in *E. coli*, it protects against E2 mediated-toxicity. Additionally, I show that four gut-derived I2 homologues bearing sequence identity levels to I2 ranging from 38% to 75% are equally capable of abrogating E2 toxicity. Using quantitative biophysical measurements, I also show that these I2 homologues physically bind E2 equally tightly pointing to the potential molecular mechanism of toxin neutralization. Lastly, through mutagenesis experiments, I found that the E2-I2 interaction is likely mediated by electrostatic forces between a small number of

residues found in the interaction interface of the two proteins. Overall, these findings demonstrate that a human gut microbiome encoded type VI secretion system effector can be neutralized by divergent immunity proteins.

ACKNOWLEDGMENTS

I would like to thank Dr. John Whitney for the opportunity he provided me with to work as a graduate student researcher in his lab for the past two years. I have been able to learn a lot more about research in basic science and beyond thanks to the detail-oriented approach in his lab and I have benefited a lot from the thought-provoking discussions we have had in our one-on-one meetings. I have also learned an extreme of insights on the importance of effective ways of communication in science in writing but most importantly in oral presentations, something Dr. Whitney emphasized many times over the course of my time in his lab. Additionally, I would like to thank Dr. Benjamin Ross for his help with the many different aspects of the project as well as being available, literally, round the clock to answer any questions I ever had about any aspect of it. I am also thankful to my committee members, Dr. Brian Coombes and Dr. Alex Hynes, for their guidance and help both inside and outside of the allocated committee meeting times.

This project would not have been successful in coming to a closure had it not been for the help of many people in the Whitney lab and the IIDR institute. In particular, I am indebted to Dr. David Sychantha for his patient approach to help me on the immensely difficult task of getting a clean ITC data, something I never thought would be achievable in such a short period of time especially during a very stressful part of the project.

I would also like to thank (soon-to-be Dr.!) Timothy Klein for his help with the structural aspect of the project, whether showing me how to set up a crystal tray for the first time or helping me go through the many shapes I would see thinking they would be protein crystals as well as always being more than willing to use the X-ray machine to confirm the presence

of protein crystals; something which sadly never came to fruition. I am also thankful to Dr. Shehryar Ahmad, my benchmate for the first year who was more than happy to help me with any questions especially at the beginning of my project when I was stressed about every subtle detail of every protocol that I was using for the first time. His presence was an asset in particular late at nights where usually few people would be to help me with my questions or simply to give me useful advice on science and beyond. I am grateful to Nathan Bullen not only for helping me get better with Adobe Illustrator (I still have a long way to go!) but also for his many insights on the details of the project during our lab meetings and beyond and for recommending important papers and studies which were essential to many aspects of hypothesis generation in my project particularly towards the end of it where I had almost lost hope in my pursuit of the mutagenesis experiments. I am also thankful to Dr. Lindsey Marmont for her help with the different aspects of protein purification and many helpful insights she shared with me. I am grateful to other members of the Whitney lab during my time there; Dirk Grabenc, Andrea Alexei, and the many undergraduates in the lab who always helped me with science and beyond. Thank you, members of the Andres lab, Monica, for helping me with the SEC-MALS experiment and Dana and Andriana, for always keeping me company during the ups and downs of my project. Finally, thank you Kartik Sacher, for always being there for me with any help be it inside or outside the lab during the past one year and a half. Having a close friend in a new city played an important role in my overall well-being and my ability to manage my time much better.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AF2: AlphaFold2

BSAP: Bacteroidales-secreted antimicrobial proteins

B. fragilis: *Bacteroides fragilis*

BME: β-mercaptoethanol

CDI: contact-dependent inhibition

CTD: C-terminal domain

DTT: dithiothreitol

E. coli: *Escherichia coli*

E-I: Effector-Immunity

GA: genetic architecture

HGT: horizontal gene transfer

ITC: isothermal titration calorimetry

MALS: multiangle light scattering

ORF: open reading frame

rAID: recombinase associated acquired interbacterial defense

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC: size-exclusion chromatography

SUKH: Syd, US22, Knr4 homology

TA: toxin-antitoxin

T6SS: Type 6 Secretion System

DECLARATION OF ACADEMIC ACHIEVEMENT

Study design, experiments, data analysis, and thesis writing were completed by Amirahmad Azhieh under the supervision of Dr. John Whitney. The bioinformatics analysis including search for the effector immunity loci shown in Figure 1 was carried out by Dr. Benjamin Ross. The SEC-MALS experiment was carried out by Monica Warner.

INTRODUCTION

Bacterial Antagonism

Bacterial antagonism is an important type of bacterial interaction observed in diverse microbial communities. In general, bacteria use two strategies to antagonize one other: 1) by producing antimicrobial molecules released into the extracellular media or 2) by secreting molecules directly into bacterial cells that they encounter using contact-dependent delivery mechanisms[1] (Garcia-Bayona et al. 2018). Examples of molecules secreted through the former strategy include small-molecule antibiotics as well as ribosomally-synthesized peptide toxins such as bacteriocins. Examples of the second strategy include the variety of toxins delivered into target cells by different secretion systems of bacteria, including the type IV, V, VI, and VII secretion systems (Souza et al 2015; Aoki et al. 2005; Hood et al. 2010; Cao et al. 2016).

Recent work has led to the characterization of many such antagonistic interactions in dense microbial environments such as the lower gut microbiome (Walter & Ley 2011). In the gut microbiome, researchers have found examples of both of the abovementioned strategies. For example, it has been demonstrated that a species of the *Blautia* genus found in the

[1] These two strategies fall under what may be regarded as "direct interbacterial antagonism" where "direct" refers to the production of a molecule that antagonizes the target bacterium by directly intoxicating it. There are also examples of indirect strategies utilized by bacteria, for example, when it comes to competition for nutrients such as production of siderophores to acquire environmental iron which can help the siderophore-producing bacteria indirectly antagonize those with no siderophores or low-iron-affinity siderophore producers (Hibbing et al. 2010).

human gut produces an antibiotic that disrupts the ability of pathogenic vancomycin-resistant *Enterococcus faecium* to establish in the gut (Kim et al. 2019). Additionally, some member species of the Bacteroidetes, the most abundant phylum in the gut microbiome (Huttenhower et al. 2012; King et al 2019), encode different Bacteroidales-secreted antimicrobial proteins (BSAP) that kill susceptible bacteria upon secretion into the extracellular milieu resulting in the susceptible bacteria being outcompeted by the BSAP-producing bacteria in the mammalian gut (Roelofs et al. 2016). Lastly, a recent genomics-based study identified a large number of genes responsible for production of many secondary metabolites with putative antimicrobial activities found in the genomes of the members of the most abundant bacterial phyla living in soil ecosystems, which is another example of a dense microbial environment of importance to human activities (Crits-Christoph et al. 2018).

When it comes to contact-dependent bacterial antagonism, the best-characterized and most prevalent pathway among Gram-negative bacteria is the type VI secretion system (T6SS), which is the central focus of my thesis and is discussed in further detail below.

Type 6 Secretion Systems

Type 6 secretion systems (T6SSs) of Gram-negative bacteria are protein-injection apparatuses first identified in the context of host-pathogen interactions (Mougous et al. 2006, Pukatzki et al. 2006). They have since come to be recognized as a system that primarily targets bacteria, which also has important implications for infection biology but not in the direct manner that was originally proposed (Hood et al. 2010). T6SS-dependent bacterial antagonism occurs in diverse environments, from isolated microbial communities

found in soil and oceanic ecosystems to the host-associated microbiota of different organisms (Peterson et al. 2020). The apparatus itself is made up of 14 structural genes encoding proteins that are assembled upon the cell coming into contact with a neighboring cell. Together, these 14 proteins make up the two major components of the system, known as the membrane complex and the phage tail-like complex. These two complexes form a tube-like conduit directing the export of toxins encoded by the genome of the T6SS-containing bacterium into the target cell where the toxins, known as effectors, perform their physiological function (Figure 1). Since the effectors are by definition toxic, the bacteria producing them express what are known as immunity proteins, which are usually encoded by a locus downstream of the effector locus to neutralize them prior to secretion in order to protect the toxin-producing cell from its own toxin or the toxins delivered intercellularly by its sister cells. The fact that these immunity proteins also protect T6SS-containing cells from kin cells producing and secreting the same effector families, allows clonal populations of bacteria to co-exist with one another in conditions that promote cell-to-cell contact (Klein et al. 2020). It is therefore plausible to consider that T6SS likely plays an important role in the human gut microbiota where cell-to-cell contact would be frequent.

Bacterial Antagonism in the Gut Microbiota

The human gut microbiome consists of a wide variety of bacteria that are found predominantly in the lower gastrointestinal tract (Huttenhower et al. 2012). It is primarily made up of four phyla of bacteria that are found across almost all gut microbiomes with Bacteroidetes representing the plurality of the Gram-negative bacteria and Firmicutes being the most abundant phylum of Gram-positive bacteria. A smaller proportion of the resident

bacteria of the human gut consist of Actinobacteria and Proteobacteria (Huttenhower et al. 2012, King et al 2019).

Initial work characterizing T6SSs in species of Bacteroidetes led to the identification of a family of T6SSs that are distinct from those of the well-characterized Proteobacterial systems that had been discovered previously, in that they lack a number of the 14 required T6SS apparatus genes mentioned above (Russell et al. 2014). Through in vitro competition assays, the same study demonstrated the functionality of the T6SS in *Flavobacterium johnsoniae*, a soil-dwelling Bacteroidetes, and in *Bacteroides fragilis*, a Bacteroidetes species found in the gut microbiome. Subsequent work has lent further support to the functionality of the *B. fragilis* T6SS system in the context of antagonizing susceptible bacteria both in vitro and, more importantly, inside the mammalian gut (Chatzidaki-Livanis et al. 2016, Wexler et al. 2016). Quantitative measurements of the activity of this T6SS revealed a very potent and active system in the dense environment of the microbiome predicted to consist of 60-600 billion effector transmission events between neighboring cells in the human gut (Wexler et al. 2016). There are also examples of at least one species of gut-colonizing pathogenic bacteria, namely *Salmonella enterica* Serovar Typhimurium, utilizing their T6SS to mediate bacterial antagonism, which the researchers demonstrated was necessary for its establishment in the gut (Sana et al. 2016). These examples all point to the importance of T6SS-mediated interbacterial antagonism as an important force shaping the bacterial composition of the gut microbiome, facilitating invasion into it by bacterial pathogens, and contributing to colonization resistance by human gut commensals as has been shown by Hecht and colleagues (Hecht et al. 2016).

To better understand the extent of T6SS-mediated antagonism in the human gut microbiome, it is imperative to identify the bacteria that interact with each other using their T6SS, especially since it has been shown that two bacteria that display antagonism in vitro may not necessarily be interacting in the mammalian gut (Wexler et al. 2016). Additionally, as I will describe later, the recent discovery of orphan T6SS immunity genes, which are clusters of immunity genes not associated with any effector genes of their own (Russell et al 2012), has raised many more questions on the nature of these interactions. For example, bacteria lacking a T6SS may evolve resistance to T6SS attack by accumulating large collections of immunity genes.

Genetic Architectures of T6SS Loci and Effector-Immunity Genotypes

The discovery of the first T6SS loci in Bacteroidetes genomes led to several studies to understand more about the distribution of such loci across genomes. A survey of the sequenced genomes of Bacteroidales identified three types of genetic architectures (GA) into which the majority of T6SS loci are organized; these T6SS subtypes are named GA1, GA2, and GA3. In addition to the different arrangements of the T6SS loci in each genetic architecture, the DNA sequences in every genetic architecture show a high level of similarity to other members of the same architecture and are different from that of the other two. The exception is the divergent regions that are variable in different genomes of the same architecture (Coyne et al. 2016). It was also found that the T6SS subtype GA1 and GA2 are found in many different species in the order Bacteroidales whereas GA3 is specific to *B. fragilis* species (Coyne et al. 2016). Subsequent studies focused on the distribution of Effector-Immunity (E-I) pairs in the divergent regions of these T6SS subtypes in

Bacteroidales genomes (Verster et al. 2017). In particular, researchers found that the variable regions of these three T6SS subtypes each consists of a small number of distinct E-I pairs indicative of very low diversity of T6SS E-I genes in individual human microbiomes (Verster et al. 2017). Of the 12 E-I pairs in the T6SS subtype GA1, the second E-I pair found in the species *Parabacteroides sp. D25* caught our interest. The E-I pair which consist of the GA1_E2 and GA1_I2 effector-immunity pair as shown in Figure 2A, is of utmost interest to us since in addition to their presence as a putative effector-immunity pair, we were able to find several I2 homologues with wide range of sequence identity to I2 as will be discussed below.

While exploring different Type Six Secretion Systems in the genomes of bacteria found in the gut microbiota, we came across a putative T6SS-associated effector that is annotated to have a domain of the WHH family of HNH nucleases (Figure 2A). Downstream of this effector, named GA1_E2, there is a putative cognate immunity locus, named GA1_I2. These loci are in the genome of the *Parabacteroides sp. D25* species mentioned above (Figure 2A). For brevity, I will henceforth refer to GA1_E2 and GA1_I2 as E2 and I2, respectively.

HNH Superfamily of Nucleases and the WHH Family

The HNH family of nucleases were first discovered in a study aimed at finding the conserved amino acid sequences of group I intron open reading frames (ORF), where in addition to these group I ORFs, several eukaryotic and prokaryotic proteins were found to contain a motif consisting of two histidine residues and an asparagine residue each separated by a number of other non-conserved residues (Shub et al. 1994; Gorbalenya et

al. 1994). Importantly, these motifs were also found to be present in a number of bacterial endonucleases including bacteriocins (Shub et al. 1994; Gorbalenya et al. 1994).

The WHH family of toxins is a sub-family of HNH nucleases. Named after the WHH amino acid residues found in the C-terminal domain of the protein, the family was discovered alongside many other subtypes of HNH nucleases as putative nucleases that are located close to loci that are part of different toxin systems in bacteria (Zhang et al 2011). As of now, very little is known about the WHH family of nucleases. No crystal structure of a member of the family has been solved and unlike a few other families of HNH nucleases listed in Table 1, such as LHH and AHH, not much is known about the mechanism of action of this toxin family.

An interesting aspect of the WHH family of nucleases shared by many other members of the HNH family of nucleases, is the cognate immunity associated with these nucleases. In their major survey of bacterial genomes using gene neighborhood analyses, Zhang et al. discovered that many nucleases of the HNH family are predicted to act as toxins and are linked to genes that encode members of the **SUKH (for Syd, US22, Knr4 homology)** domain superfamily (Zhang et al 2011). They hypothesize that these SUKH domain-containing genes function as immunity proteins associated with the HNH toxins to protect the cells producing these toxins from self-toxicity, similar to the concept of effector-immunity mentioned above. This is an unusual observation in the toxin-immunity field because it has been found that most immunity proteins, even if they neutralize toxins with similar modes of action, possess little to no homology with one another and sometimes even adopt entirely different protein folds (Tang et al, 2018). Thus, the fact that a single

protein domain family, SUKH, is found next to thousands of different predicted nucleases suggests that the molecular mechanism used to inhibit this type of enzymatic activity is conserved.

SUKH Family of Proteins

The Smi1/Knr4 protein was first discovered in the eukaryotic model organism *Saccharomyces cerevisiae*. Smi1 was identified through a genetic screen as a protein whose deletion led to temperature-sensitive cell cycle arrest during the S-phase (Fishel et al. 1993). More insights were gained through a subsequent screen-based method that selected for genes that confer resistance against K9, a toxin produced by *Williopsis saturnus* var. mrakii, a member of a distinct genus of yeasts, whose target is the *S. cerevisiae* cell wall (Hong et al. 1994). The screen identified a gene with 100% identity to Smi1 which the researchers named Knr4, for K-nine-resistance 4, referring to the fourth mutant which showed resistance against the K9 toxin (Hong et al. 1994). Further genetic and phenotypic experiments identified Smi1/Knr4 as a gene that is implicated in the pathway leading to the synthesis of 1,3-β-glucan, an important component of the yeast cell wall (Hong et al. 1994). Subsequent studies found a number of physical interactions of the Smi1/Knr4 protein with proteins whose functions fall under cell wall maintenance and biogenesis, cell polarity and bud emergence, as well as other proteins which may have a role in the stability and/or degradation of proteins, indicating that Knr4 is an important protein with many cellular roles through the variety of its physical interactions with a large number of proteins (Basmaji et al. 2005).

The main takeaway from the yeast studies is that Knr4 is suggested to be an important hub protein coordinating the activity of many other proteins in the cell responsible for such diverse functions as cell wall synthesis and cell cycle through its myriad of physical interactions with those proteins (Martin-Yken et al. 2016). That suggests that in other domains of life, any protein with a domain characteristic of the Smi1/Knr4 protein, may play important roles through its ability to physically interact with a large number of proteins. In light of the observation of these SUKH-domain containing putative immunity genes next to a variety of potential nuclease toxins, it will be important to see whether proteins containing such domains will have the ability to bind a variety of other prokaryotic proteins in the process of toxin inhibition.

The first observation of proteins containing Smi1/Knr4 domains in prokaryotes was reported by Zhang and colleagues in a bioinformatics survey of Smi1/Knr4 homologs in bacteria (Zhang et al. 2011). Through their extensive bioinformatic as well as structural similarity searches aided by a crystal structure of *Bacillus subtilis* protein containing the Smi1/Knr4 domain, they found a number of homologous proteins, namely Syd, US22 and Knr4, in eukaryotes, prokaryotes, and DNA viruses pointing to their widespread occurrence across kingdoms (Zhang et al. 2011). Using this general family of proteins, they named this superfamily of conserved domains SUKH (short for the Syd, US22, Knr4 homology). As mentioned previously, another important finding of their analysis was the association of the proteins with the SUKH domain with the HNH family of nucleases that Zhang et al. postulated act as cognate immunity genes for these toxins (Zhang et al. 2011). Interestingly, we found that I2, the putative cognate immunity of E2, is predicted to consist of two tandem

SUKH domains. We therefore found it to be of interest to study this effector along with its cognate immunity to understand more about the molecular nature of the interaction of these two protein families. However, before I move on with my characterization of these proteins, I first need to introduce the concept of orphan immunity.

Orphan Immunity proteins

As mentioned earlier (in the Type 6 Secretion System section), in the context of antibacterial toxin systems, each toxin-producing bacterium produces an immunity protein that neutralizes the activity of the toxin, usually by binding directly to the active site of the toxin (Hersch et al. 2020). However, many bacteria that do not possess the toxin of a given toxin-immunity system might still encode immunity proteins that can neutralize other toxins directed to them by other bacterial toxin systems. Since they are not genetically linked to any upstream toxins, they are referred to as 'orphan' immunity proteins.

The presence of immunity genes not linked to cognate effectors was first observed in the context of two-partner secretion systems, an example of Contact-Dependent Inhibition (CDI) systems found in Proteobacteria (Zhang et al. 2011). This finding was subsequently followed by the first observation of immunity loci that were not found adjacent to effector genes in the context of the T6SSs of several Gram-negative Proteobacteria (Russell et al. 2012). Several years later, the first experimental demonstration that orphan immunity genes in different *B. fragilis* strains can neutralize the effectors produced by other antagonizing *B. fragilis* strains harbouring the T6SS weaponry was published (Wexler et al. 2016). This finding has been further supported by the recent discovery of large clusters of putative interbacterial defense genes, including T6SS orphan immunity genes, across many species

of Bacteroidetes found in the gut (Ross et al 2019). These gene clusters, which have now been named recombinase associated Acquired Interbacterial Defense (rAID) systems, consist of a recombinase gene followed by a number of orphan immunity genes homologous to T6SS immunity genes as well as immunity genes from other antibacterial toxin delivery systems (Figure 2B). The presence of large numbers of these gene clusters in bacterial genomes derived from the gut microbiota has raised a number of questions, chief among them being whether or not the orphan immunity genes present in these clusters are capable of conferring to the species harbouring them the ability to produce proteins that would be used to defend against the T6SS-mediated attack of other neighboring bacteria through the neutralization of incoming antibacterial toxins.

Homologues of I2

As mentioned before, E2 and I2 are a putative effector-immunity pair found in the genome of the microbiome species *Parabacteroides sp. D25*. We observed early on that I2 has a number of homologous genes in several putative rAID clusters found in human gut bacteria (Figure 2C). One of these rAID clusters, which contains two homologues of I2, is found in a *B. fragilis* strain called BIOML-A28 that was isolated from a healthy human gut microbiome (Figure 2D, lower panel) (Zhao et al. 2019). Interestingly, the same microbiome sample contains a strain of *Bacteroides uniformis* (*B. uniformis*) named BIOML-A5 harbours a bicistron encoding E2 and I2 (Figure 2D upper panel) suggesting that the orphan immunity homologues of I2 in BIOML-A28 might confer resistance against T6SS-mediated attack from *B. uniformis*. We named these two orphan immunity homologues Orf3 and Orf5 based on the position of these genes within the rAID cluster.

Using a BLAST search, we found additional homologues of I2 with closer identity levels to *Parabacteroides sp. D25* I2 than the orphan I2s. Two of these homologues, possessing 67% and 73% identity to I2 respectively, are from strains of *Phocaeicola vulgatus* and *Bacteroides uniformis* species and are both part of a likely T6SS effector-immunity bicistron (Figure 2E). To reflect the strains these immunity genes were isolated from, we decided to name them I2_PV and I2_BU to designate the *Phocaeicola vulgatus* and *Bacteroides uniformis* homologues, respectively.

Project Goals

My MSc thesis sought to address the following questions:

1. Does E2 encode an antibacterial toxin and can its toxicity be protected against by I2, lending support to their predicted roles as a T6SS effector-immunity pair?
2. Can any of the identified I2 homologues proteins found in other species of gut bacteria protect against E2-mediated toxicity to a similar extent as its cognate I2?
3. What is the molecular mechanism of neutralization of E2 by I2?

MATERIALS AND METHODS

Cloning and Plasmid Construction

All cloning reactions were performed using restriction cloning. The primers used for individual cloning reactions as well as the restriction enzymes used are all listed in Table S1. For the toxicity assays, two types of plasmids were used. If used for the toxicity assays, all toxin genes were cloned in pSCrhaB2 plasmids (Cardona & Valvano 2005). All immunity genes were cloned in pPSV39-CV plasmids (Rietsch et al., 2005). For all cloning reactions carried out for protein expression for setting up crystal screens, the plasmid used was pETDuet where the effector was cloned into multiple cloning site #1 and the immunity into multiple cloning site #2. To purify the effector protein for ITC experiments, the plasmid used was also pETDuet. However, for every single purification of the immunity proteins to be used in ITC, the pET29b plasmid was used.

The cloned plasmids were transformed into two types of chemically competent cells. For the toxicity assays, the transformation was done in XL1-blue cells. For the protein purification analyses, both for setting up crystal trays as well as to use the protein for ITC, the transformation was done in BL21 pLysS cells. Site-specific mutants used in this study were generated by overlap extension PCR. All plasmids were sequenced by Genewiz Incorporated.

Toxicity Assays

All toxicity assays were performed by serial dilution of the corresponding strains on two types of plates. Using a 96-well plate, 200 µL of overnight cultures of individual strains were added to the wells in the first row and were serially diluted 1/10 in each subsequent row. To assay for toxicity, the different dilutions of the strains were spotted onto an LB-agar plate which contained 200 µg/mL of Trimethoprim (selecting for the pSCrhaB2 plasmid), 15 µg/mL of Gentamicin (selecting for the pPSV39-CV plasmid), and 0.1% (w/v) of rhamnose to induce the expression of the toxin. To assay for the ability of the immunity proteins to protect against the toxin, the different dilutions of the strains were spotted onto an LB-agar plate containing 200 µg/mL of Trimethoprim, 15 µg/mL of Gentamicin, 0.1% of rhamnose, in addition to 0.5 µM of IPTG to induce the expression of the immunity proteins.

Protein Expression and purification

For the expression and copurification of effector-immunity complexes, the nucleotide sequence of the genes of interest were cloned into pETDuet-1 vectors. All Bacteroidetes loci used were codon optimized. Cells were grown overnight and diluted in a 1/50 ratio into 1 L Fernbach flasks and allowed to grow to approximately an OD600 of 0.6. They were then induced by 1 mM of IPTG and expressed for at least three hours before being centrifuged to collect the pellets. For the nickel-column based purification, cells were resuspended in lysis buffer containing 10 mM imidazole, 150 mM NaCl, and 50 mM Tris-HCl at pH 8.0, lysed, and the tagged proteins were eluted off the column using a solution containing 150 mM NaCl, 50 mM Tris-HCl at pH 8.0, and 400 mM imidazole. Protein

samples were further purified by size exclusion chromatography using a HiLoad 16/600 Superdex 200 column connected to an AKTA protein purification system (GE Healthcare) and concentrated to their respective concentrations.

For the expression of individual immunity proteins, the nucleotide sequence of the genes of interest were cloned into pET-29 vectors. The rest of the expression and purification steps were conducted in a similar manner as the effector-immunity complex. The only difference was after induction, the cells were allowed to grow overnight at 18 °C as I discovered it led to higher level expression of individual proteins. Additionally, reducing agents were needed to abrogate any potential disulfide bond formation. For that, 1 mM of dithiothreitol (DTT) was added to the lysis and elution buffers and 1 mM β-mercaptoethanol (BME) was used on the HPLC. DTT was used on the nickel column due to the incompatibility of using BME on the resin.

For the purification of individual effector protein $E2_{medCTD}$, I purified the $E2_{medCTD}$-I2 complex but before eluting the HIS-tagged protein complex I denatured the immunity protein by washing the column with a lysis buffer that contained 6 M guanidinium chloride. After that, the column was washed three times to remove any residual denaturing reagent. The remaining solutions on the column were then eluted off the column as before. Another important difference was that each solution was prepared at pH 7.0 due to the effector protein $E2_{medCTD}$ being most stable at that pH.

For ITC, $E2_{medCTD}$ and the different immunity proteins were further purified using SEC at pH 7.0 so both proteins are in the same buffer condition to prevent the ITC instrument from registering any heat release due to the differences in the chemical composition of the

solution containing the immunity protein compared to the solution containing the effector protein.

SEC-MALS

Size exclusion chromatography with multi-angle laser static light scattering was performed on the $E2_{medCTD}$-I2 complex. The concentration of the protein complex used was 2 mg/ml. The protein was further purified using a Superdex 200 column (GE Healthcare). MALS was conducted using a MiniDAWN and Optilab system (Wyatt Technologies). Data was collected and analyzed using the Astra software package (Wyatt Technologies).

ITC

ITC measurements were performed with a MicroCal PEAQ-ITC microcalorimeter (Malvern). Titrations were carried out with 125 μM of the immunity protein in the syringe and 20 μM of $E2_{medCTD}$ in the cell. The titration experiment consisted of one 0.4-μl injection followed by 24 1.5-μl injections with 150-s intervals between each injection. The ITC data were analyzed using the Origin software (version 5.0, MicroCal, Inc.) and fit using a single-site binding model.

Sequence Alignments and 3D-structure Prediction

All sequence alignments were performed using ESPript 3.0 (Robert & Gouet 2014). The likely position of secondary structural features was based on the AlphaFold predicted model of the 3D structure of the proteins.

For the prediction of the 3D-strcuture of the proteins, I used the latest version of AlphaFold 2.0 (Jumper et al. 2021) using the online resource:

https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb

Western Blot

Western blot analyses of protein samples were performed using a SDS-PAGE gel and buffer system and a standard western blotting protocol. After SDS-PAGE separation, proteins were wet-transferred to 0.45 μm nitrocellulose membranes (100 V for 30 mins, 4°C). The samples were then analyzed with Western blot using the protein-specific rabbit primary antibodies α-VSVG (1:5000, 1hr) and a goat α-rabbit secondary antibody (Sigma, 1:5000, 45 minutes). Western blots were imaged using a ChemiDoc System (Bio-Rad).

Growth Assay

Individual strains were grown overnight at 37°C in LB with shaking. The day after, 1 μL of the overnight culture was added into 200 μL LB broth and grown at 37°C with shaking. OD600 readings were taken every 15 minutes using an Epoch 2 plate reader (Biotek) until the OD600 of the strains were close to 0.3 at which point the expression of toxin/immunity was induced. Afterwards, the strains were allowed to grow at the same condition for 15 more hours. For every given strain and condition, the experiment was repeated three times and the average growth was shown in the graph. The error bars were calculated using the standard deviation of the three experiments.

RESULTS

My first goal was to understand if the *Parabacteroides sp.* D25 E2-I2 function as a toxin-immunity pair. To address this question, I began my project by setting out to perform toxicity assays in *Escherichia coli* where I attempted to clone the toxin domain of E2 for heterologous expression. For the majority of the T6SS effectors identified so far, the toxic domain is located within the C-terminal domain (CTD) of the effector locus. Guided by this precedent, I designed three different truncations of the effector CTD and tested if any of them were toxic when cloned into *E. coli* and subsequently expressed. As shown in Figure 3A, one of these truncations, henceforth referred to as $E2_{tox}$, proved to be toxic to *E. coli* and thus was used for all subsequent toxicity assays. Since most T6SS effectors identified to date are encoded upstream of their cognate immunity proteins, I next set out to test whether $E2_{tox}$-mediated toxicity could be protected against by its downstream open reading frame (I2) (Figure 2A). As shown in Figure 3A, expression of I2 alleviates the toxic effects of $E2_{tox}$ and this capability is specific to I2 as I5, an immunity protein that confers immunity to another family of T6SS effectors found in the human gut microbiome, was unable to protect against $E2_{tox}$ (Figure 3A).

As shown in Figure 2, there are four homologues of I2 with varying degrees of identity to I2 that we reasoned would allow us to test the capacity of divergent I2-like proteins to restore growth to a strain that expresses $E2_{tox}$. Two of these I2 genes, I2_PV and I2_BU, are found in effector-immunity bicistrons while the other two, Orf3 and Orf5, are found in

a rAID cluster suggesting they may be orphan immunity genes. These genes have levels of identity to *Parabacteroides sp. D25* I2 ranging from 40% to approximately 75% and our prediction was that the I2 homologues possessing greater identity level to I2 would be better able to protect against $E2_{tox}$-mediated toxicity.

I set out to clone these immunity genes and perform toxicity assays, which would allow me to test the capacity of each of the immunity genes to restore *E. coli* growth. As shown in Figure 3B, I found that all these four homologues of I2 are capable of protecting against $E2_{tox}$-mediated toxicity just as well as its cognate I2 suggesting that surprisingly, despite the wide range of sequence identity levels to cognate I2, all four of these I2 homologues may be effective in conferring resistance against E2-mediated attack. I next sought to identify a non-protective I2 homologous protein and to do so, I searched for more distant homologues of I2 to see if a greater difference in the level of sequence identity to I2 would lead to the identification of a SUKH immunity protein unable to protect against $E2_{tox}$. I identified a homologue in a *Methylomonas* species called LW13, which belongs to the phylum Proteobacteria instead of Bacteroidetes, which has an approximately 30% identity level to *Parabacteroides sp. D25* I2. Using the same cloning and toxicity assay as described previously, I was able to show that this I2 homologue, henceforth referred to as I2_Meth, was incapable of protecting against $E2_{tox}$-mediated toxicity (Figure 3B). In summary, my toxicity data show that while all four I2 homologous genes from Bacteroidetes protect against E2-mediated toxicity, a Proteobacterial I2 homologue does not.

Importantly, the *Parabacteroides sp. D25* cognate immunity I2 is made up of two SUKH domains. The other four Bacteroidetes I2 homologues all capable of protecting against

toxicity are made up of two SUKH domains as well. By contrast, I2_Meth consists of only one SUKH domain resulting in it being only half the size as I2 and its other homologues. This observation raised the possibility that the reason I2_Meth does not protect against E2 toxicity was due to its 'incomplete' length. However, I noted using sequence alignments that the C-terminal SUKH domains of the two-domain I2 proteins has much higher levels of identity among the various I2 homologous than the N-terminal SUKH domain suggesting that this domain might be sufficient to protect against $E2_{tox}$ (Figure 3C). To test this, I cloned each SUKH domain in isolation and transformed them into *E. coli* cells along with $E2_{tox}$ to perform toxicity assays. As shown in Figure 3D, the C-terminal SUKH domain is both necessary and sufficient to protect against $E2_{tox}$-mediated toxicity. This result suggests that the inability of I2_Meth to protect against $E2_{tox}$-mediated toxicity is not due it lacking an N-terminal SUKH domain but rather its sequence divergence from *Parabacteroides sp. D25* I2.

Having identified a collection of I2 homologues capable or incapable of protecting against $E2_{tox}$-mediated toxicity, I next proceeded to study the process of toxin neutralization in more detail. Based on the existing data in the literature, there are two possible mechanisms of toxin inhibition by I2-like immunity proteins that I needed to consider (Klein et al. 2020). First, I2 could be physically interacting with $E2_{tox}$ and by doing so, would make the enzyme incapable of performing its biochemical toxic function(s). Alternatively, immunity could be an enzyme that functions by biochemically reversing the activity of the toxin. To test the first possibility, I cloned three truncations of my effector E2 from its CTD, which I named short, medium, and long for their respective lengths (the long construct being $E2_{tox}$

from my previously described toxicity assays) and tested their ability to copurify with I2. I2 was found to bind to both the long and medium constructs indicating that E2 and I2 interact strongly. This result suggests that physical association is the mechanism underlying effector neutralization by the immunity protein. The result for the medium construct is shown in Figure 4A.

To understand more about the nature of interaction and to uncover the molecular mechanism of toxin neutralization, I next attempted to solve the structure of the E2-I2 complex. The overexpression of the toxin-immunity complex was most successful when the medium truncation of the E2 C-terminal domain (henceforth referred to as $E2_{medCTD}$) was co-expressed with I2 (Figure 4A). Therefore, I first tried to purify this complex and use the resulting protein for crystallization experiments. Despite the high level of expression of $E2_{medCTD}$ and the successful co-purification of the $E2_{medCTD}$-I2 complex by size-exclusion chromatography (SEC), I was not able to grow any diffraction quality crystals from the 384 screening conditions routinely used for protein crystallography in our lab.

Consequently, I turned my attention to finding another complex that might be more conducive to crystallization. Having determined that the C-terminal SUKH domain of I2 is necessary and sufficient to rescue $E2_{tox}$-mediated toxicity, I reasoned that the CTD of I2 is almost certainly also necessary and sufficient for its interaction with $E2_{tox}$ and by extension $E2_{medCTD}$. At around the same time, researchers released the latest version of a protein 3D-structure prediction software system named TrRosetta (Du et al. 2021). Using TrRosetta, I input I2 to learn more about its likely 3D-structure, and more specifically, the whereabouts

of the linker region connecting its two SUKH domains. Based on the output of trRosetta, I decided to truncate I2 at three different positions, starting at the 161st, 175th and 182nd amino acid positions through to the end of the protein. I then co-expressed each of these three truncations with $E2_{medCTD}$ and examined their ability to co-purify with $E2_{medCTD}$. In support of the trRosetta model, all three truncations copurified with $E2_{medCTD}$. I moved on with further purifying these complexes using SEC and setting up crystal screens in the same manner as I did previously for full-length I2. Unfortunately, none of these complexes formed crystals.

Having not been able to crystallize the $E2_{medCTD}$-immunity complex, I next decided to use the I2_PV and I2_BU homologues of I2 as the binding partners of $E2_{medCTD}$. In doing so, I would not only be able to confirm that these I2 homologues also bind tightly to $E2_{medCTD}$ like the cognate immunity protein but I would also be able to test if perhaps any of these homologues would be more conducive to crystallization as part of an effector-immunity complex. From these experiments, I was able to show that $E2_{medCTD}$ binds tightly to both I2_PV and I2_BU, supporting a similar physical inhibition mechanism of toxin neutralization. Unfortunately, neither of these two new complexes resulted in diffraction quality crystals suggesting that obtaining an experimental structure of $E2_{tox}$ with any of the characterized I2 homologues would be a difficult path moving forward.

To increase my chances of crystallization, I next turned my attention to other E2 toxin homologues beyond the *Parabacteroides sp. D25* protein. The *Methylomonas* sp. immunity protein used as a non-protective I2 homologue in my toxicity assays is itself the cognate immunity of a putative effector-immunity pair in this organism. In order to ensure that this

putative effector is toxic to *E. coli,* I tested three truncations of its CTD for growth inhibition. Much like the *Parabacteroides* effector, one of these truncations proved to be toxic in *E. coli* and is henceforth referred to as E2_Meth. Additionally, I was able to show that this toxicity is abrogated when the immunity protein, I2_Meth, is co-expressed with E2_Meth. These results mean that the E2_Meth and I2_Meth loci likely function as an effector-immunity for a putative T6SS in *Methylomonas*. Therefore, I embarked on protein crystallization for this protein complex in the hopes that it would be more amenable to structure determination.

First, I determined that the short and medium immunity constructs, referred to as Short E2_Meth and Med E2_Meth, express well in *E. coli*. I was able to show the Med I2_Meth copurifies with E2_Meth in the same manner that E2 co-purifies with I2. I further purified the *Methylomonas* complex and set up crystallization screens, which unfortunately again did not result in the formation of crystals.

For my last attempt at crystallization, I decided to co-express the short E2_Meth truncation along with I2_Meth. Unfortunately, however, my efforts to obtain diffraction quality crystals using this complex proved unsuccessful. In summary, despite my work on searching for an E2-I2 effector-immunity complex that is abundantly expressed, efficiently co-purified, and crystalized in one of our 384 buffer conditions, I was unable to determine the crystal structure of this effector-immunity pair. Table 2 summarizes my crystallization efforts including the starting concentration of the protein complexes used for each round of crystal screening.

Having been unsuccessful at solving the crystal structure of the E2-I2 effector-immunity complex, I decided to study the molecular nature of their interaction using other biophysical methods. One strategy that would allow us to quantitatively measure the affinity between E2 and the protective and non-protective homologues of I2 is Isothermal Titration Calorimetry (ITC). For this line of investigation, I expressed and purified the individual effector and immunity proteins. These proteins included the medium truncation of the E2 C-terminal toxin domain $E2_{medCTD}$, its cognate immunity I2, the four microbiome-derived homologues of I2 that protect against $E2_{tox}$-mediated toxicity, and finally, I2_Meth, the SUKH immunity gene that does not protect against $E2_{tox}$. I then used either the cognate immunity or each of the immunity homologues as my titrant and gradually added it to $E2_{medCTD}$ to measure binding as discussed in the methods section. As shown in Table 3 and Figure 4B-D, the four I2 immunity homologues that did protect against $E2_{tox}$-mediated toxicity, also interact with $E2_{medCTD}$ with approximately the same nanomolar binding affinity as the cognate I2 immunity. Additionally, I2_Meth showed no detectable binding to $E2_{medCTD}$ supporting our findings from the copurification studies that indicated that the neutralization of $E2_{tox}$-mediated toxicity is mediated through direct binding of the effector with the immunity protein.

In summary, the combination of the cell-based toxicity assays as well as the ITC experiments using purified proteins collectively demonstrate that despite the differing levels of sequence identity between I2 and the four homologues from Bacteroidetes, they all neutralize $E2_{tox}$-mediated toxicity to the same extent. Furthermore, this neutralization is very likely due to the direct binding of $E2_{tox}$ to each of the immunity proteins and I found

that the binding affinity of all four protective homologues towards $E2_{tox}$ is approximately the same, regardless of their level of sequence identity to the cognate immunity protein, I2.

Having established that a panel of divergent I2 immunity proteins maintain the ability to interact with E2 with comparable affinity, I was curious as to what the molecular features of these immunity proteins are that facilitate their binding to $E2_{tox}$. As my efforts to solve the structure of $E2_{tox}$-I2 proved unsuccessful, I had to find other ways to predict the potential interaction interface between these proteins. To this end, I employed the recently released AlphaFold2 (AF2) software package (Jumper et al. 2021). Published in mid-2021, this AI-assisted structure prediction software predicts the 3D-structure of proteins with unprecedented accuracy. I input my cognate I2 immunity protein as well as the five other I2 homologues (protective and non-protective) into AF2. Interestingly, as shown in Figure 4D, there do not seem to be many differences between the cognate immunity I2 and the rest of the homologues including, surprisingly, I2_Meth, which neither protected against $E2_{tox}$-mediated toxicity nor showed a measurable binding affinity towards $E2_{medCTD}$. This suggested to me that the mode of interaction between the two proteins is not simply reliant on protein shape but depends on some other physico-chemical feature of the proteins that differs between I2_Meth and the rest of the I2 homologues.

Therefore, I generated a sequence alignment of I2 and the five non-cognate immunity proteins to find regions of I2 that are the same among the protein sequences that protect against $E2_{tox}$-mediated toxicity but are different from the non-protective I2_Meth homologue. As shown in Figure 5, there are 27 residues that meet that criterion. Using ChimeraX (Petterson et al. 2021) and the AF2-predicted structure of I2, I next attempted to

refine the list of candidate residues by identifying those that are predicted to occur on the surface of the protein. I further narrowed down my list by selecting residues that have markedly different biochemical properties than what is found in I2 (e.g. oppositely charged amino acid side chain, etc.). I was left with five residues that were confidently predicted to be surface exposed and possess R-groups with substantially different properties compared to their I2_Meth counterparts. I mutated these residues in I2 to amino acid residues bearing substantially differing properties using site-directed mutagenesis. Once cloned, these I2 variants were used in toxicity assays to compare their ability to protect against $E2_{tox}$-mediated toxicity relative to wild-type I2. Surprisingly, all single-site variants exhibited the same level of protection as I2 (data not shown). Furthermore, binary combinations of these mutations did not alter the ability of I2 to protect against $E2_{tox}$. Based on these results, I concluded that the buried surface area of the E2-I2 interaction must be quite substantial and that I may need to make more drastic substitutions in the I2 sequence in order to abrogate its ability to bind $E2_{tox}$ and neutralize $E2_{tox}$-mediated toxicity.

At around the same time I was conducting the above mutagenesis studies, AlphaFold2 released their structure prediction software to include the accurate prediction of two-protein complexes. Taking advantage of the new update, I decided to input my two proteins of interest to see if AlphaFold2 could predict a structure of the complex that I could use to predict important residues in the interaction interface. Figure 6A shows the AF2-predicted structure of the $E2_{tox}$-I2 complex. If this structure is accurate, it appears that only two out of the five residues previously mutated exist in the interaction interface perhaps explaining why the previous mutagenesis approach was unsuccessful.

Based on this new information, I decided to do another round of mutagenesis this time with the help of the aforementioned AF2-predicted E2-I2 complex structure. Additionally, I was less conservative with the types of residues I selected when it came to their surface exposure; as long as a given residue was chemically different enough between the non-protective I2_Meth and the protective I2 homologues and was located in the predicted-interaction interface, it was eligible to be included as a candidate residue.

In order to be more systematic in my mutagenesis, I also used the model to predict a possible mode of interaction between the effector and immunity protein. One possibility would have been specific shape-based structural features that facilitate binary interaction between the two proteins. This prediction, however, implies that I2_Meth would be structurally different from the protective I2 homologues. However, based on the analysis discussed above and the predicted structures shown in Figure 6B, this does not appear to be the case. Therefore, another possible mode of interaction that I considered was based on the biochemical features of the residues located in the interaction interface of the two proteins. One striking feature I noticed is the highly electrostatic nature of E2 and I2. Indeed, as shown in Figure 7A, the predicted interaction interface between the two proteins appears to predominantly consist of oppositely charged residues, implying a possible role for electrostatics in driving the specificity of the interaction between the two proteins. More specifically, the interaction interface of the immunity protein consists of basic residues that interact with acidic residues on the effector, resulting in the physical interaction between the two proteins.

To test my proposed electrostatic model of interaction, I initiated mutagenesis on acidic residues on I2 and replaced them with either of the basic residues; arginine or lysine. I found a variable region in I2 that contained several acidic residues not found in I2_Meth, in particular residues E228, D229, D231, E234, and N236. As shown in Figure 7B, upon mutation of these five residues, the growth rate of a strain expressing both $E2_{tox}$ and the mutated I2 demonstrated growth inhibition that was close to that of a strain expressing just the $E2_{tox}$ alone. Using a Western blot analysis, I also confirmed that this growth defect is not due to the inability of the strain to express the mutated I2 variant (Figure 7C). Overall, these results support my hypothesis that the acidic residues located on I2 play a critical role in conferring binding specificity to $E2_{tox}$ and thus confer the ability to protect against its toxic activity in *E. coli* and that the physical interaction between $E2_{tox}$ and I2 is mediated by electrostatic interactions.

DISCUSSION

Diverse Immunity Proteins Protect Against Single Effectors

In their initial discovery of rAID clusters, Ross et al. found that the rAID-encoded orphan immunity genes range anywhere from having 90% identity to just over 30% identity to cognate immunity proteins (Ross et al. 2019). This large range of sequence identity between the cognate immunity proteins and orphan immunity proteins raises a number of interesting questions. First, and perhaps most obvious, is whether all of these orphan immunity genes that exist as part of rAID clusters encode functionally useful proteins; that is to say whether they confer to the strains harbouring them the ability to protect against effector-mediated attacks by the T6SSs of other bacteria in the environment of the human gut. Another question is whether the differences in percent sequence identity are useful in terms of predicting *a priori* the degree to which a given rAID-encoded immunity protein can neutralize an incoming effector. More specifically, do immunity proteins with lower levels of sequence identity to their corresponding effector proteins protect against the toxicity of the corresponding cognate effector less effectively than the cognate immunity protein itself or compared to other orphan immunity homologues with higher levels of identity to the cognate immunity gene, or do they possess approximately the same level of protection. From a bacterial competition standpoint, when injected with a given effector, would strains encoding an orphan immunity gene with low sequence identity to the cognate immunity grow slower than one with high sequence identity?

The findings from my investigation of a specific family of effector-immunity proteins abundantly found in the human gut microbiome have shown that at least in one case, large differences in percent sequence identity do not significantly affect the capacity of immunity proteins to protect against effector-mediated toxicity. Specifically, my results indicate that in *E. coli*, E2-mediated toxicity is protected against by its cognate immunity protein as well as by all other tested homologues from Bacteroidetes, which possess as low as 38% sequence identity to cognate immunity. This is approximately equal to the percent identity that Ross and colleagues had reported in their paper to be the lowest percent identity found in the rAID clusters of the sequences they surveyed (Ross et al 2019), suggesting the possibility that many orphan immunity genes, regardless of their sequence identity to the corresponding cognate immunity gene, possess potent effector-neutralizing capacity. However, one important caveat of my thesis work is that all my experiments were carried out in *E. coli*. The heterologous nature of this system and the fact that I was overexpressing proteins from inducible plasmids means that artificially high amounts of proteins likely exist in the cell. This in turn could result in ratios of immunity and effector proteins that might be far different from those in nature. Therefore, future studies should focus on measuring the ability of the various I2 immunity proteins examined in this work to protect against effector toxicity in their native context; that is, when attacked by a Bacteroidetes strain encoding the effector as part of its T6SS. This could be done using techniques such as bacterial competition assays using both wild-type strains and strains lacking E2 (attacker) or lacking the various rAID-encoded I2 genes (defender). Even then, the results of such assays alone may not give an accurate representation of what is happening in the

dense environment of the human gut microbiota, with a huge number of different bacterial species and strains (Huttenhower et al. 2012; King et al 2019) as well as constantly changing environmental conditions. To most accurately measure the ability of these immunity proteins to protect against their corresponding effectors, future studies should focus their efforts on studying the growth rate of immunity protein-containing bacteria in a densely populated bacterial environment. To best reflect the conditions found in the human gut, mouse model studies such as those performed by Comstock and colleagues and Wexler and colleagues (Chatzidaki-Livanis et al. 2016; Wexler et al. 2016) are highly needed to better understand the physiological significance of rAID-encoded T6SS effector resistance. Another outstanding question is regarding the *B. fragilis* BIOML-A5 rAID cluster that contains both the Orf3 and Orf5 I2 immunity proteins. In light of my data showing that either one of them can fully protect against $E2_{tox}$-mediated toxicity in *E. coli* and binds to $E2_{tox}$ with the same approximate affinity, the question that arises is what is the significance of having two closely related immunity proteins in the same rAID cluster? Importantly, as shown in Figure 8, the two proteins are identical in their C-terminal SUKH domains. In cognate I2, I demonstrated that the C-terminal SUKH domain is both necessary and sufficient for protection against $E2_{tox}$-mediated toxicity, suggesting the same is likely true for Orf3 and Orf5. One possibility is that in the context of the gut microbiota, individual cells of *B. fragilis* BIOML-A5 are injected with large doses of effectors that belong to the same family as $E2_{tox}$, namely the WHH family of HNH nucleases. Even though not much is known about the regulatory aspects of the genes inside rAID clusters, the presence of two copies of an immunity gene suggests that there are twice as many immunity proteins

available to neutralize the T6SS attack of another strain. Alternatively, the divergence of the N-terminal SUKH domains of Orf3 and Orf5 may be helpful in combating the toxic effects of a diversity of E2 homologous toxins.

Another important point to consider is that even though I have shown the importance of only one of the SUKH domains in the protection against effector-induced toxicity, it is still quite curious that every one of these immunity genes found in human gut bacteria is comprised of two SUKH domains. As Zhang et al. have pointed out from their gene neighborhood analyses (Zhang et al 2011), an important feature of SUKH-domain containing proteins is that they occur next to many other types of nucleases in addition to the members of the HNH superfamily, suggesting that they may have the ability to act as immunity genes against many such toxins in the context of different interbacterial antagonism pathways. Therefore, I speculate that the N-terminal SUKH domain of I2 homologous proteins may be functionally important in neutralizing other such toxins targeted at the bacteria harbouring them. Alternatively, the N-terminal SUKH domain could have a regulatory function on the action of other cytosolic nucleases that are not secreted from the bacterial cell that have other physiological functions. SUKH-mediated inhibition could serve as a regulatory switch that upon some sort of physiological cue results in the release of the SUKH domain and the activation of the nuclease.

The Mutational Landscape of Immunity Proteins

The question of how effector and immunity proteins of a given T6SS interact with one another is one of biochemical interest to the large field of protein-protein interactions. A number of studies have revealed crucial insights on the nature of such interactions in

different bacterial protein toxin-immunity complexes as well as how mutations in each protein component of the complex can play a role in disrupting or altering the specificity of the interaction. One of the first studies that investigated the interaction of an effector-immunity pair was in the context of a *Pseudomonas aeruginosa* T6SS. This study found that the immunity protein, Tsi2, tightly binds its cognate effector, Tse2, and that this interaction is easily disrupted by the mutation of single acidic residues on Tsi2 located at the binding interface of the two proteins into a basic residue (Li et al. 2012) suggesting that surface electrostatics can play an important role in the physical interaction of an effector-immunity pair.

A similar phenomenon has been observed in toxin-antitoxin (TA) systems. In TA systems, the toxin is a non-secreted protein whose activity can inhibit cell growth under certain conditions but otherwise is inhibited by a cognate antitoxin. In the well-characterized *E. coli* MazF-MazE TA system, the MazE antitoxin has a highly acidic C-terminus that interacts with MazF, which is thought to be the same site that MazF interacts with its RNA substrates, explaining how MazE inhibits MazF toxicity inside the cell (Yamaguchi et al. 2011; Kamada et al. 2003).

The above examples, in addition to my findings on the mode of interaction between E2 and I2, suggest that complementary surface electrostatics could be the driving factor behind the protein-protein interactions that take place in numerous different bacterial toxin systems where the activity of the immunity/antitoxin protein is primarily to inhibit the enzymatic activity of its binding partner.

Understanding the exact mutational landscape of the specificity in the interaction interface between two proteins is still an intensely studied question in protein biochemistry. In one of the early systematic investigations of protein-protein interactions at the amino acid residue level, researchers showed that there is a very strong selective pressure in what they referred to as "covarying residues" at the interaction interfaces of mitochondrial and bacterial protein complexes; a finding that suggests that contacting residues at the interface of two interacting proteins evolve simultaneously in order to preserve the nature of the physical interaction between the two proteins (Ovchinnikov et al. 2012).

Based on computational analyses, other researchers have carried out some of the most comprehensive systematic studies of the nature of specificity in protein-protein interactions using bacterial toxins as model systems. For example, the ParD-ParE TA system is found encoded in the chromosomes of many bacteria, where ParE is the toxin and ParD the antitoxin (Fiebig et al. 2010; Leplae et al. 2011). In a series of studies, researchers have investigated the features of the residues at the interaction interface of this toxin-antitoxin pair that confers a stable interaction between the two proteins (Aakre et al. 2015; Lite et al. 2020; Ding et al 2022). After identifying coevolving residues using the computational method designed by Baker et al., the authors set out to test the importance of these residues in conferring interaction specificity using experimental approaches. Using *E. coli* as a biological system, they showed that the toxin and antitoxin in the ParD-ParE systems from several different bacteria are highly specific for one another but that mutating only four residues in the antitoxin ParD3 was sufficient to change its specificity from its cognate toxin, ParE3, to other toxins in the ParD-ParE systems such as ParE2 or ParE1 (Aakre et

al. 2015). Then, after showing that the ParE2-ParD2 system is a result of gene duplication and subsequent divergence from the ParE3-ParD3 system, Lite et al. created an exhaustive library based on the three different residues between ParD2 and ParD3 containing 8000 variants that include all amino acids at each position in ParD3 and they discovered that there are a large number of variants that can interact with both ParE2 and ParE3 toxins in a promiscuous manner (Lite et al. 2020). These results suggest two things: that again the mutation of a few residues is sufficient to change the specificity of the interaction of a protein partner from one partner to another, and secondly, that many amino acid substitutions create an intermediate state where a protein will have the capacity to bind two homologous partners perhaps explaining the mutational paths for evolution of interaction specificity.

Additionally, the interaction specificity of two proteins for one another can be conducted by identifying residues on the surface of the proteins that mediate specific interaction between the cognate pair (positive residues) as well as those which prevent one protein from binding to the non-cognate protein partners (negative residues) (Schreiber & Keating 2010). Interestingly, Lite et al. also show that the three residues initially found to be necessary and sufficient for changing of the specificity all serve as positive elements whereas only two serve as negative elements demonstrating how each residue can have differing roles in the interaction interface of two interacting partners (Lite et al. 2020).

My work on the physical interaction of E2 and I2 suggests that only a small portion of the immunity protein plays a role in conferring the specificity of the interaction between the immunity protein and its cognate effector, which is crucial for the toxin neutralization

ability of the immunity protein. Specifically, the five residues E228, D229, D231, E234, and N236 are sufficient to mediate the specific interaction between E2 and I2 and the mutations to residues with opposite electrostatic features results in the abrogation of the tight interaction between the two, suggesting that a limited number of residues are at play similar to the ParD3-ParE3 case discussed above.

Modes of Interaction of Two Proteins in a Protein Complex

As has been also described in the work on the ParD3-ParE3 TA system, the biochemical features of the important residues at the interface play an important role on whether a given variant of the ParD3 antitoxin can interact with its cognate ParE3, a non-cognate ParE2, or both. It should be noted however that there are other forces that may play a role in conferring stability to a protein complex. One such example, is the presence of "interface add-ons", which are structural elements, such as loops or secondary structures, present in the interaction interface of two proteins that help with the stability of the complex (Plach et al 2017). Most recently, in the three paralogous TacAT systems, which are TA systems found in *Salmonella enterica* serovar Typhimurium, the presence of an α-helical secondary structure add-on element was found to be important in insulating different types of paralogous toxin-antitoxin pairs from one another (Grabe et al. 2021). In summary, despite several studies, including my own, pointing to electrostatic forces as important determinants of specificity for protein-protein interactions in the context of bacterial toxin systems and beyond, these examples show that my findings cannot necessarily be extended to all effector-immunity pairs in the various T6SS's found in numerous bacteria.

Lastly, another important insight that has come out of the ParD-ParE studies is the importance of residues that reside outside of the interaction interface of two proteins in conferring to one protein the ability to physically interact with different variants of its partner. Using high-throughput mutagenesis, researchers have identified that mutations in the ParE3 toxin in many residues that exist outside of its interaction interface with the ParD3 antitoxin help define how it interacts with different variants of the antitoxin (Ding et al. 2022). This finding suggests that there is much more nuance to the evolutionary paths of co-evolving proteins than just the accumulation of mutations at their interaction interface. In summary, more systematic studies are needed to shed light on the nature of effector-immunity interactions in T6SS and other bacterial toxin systems as well as to understand what drives the ability of either protein partner to specifically interact with its cognate partner or to be able to accommodate a large number of non-cognate interactions with homologues of its cognate immunity as I have observed for I2.

Biochemical Activity of E2 Effectors and How I2 Immunity Blocks Its Activity

The E2 effector belongs to the HNH family of nucleases and contains a conserved WHH region within its C-terminal domain. Little has been done on this toxin's mode of action and how its activity stops cell growth. However, during my E2-I2 crystallization efforts I found that the E2-I2 complex associates with DNA (data not shown). Whether this interaction is biologically meaningful and how it would lead to inhibition of cell growth are outstanding questions that require further investigation. Additionally, it would be interesting to determine if the interaction interface of I2 and E2 overlaps with where E2

interacts with DNA similar to the MazF-MazE TA system whereby the MazE antitoxin interacts with MazF where it is thought to interact with its RNA substrates (Zhang et al 2003; Zhang et al 2005).

Adaptive Evolution of Orphan Immunity Genes

Over the course of our bioinformatic analyses and based on a previous study by Zhao and colleagues (Zhao et al 2019), we found that some genes present in the rAID clusters have acquired a number of point mutations over time, suggesting they may have adaptive functions. However, based on our observation that both the Orf3 and Orf5 orphan immunity proteins can protect against $E2_{tox}$-mediated toxicity just as well as cognate I2, the observed point mutations are probably more likely to be neutral. Furthermore, now that I have defined the interaction interface for this effector-immunity pair, a more focused effort could be made on what types of mutations would have the potential to make the affinity between the immunity and the effector proteins stronger based on the fact that only a limited number of residues seem to be involved. Additionally, if for this family of effectors, we cannot identify orphan immunity genes that "evolve" over time to better neutralize it, future studies could more specifically focus on finding mutations that occur in the interaction interface of the two proteins to look for potential adaptive mutations.

REFERENCES

Aakre, C.D., Herrou, J., Phung, T.N., Perchuk, B.S., Crosson, S. and Laub, M.T., 2015. Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell*, 163(3), pp.594-606.

Alcoforado Diniz, J. and Coulthurst, S.J., 2015. Intraspecies competition in *Serratia marcescens* is mediated by type VI-secreted Rhs effectors and a conserved effector-associated accessory protein. *Journal of bacteriology*, 197(14), pp.2350-2360.

Aoki, S.K., Pamma, R., Hernday, A.D., Bickham, J.E., Braaten, B.A. and Low, D.A., **2005.** Contact-dependent inhibition of growth in *Escherichia coli*. *Science*, *309*(5738), pp.1245-1248.

Basmaji, F., Martin-Yken, H., Durand, F., Dagkessamanskaia, A., Pichereaux, C., Rossignol, M. and Francois, J., 2006. The 'interactome'of the Knr4/Smi1, a protein implicated in coordinating cell wall synthesis with bud emergence in *Saccharomyces cerevisiae*. *Molecular Genetics and Genomics*, 275(3), pp.217-230.

Cao, Z., Casabona, M.G., Kneuper, H., Chalmers, J.D. and Palmer, T., 2016. The type VII secretion system of Staphylococcus aureus secretes a nuclease toxin that targets competitor bacteria. *Nature microbiology*, *2*(1), pp.1-11.

Cardona, S.T. and Valvano, M.A., 2005. An expression vector containing a rhamnose-inducible promoter provides tightly regulated gene expression in *Burkholderia cenocepacia*. *Plasmid*, 54(3), pp.219-228.

Chatzidaki-Livanis, M., Geva-Zatorsky, N. and Comstock, L.E., 2016. *Bacteroides fragilis* type VI secretion systems use novel effector and immunity proteins to antagonize human gut Bacteroidales species. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(13), pp.3627-3632.

Coyne, M.J., Zitomersky, N.L., McGuire, A.M., Earl, A.M. and Comstock, L.E., 2014. Evidence of extensive DNA transfer between bacteroidales species within the human gut. *MBio*, *5*(3), pp.e01305-14.

Coyne, M.J., Roelofs, K.G. and Comstock, L.E., 2016. Type VI secretion systems of human gut Bacteroidales segregate into three genetic architectures, two of which are contained on mobile genetic elements. *BMC genomics*, 17(1), pp.1-21.

Crits-Christoph, A., Diamond, S., Butterfield, C.N., Thomas, B.C. and Banfield, J.F., 2018. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature*, *558*(7710), pp.440-444.

Ding, D., Green, A.G., Wang, B., Lite, T.L.V., Weinstein, E.N., Marks, D.S. and Laub, M.T., 2022. Co-evolution of interacting proteins through non-contacting and non-specific mutations. *Nature Ecology & Evolution*, 6(5), pp.590-603.

Du, Z., Su, H., Wang, W., Ye, L., Wei, H., Peng, Z., Anishchenko, I., Baker, D. and Yang, J., 2021. The trRosetta server for fast and accurate protein structure prediction. *Nature protocols*, 16(12), pp.5634-5651.

Fiebig, A., Castro Rojas, C.M., Siegal-Gaskins, D. and Crosson, S., 2010. Interaction specificity, toxicity and regulation of a paralogous set of ParE/RelE-family toxin–antitoxin systems. *Molecular microbiology*, 77(1), pp.236-251.

Fishel, B.R., Sperry, A.O. and Garrard, W.T., 1993. Yeast calmodulin and a conserved nuclear protein participate in the in vivo binding of a matrix association region. *Proceedings of the National Academy of Sciences of the United States of America*, 90(12), pp.5623-5627.

García-Bayona, L. and Comstock, L.E., 2018. Bacterial antagonism in host-associated microbial communities. *Science*, *361*(6408).

Gorbalenya, A.E., 1994. Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family. *Protein Science*, 3(7), pp.1117-1120.

Grabe, G.J., Giorgio, R.T., Hall, A.M., Morgan, R.M., Dubois, L., Sisley, T.A., Rycroft, J.A., Hare, S.A. and Helaine, S., 2021. Auxiliary interfaces support the evolution of specific toxin–antitoxin pairing. *Nature Chemical Biology*, 17(12), pp.1296-1304.

Hecht, A.L., Casterline, B.W., Earley, Z.M., Goo, Y.A., Goodlett, D.R. and Bubeck Wardenburg, J., 2016. Strain competition restricts colonization of an enteric pathogen and prevents colitis. *EMBO reports*, 17(9), pp.1281-1291.

Hersch, S.J., Manera, K. and Dong, T.G., 2020. Defending against the type six secretion system: beyond immunity genes. *Cell Reports*, 33(2), p.108259.

Hibbing, M.E., Fuqua, C., Parsek, M.R. and Peterson, S.B., 2010. Bacterial competition: surviving and thriving in the microbial jungle. *Nature reviews microbiology*, 8(1), pp.15-25.

Holberger, L.E., Garza-Sánchez, F., Lamoureux, J., Low, D.A. and Hayes, C.S., 2012. A novel family of toxin/antitoxin proteins in *Bacillus* species. *FEBS letters*, 586(2), pp.132-136.

Hong, Z., Mann, P., Brown, N.H., Tran, L.E., Shaw, K.J., Hare, R.S. and DiDomenico, B., 1994. Cloning and characterization of KNR4, a yeast gene involved in (1, 3)-beta-glucan synthesis. *Molecular and cellular biology*, 14(2), pp.1017-1025.

Hood, R.D., Singh, P., Hsu, F., Güvener, T., Carl, M.A., Trinidad, R.R., Silverman, J.M., Ohlson, B.B., Hicks, K.G., Plemel, R.L. and Li, M., 2010. A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell host & microbe*, *7*(1), pp.25-37.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl, A.M., FitzGerald, M.G., Fulton, R.S. and Giglio, M.G., 2012. Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), p.207.

Jana, B., Fridman, C.M., Bosis, E. and Salomon, D., 2019. A modular effector with a DNase domain and a marker for T6SS substrates. *Nature communications*, 10(1), pp.1-12.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. and Bridgland, A., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), pp.583-589.

Kamada, K., Hanaoka, F. and Burley, S.K., 2003. Crystal structure of the MazE/MazF complex: molecular bases of antidote-toxin recognition. *Molecular cell*, 11(4), pp.875-884.

Kim, S.G., Becattini, S., Moody, T.U., Shliaha, P.V., Littmann, E.R., Seok, R., Gjonbalaj, M., Eaton, V., Fontana, E., Amoretti, L. and Wright, R., 2019. Microbiota-derived lantibiotic restores resistance against vancomycin-resistant *Enterococcus*. *Nature*, *572*(7771), pp.665-669.

King, C.H., Desai, H., Sylvetsky, A.C., LoTempio, J., Ayanyan, S., Carrie, J., Crandall, K.A., Fochtman, B.C., Gasparyan, L., Gulzar, N. and Howell, P., 2019. Baseline human gut microbiota profile in healthy people and standard reporting template. *PloS one*, 14(9), p.e0206484.

Klein, T.A., Ahmad, S. and Whitney, J.C., 2020. Contact-dependent interbacterial antagonism mediated by protein secretion machines. *Trends in microbiology*, *28*(5), pp.387-400.

Koskiniemi, S., Lamoureux, J.G., Nikolakakis, K.C., t'Kint de Roodenbeke, C., Kaplan, M.D., Low, D.A. and Hayes, C.S., 2013. Rhs proteins from diverse bacteria mediate intercellular competition. *Proceedings of the National Academy of Sciences*, 110(17), pp.7032-7037.

Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Dreze, P. and Van Melderen, L., 2011. Diversity of bacterial type II toxin–antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic acids research*, 39(13), pp.5513-5525.

Li, M., Le Trong, I., Carl, M.A., Larson, E.T., Chou, S., De Leon, J.A., Dove, S.L., Stenkamp, R.E. and Mougous, J.D., 2012. Structural basis for type VI secretion effector recognition by a cognate immunity protein. *PLoS pathogens*, 8(4), p.e1002613.

Lite, T.L.V., Grant, R.A., Nocedal, I., Littlehale, M.L., Guo, M.S. and Laub, M.T., 2020. Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *Elife*, 9: e60924.

Ma, L.S., Hachani, A., Lin, J.S., Filloux, A. and Lai, E.M., 2014. *Agrobacterium tumefaciens* deploys a superfamily of type VI secretion DNase effectors as weapons for interbacterial competition in planta. *Cell host & microbe*, 16(1), pp.94-104.

Martin-Yken, H., François, J.M. and Zerbib, D., 2016. Knr4: a disordered hub protein at the heart of fungal cell wall signalling. Cellular microbiology, 18(9), pp.1217-1227.

Mougous, J.D., Cuff, M.E., Raunser, S., Shen, A., Zhou, M., Gifford, C.A., Goodman, A.L., Joachimiak, G., Ordoñez, C.L., Lory, S. and Walz, T., 2006. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science*, *312*(5779), pp.1526-1530.

Ovchinnikov, S., Kamisetty, H. and Baker, D., 2014. Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *Elife*, 3, p.e02030.

Pei, T.T., Li, H., Liang, X., Wang, Z.H., Liu, G., Wu, L.L., Kim, H., Xie, Z., Yu, M., Lin, S. and Xu, P., 2020. Intramolecular chaperone-mediated secretion of an Rhs effector toxin by a type VI secretion system. *Nature communications*, 11(1), pp.1-13.

Peterson, S.B., Bertolli, S.K. and Mougous, J.D., 2020. The central role of interbacterial antagonism in bacterial life. Current Biology, 30(19), pp.R1203-R1214.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H. and Ferrin, T.E., 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1), pp.70-82.

Pissaridou, P., Allsopp, L.P., Wettstadt, S., Howard, S.A., Mavridou, D.A. and Filloux, A., 2018. The Pseudomonas aeruginosa T6SS-VgrG1b spike is topped by a PAAR protein eliciting DNA damage to bacterial competitors. *Proceedings of the National Academy of Sciences of the United States of America*, 115(49), pp.12519-12524.

Plach, M.G., Semmelmann, F., Busch, F., Busch, M., Heizinger, L., Wysocki, V.H., Merkl, R. and Sterner, R., 2017. Evolutionary diversification of protein–protein interactions by

interface add-ons. *Proceedings of the National Academy of Sciences*, 114(40), pp.E8333-E8342.

Pukatzki, S., Ma, A.T., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W.C., Heidelberg, J.F. and Mekalanos, J.J., 2006. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proceedings of the National Academy of Sciences*, *103*(5), pp.1528-1533.

Rankin, D.J., Rocha, E.P. and Brown, S.P., 2011. What traits are carried on mobile genetic elements, and why?. *Heredity*, 106(1), pp.1-10.

Rietsch, A., Vallet-Gely, I., Dove, S.L. and Mekalanos, J.J., 2005. ExsE, a secreted regulator of type III secretion genes in *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America*, 102(22), pp.8006-8011.

Ross, B.D., Verster, A.J., Radey, M.C., Schmidtke, D.T., Pope, C.E., Hoffman, L.R., Hajjar, A.M., Peterson, S.B., Borenstein, E. and Mougous, J.D., 2019. Human gut bacteria contain acquired interbacterial defence systems. *Nature*, *575*(7781), pp.224-228.

Russell, A.B., Singh, P., Brittnacher, M., Bui, N.K., Hood, R.D., Carl, M.A., Agnello, D.M., Schwarz, S., Goodlett, D.R., Vollmer, W. and Mougous, J.D., 2012. A widespread bacterial type VI secretion effector superfamily identified using a heuristic approach. *Cell host & microbe*, *11*(5), pp.538-549.

Russell, A.B., Wexler, A.G., Harding, B.N., Whitney, J.C., Bohn, A.J., Goo, Y.A., Tran, B.Q., Barry, N.A., Zheng, H., Peterson, S.B. and Chou, S., 2014. A type VI secretion-related pathway in Bacteroidetes mediates interbacterial antagonism. *Cell host & microbe*, *16*(2), pp.227-236.

Robert, X. and Gouet, P., 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research*, 42(W1), pp.W320-W324.

Roelofs, K.G., Coyne, M.J., Gentyala, R.R., Chatzidaki-Livanis, M. and Comstock, L.E., 2016. Bacteroidales secreted antimicrobial proteins target surface molecules necessary for gut colonization and mediate competition in vivo. *MBio*, *7*(4), pp.e01055-16.

Sana, T.G., Flaugnatti, N., Lugo, K.A., Lam, L.H., Jacobson, A., Baylot, V., Durand, E., Journet, L., Cascales, E. and Monack, D.M., 2016. Salmonella Typhimurium utilizes a T6SS-mediated antibacterial weapon to establish in the host gut. *Proceedings of the National Academy of Sciences of the United States*, *113*(34), pp.E5044-E5051.

Santos, M.N.M., Cho, S.T., Wu, C.F., Chang, C.J., Kuo, C.H. and Lai, E.M., 2020. Redundancy and specificity of type VI secretion vgrG loci in antibacterial activity of *Agrobacterium tumefaciens* 1D1609 strain. *Frontiers in microbiology*, 10, p.3004.

Schreiber, G. and Keating, A.E., 2011. Protein binding specificity versus promiscuity. *Current opinion in structural biology*, 21(1), pp.50-61.

Shub, D.A., Goodrich-Blair, H. and Eddy, S.R., 1994. Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. *Trends in biochemical sciences*, 19(10), pp.402-404.

Souza, D.P., Oka, G.U., Alvarez-Martinez, C.E., Bisson-Filho, A.W., Dunger, G., Hobeika, L., Cavalcante, N.S., Alegria, M.C., Barbosa, L.R., Salinas, R.K. and Guzzo, C.R., 2015. Bacterial killing via a type IV secretion system. *Nature communications*, *6*(1), pp.1-9.

Tang, J.Y., Bullen, N.P., Ahmad, S. and Whitney, J.C., 2018. Diverse NADase effector families mediate interbacterial antagonism via the type VI secretion system. *Journal of Biological Chemistry*, 293(5), pp.1504-1514.

Verster, A.J., Ross, B.D., Radey, M.C., Bao, Y., Goodman, A.L., Mougous, J.D. and Borenstein, E., 2017. The landscape of type VI secretion across human gut microbiomes reveals its role in community composition. *Cell host & microbe*, 22(3), pp.411-419.

Walter, J., & Ley, R. (2011). The Human Gut Microbiome: Ecology and Recent Evolutionary Changes. *In Annual Review of Microbiology* (Vol. 65, Issue 1, pp. 411–429). Annual Reviews.

Wexler, A.G., Bao, Y., Whitney, J.C., Bobay, L.M., Xavier, J.B., Schofield, W.B., Barry, N.A., Russell, A.B., Tran, B.Q., Goo, Y.A. and Goodlett, D.R., 2016. Human symbionts

inject and neutralize antibacterial toxins to persist in the gut. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(13), pp.3639-3644.

Yamaguchi, Y. and Inouye, M., 2011. Regulation of growth and death in Escherichia coli by toxin–antitoxin systems. *Nature Reviews Microbiology*, 9(11), pp.779-790.

Zhang, Y., Zhang, J., Hara, H., Kato, I. and Inouye, M., 2005. Insights into the mRNA cleavage mechanism by MazF, an mRNA interferase. *Journal of Biological Chemistry*, 280(5), pp.3143-3150.

Zhang, Y., Zhang, J., Hoeflich, K.P., Ikura, M., Qing, G. and Inouye, M., 2003. MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. *Molecular cell*, 12(4), pp.913-923.

Zhang, D., Iyer, L.M. and Aravind, L., 2011. A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. *Nucleic acids research*, 39(11), pp.4532-4552.

Zhao, S., Lieberman, T.D., Poyet, M., Kauffman, K.M., Gibbons, S.M., Groussin, M., Xavier, R.J. and Alm, E.J., 2019. Adaptive evolution within gut microbiomes of healthy people. *Cell host & microbe*, *25*(5), pp.656-667.

TABLES

Table 1: List of the genes with demonstrated or putative nuclease activity discovered to act as the effector of secretion systems of various bacteria. Put. Means putative.

Study	*Species (Gram)*	Gene Name	Nuclease Family	Target
Holberger et al. (2012)	*B. subtilis(+)*	YobL	HNH with a conserved LHH	rRNA
Holberger et al. (2012)	*B. subtilis(+)*	YxiD	HNH	rRNA
Holberger et al. (2012)	*B. subtilis(+)*	YqcG	HNH with a Conserved GHE	rRNA
Holberger et al. (2012)	*B. cereus(+)*	BC_0920	EndoU nuclease	rRNA and tRNA
Holberger et al. (2012)	*B. subtilis(+)*	YeeF	Endonuclease NS_2	RNA (Put.)
Holberger et al. (2012)	*B. subtilis(+)*	YokI	HNH with a SHH signature	RNA (Put.)
Koskiniemi et. al (2013)	*D. dadantii(-)*	RhsB	HNH	DNA
Koskiniemi et. al (2013)	*D. dadantii(-)*	RhsA	Endonuclease NS_2	DNA
Koskiniemi et. al (2013)	*B. subtilis(+)*	WapA	HNH with a conserved LHH	tRNA
Ma et al. (2014)	*A. tumefaciens(-)*	TdeI and TdeII	Ntox15	DNA
Alcoforado Diniz et al. (2015)	*S. marcescens(-)*	Rhs2	HNH	DNA
Cao et al. (2016)	*S. aureus(+)*	EsaD	Endonuclease NS_2	DNA
Pissaridou et al. (2018)	*P. aeruginosa*	Tse7	HNH with a GHH signature	DNA
Jana et al. (2019)	*V. parahaemolyticus (-)*	V12_14465	PD-(D/E)XK nuclease superfamily	DNA
Jana et al. (2019)	*B. cereus(+)*	BC3021	PD-(D/E)XK nuclease superfamily	DNA

Santos et al. (2020)	*A. tumefaciens(-)*	V2a	HNH with a conserved AHH	DNA
Pei et al. (2020)	*A. dhakensis(-)*	TseI	Tox-HNH-EHHH	DNA

Table 2: List of protein complexes for which crystal trays were set up

Protein Complex	**Concentration Used (mg/ml)**	**Protein:Crystalization Solution**
E2 & I2	10	1:1
E2 & I2_PV	12	1:1
E2 & I2_PV	15	1:1
E2 & I2_BU	11	1:1
E2 & Short I2	10	1:1
E2 & Medium I2	10	1:1
E2 & Long I2	10	1:1
Med E2_Meth _CTD & I2_Meth	7	1:1
Short E2_Meth _CTD & I2_Meth	8	1:1

Table 3: Summary of the results from the ITC experiments shown in Figure 4D

Name	Identity to I2	Physical Interaction with E2	K_D Value (nM)
I2	100%	Yes	27
I2_PV	76%	Yes	5
I2_BU	68%	Yes	18
Orf3	37%	Yes	2
Orf5	39%	Yes	4.5
I2_Meth	28%	No	N/A

FIGURES

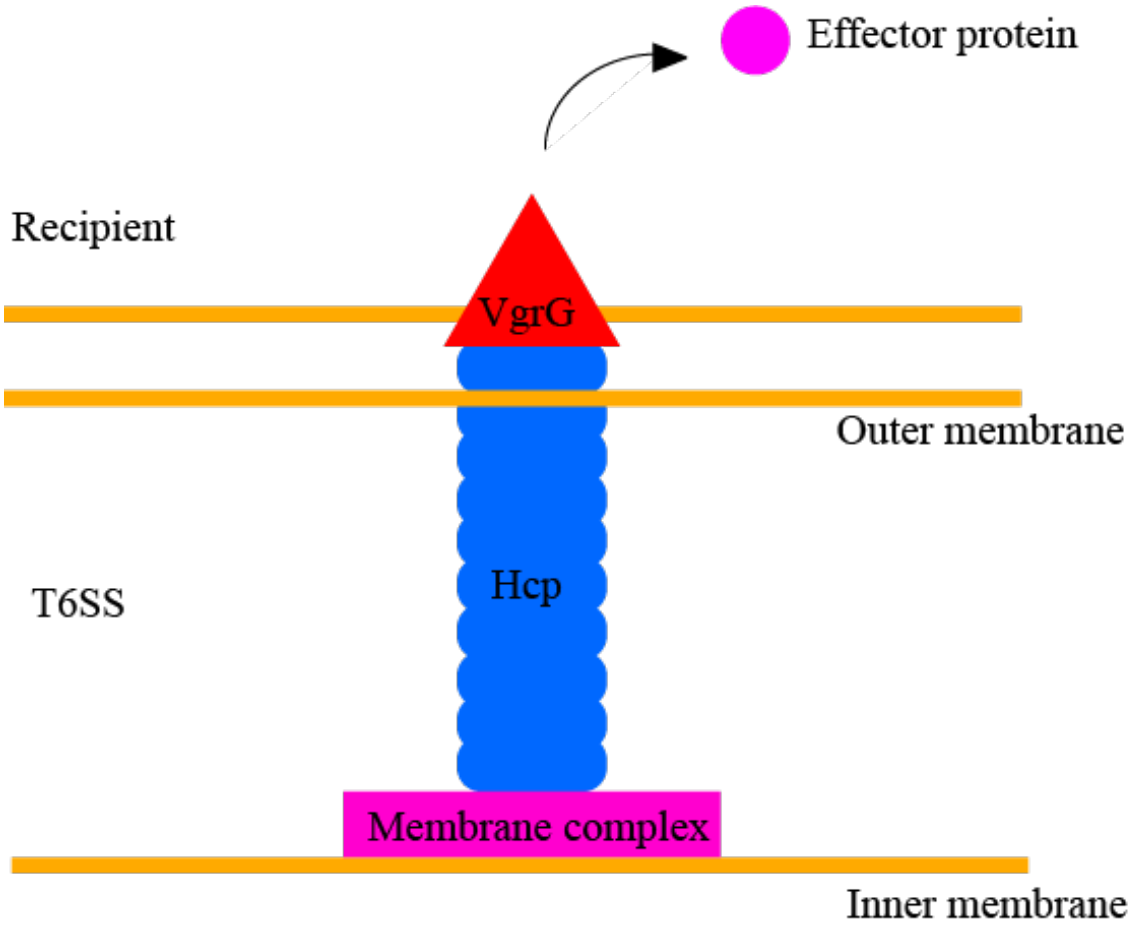


Figure 1: Schematic of a type 6 secretion system. Upon coming into contact with the target cell, the apparatus is assembled where the effectors, depending on their types, either associated with the Hcp hexamers or the VgrG protein which is capped by a PAAR spike forming a VgrG-PAAR protein complex. Thereafter, the sheath is contracted resulting in the release of the Hcp proteins and by extension the effector in the target cell (Klein et al 2020).

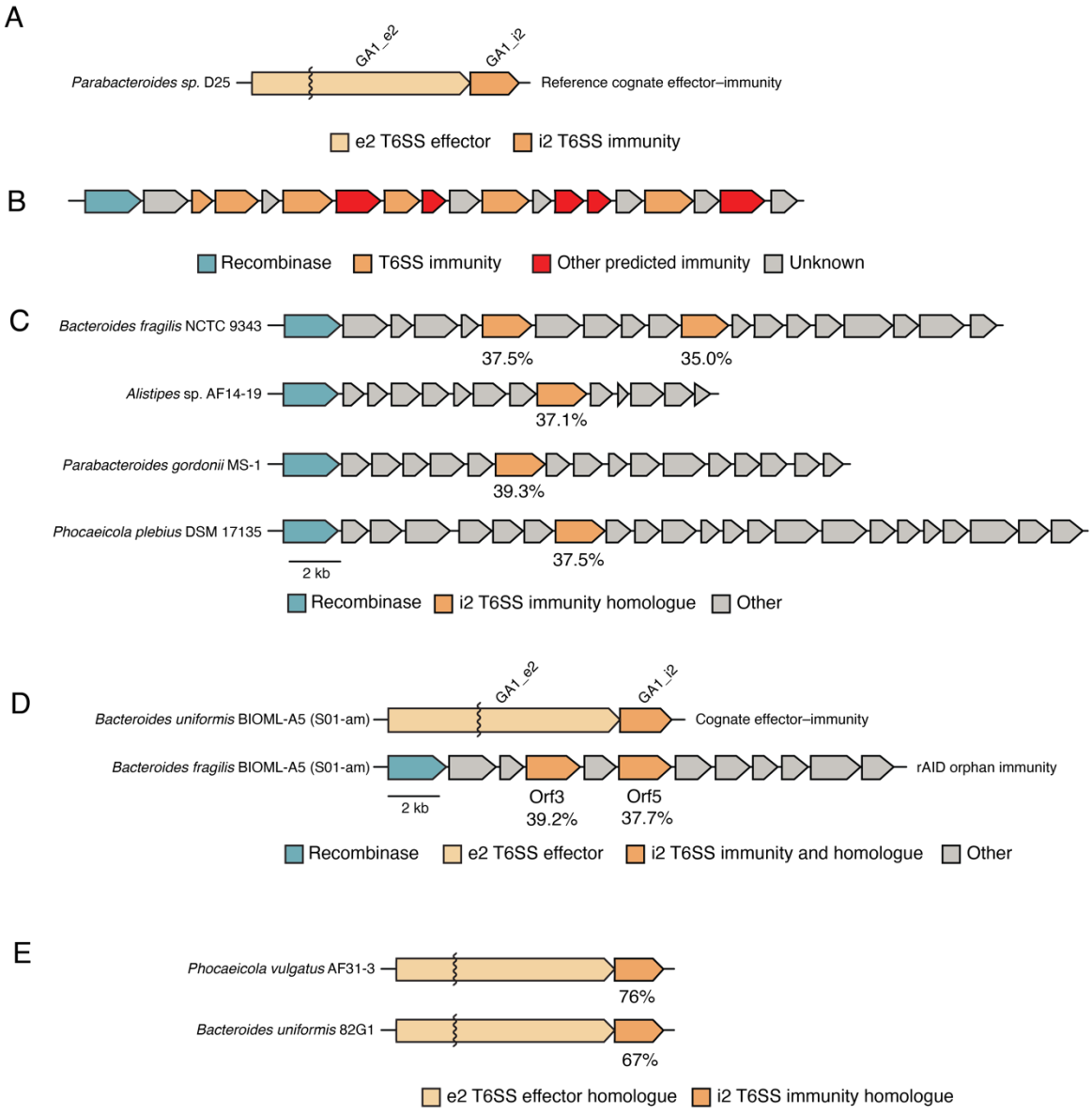


Figure 2: The genomic context of the effector and immunity loci referred to in the study. The zigzag lines indicate gaps due to an incomplete sequence assembly. A. The E2-I2 effector-immunity pair we came across initially. E2, contains a WHH nuclease domain towards its CTD and I2 consists of two SUKH domains. **B.** Schematic of a hypothetical rAID cluster. Each cluster starts with a recombinase gene followed by loci with different levels of sequence identity to cognate immunity genes of different bacterial

systems. Each locus is separated from the next by repetitive intergenic sequences (Ross et al. 2019). **C.** Some rAID clusters of different bacterial species found in the gut that contain homologues of I2. This includes the strain NCTC9343 of *B. fragilis*, one of the most widely used lab strains of *B. fragilis,* that contains two I2 homologues with identity levels of 37% and 35% to I2 respectively. The sequence identity levels of I2 homologues to I2 are written underneath each I2 homologue. **D.** The E2-I2 locus found in a recent study (Zhao et al. 2019) are in a strain of *B. uniformis*. The same human gut metagenomic samples include a *B. fragilis* strain whose genome contains a rAID cluster which consists of two I2 homologues, named Orf3 and Orf5, which are 39% and 37% identical to I2 respectively. The sequence identity levels of I2 homologues to I2 are written underneath each I2 homologue. **E.** Closer homologues of I2 are also found in gut samples in two strains of *P. vulgatus* and *B. uniformis* both encoding homologues of I2 with sequence identity levels of 76% and 67% to I2 respectively, as part of a putative T6SS effector-immunity pair. The sequence identity levels of I2 homologues to I2 are written underneath each I2 homologue.

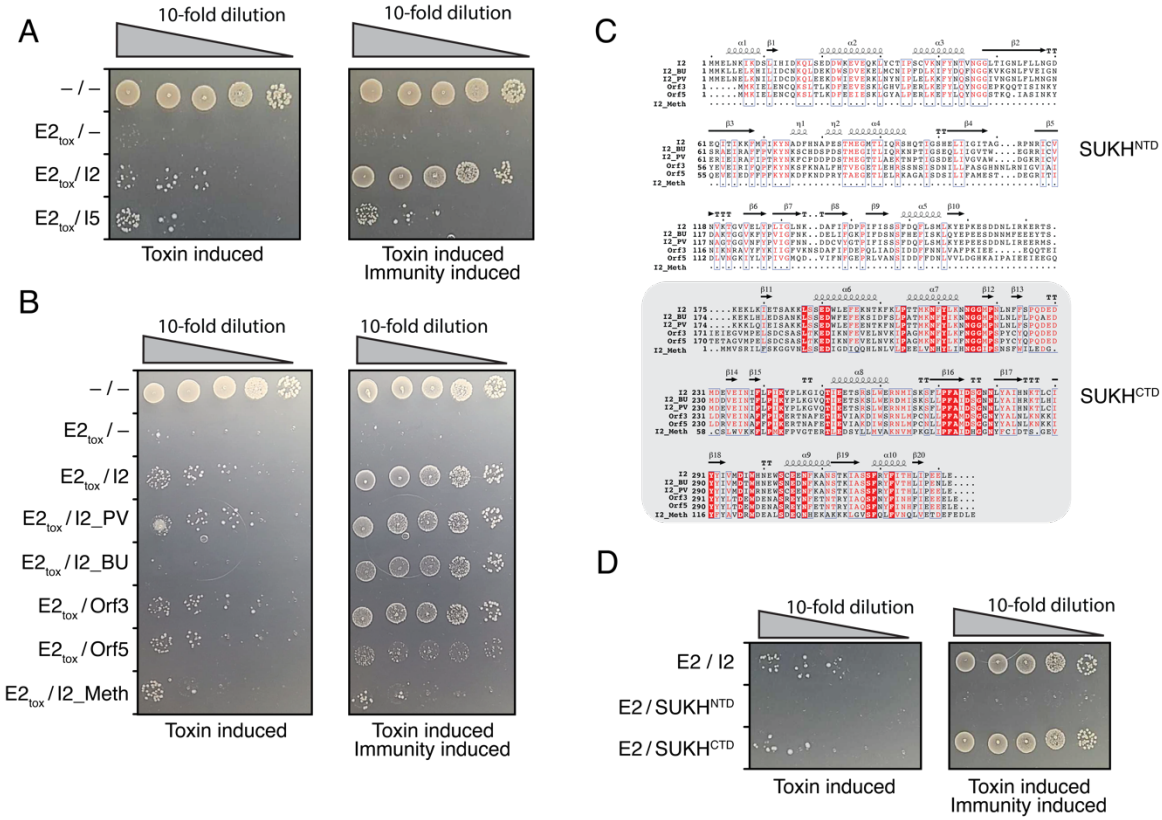


Figure 3: Toxicity assays to quantify the level of toxicity of the $E2_{tox}$ in the presence of different immunity proteins in *E. coli*. A. CFU (colony-forming units) plating of *E. coli* cells expressing the plasmids shown on the left side of the plates. Cells were plated in 1:10 dilutions. I2 is the cognate immunity of E2 and I5 is a cognate immunity gene found in the *Bacteroides fragilis* YCH46 strain, used here to assess whether the ability to protect against $E2_{tox}$-mediated toxicity is specific to I2. The negative sign refers to an empty plasmid **B.** CFU plating of *E. coli* cells expressing the plasmids shown on the left side of the plates. The assay was carried out in the same way as panel **A. C.** The sequence alignment of I2 and the rest of the I2 homologues. The alignment is separated into two sections corresponding to the two SUKH domains of I2. **D.** CFU plating of *E. coli* cells expressing the plasmids shown on the left side of the plates. $SUKH^{NTD}$ and $SUKH^{CTD}$ refers to the SUKH domains closer to the N-terminus and the C-terminus of I2 respectively.

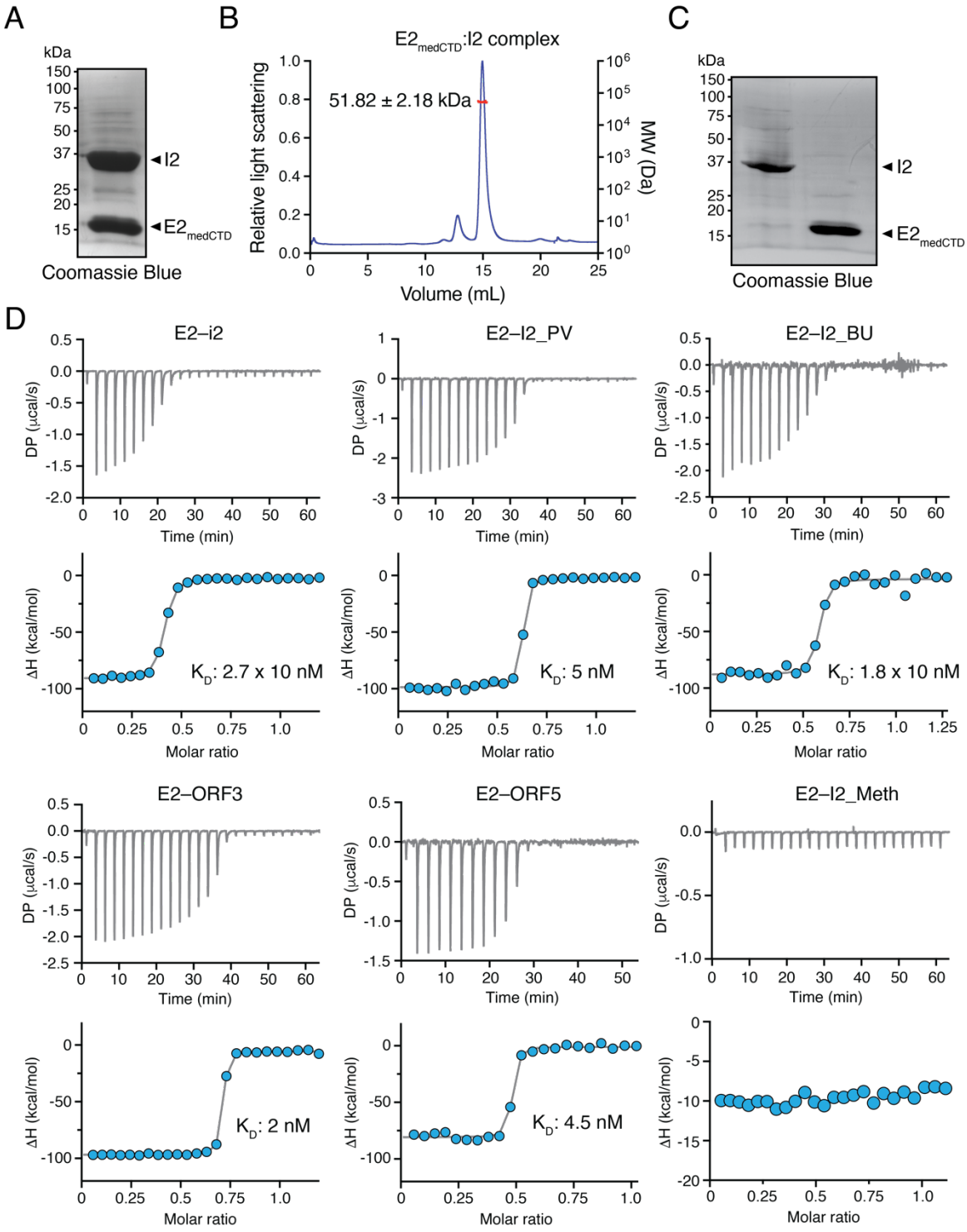

A
kDa
150
100
75
50
37
25
20
15
I2
E2medCTD
Coomassie Blue
B
E2medCTD:I2 complex
51.82 ± 2.18 kDa
Relative light scattering
MW (Da)
Volume (mL)
C
kDa
I2
E2medCTD
Coomassie Blue
D
E2–i2
E2–I2_PV
E2–I2_BU
E2–ORF3
E2–ORF5
E2–I2_Meth
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
Molar ratio
KD: 2.7 x 10 nM
KD: 5 nM
KD: 1.8 x 10 nM
KD: 2 nM
KD: 4.5 nM

Figure 4: I2 and its Bacteroidetes homologues directly bind $E2_{medCTD}$ A. The SDS-PAGE gel showing the result of the copurification of I2 and $E2_{medCTD}$ where $E2_{medCTD}$ is attached to a polyhistidine tag at its N-terminus. The molecular weight of I2 and $E2_{medCTD}$ are approximately 38.1 KDa and 14.6 KDa respectively. **B.** The SEC-MALS (size exclusion chromatography- multiangle light scattering) result from the copurification experiment shown in **A. C.** The SDS-PAGE gel from purification of the individual proteins from a sample effector-immunity pair used for the ITC experiments shown in **D.** (here, I2 and $E2_{medCTD}$) as described in the methods section. **E.** The result of the ITC experiments for each of the effector immunity pairs. For each pair, the upper graph demonstrates the heat released per unit time and the lower graph is the integration of that result. K_d stands for the dissociation constant calculated from the lower graph.

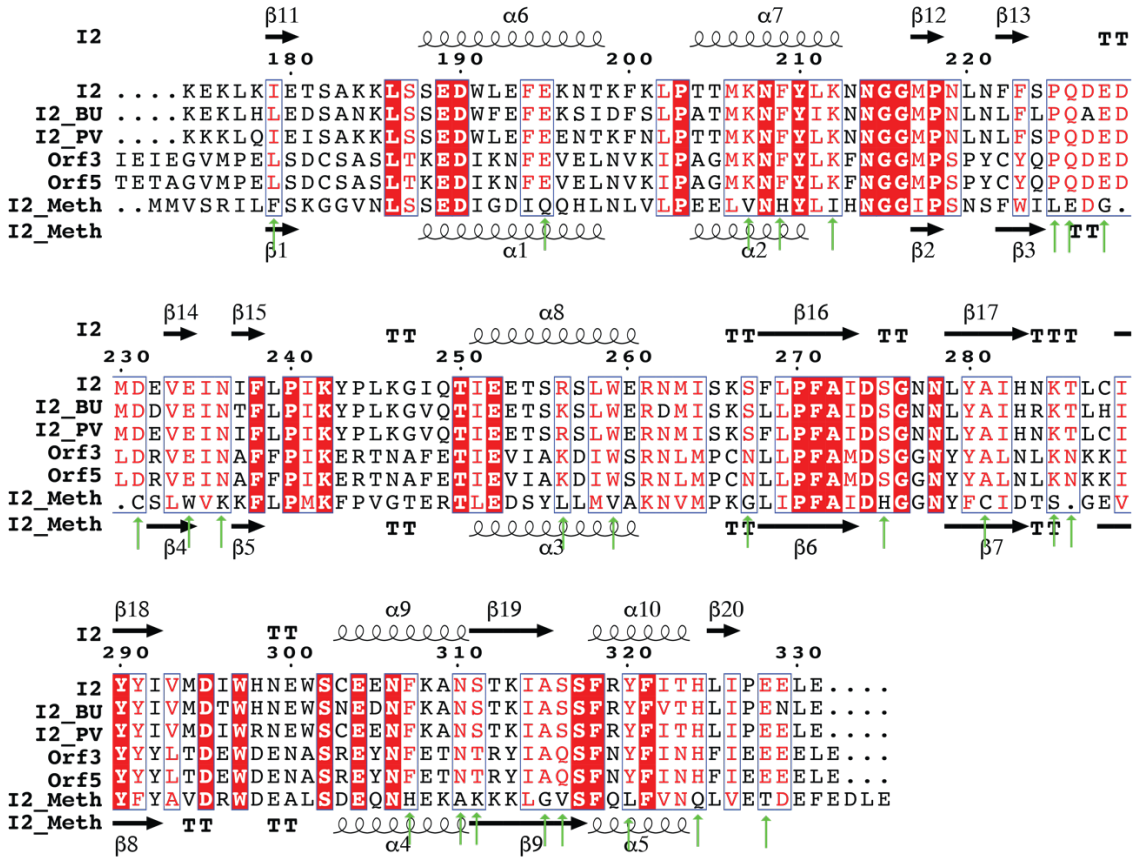
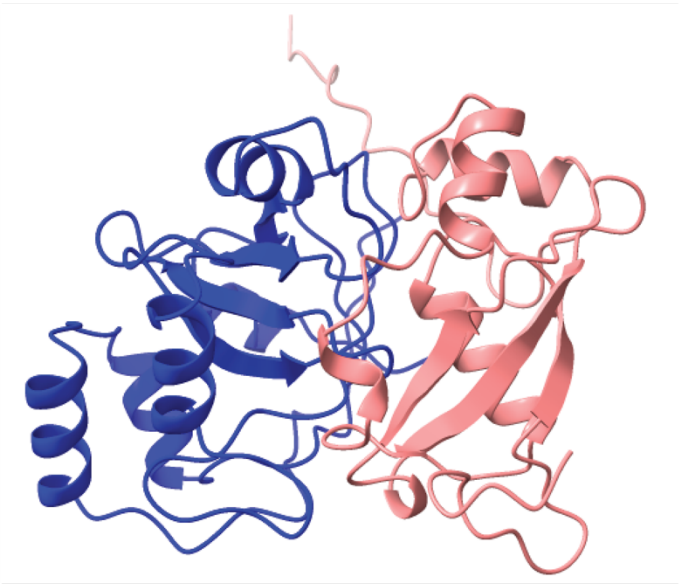


Figure 5: The sequence alignment of I2 and its homologues reveal 27 residues that are different in I2_Meth. The sequence alignment of I2 and its five homologues investigated in the study has been shown here. Only the part of the alignment is shown which corresponds to the SUKH domain closer to the C-terminus of I2 corresponding to the lower box in Figure 3C. The predicted secondary structure of the protein is derived from the AlphaFold predicted structure inputted into the software. The red boxes denote residues where all 6 proteins encode the same amino acid whereas the blue boxes denote where at least one protein encodes a different amino acid.

A



B

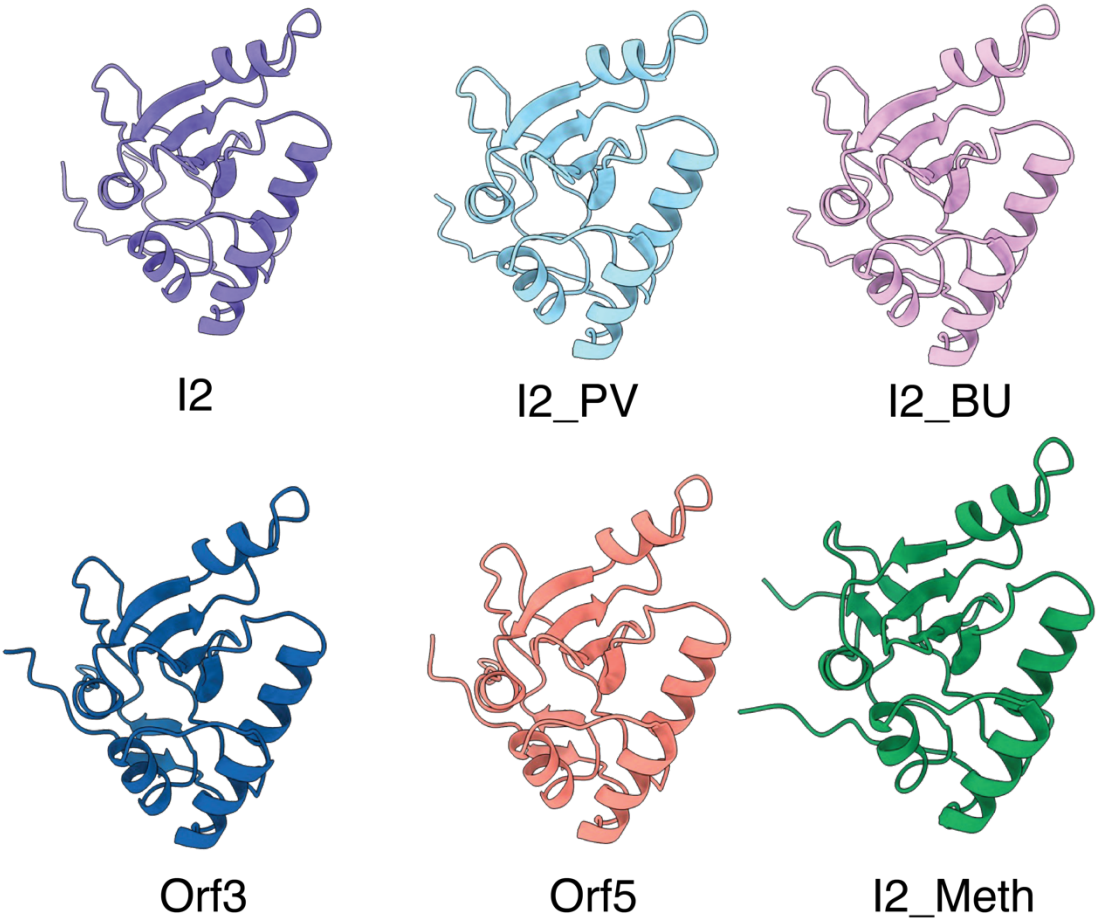


Figure 6: AlphaFold-derived models of protein tertiary structures suggest I2 interacts with $E2_{tox}$ and the five I2 homologues closely resemble the I2 structure. A. The structure of the I2- $E2_{tox}$ complex is shown where blue ribbon diagram corresponds to the predicted structure of $E2_{tox}$ and the red to that of I2. The I2 predicted structure shown corresponds only to the SUKH domain closer to the C-terminus of I2 corresponding to the lower box in Figure 3C. **B.** The AlphaFold-predicted structure of I2 and its 5 homologues investigated in this study. As with **A.**, all these structures are from that of the SUKH domain closer to the C-terminus of I2 corresponding to the lower box in figure 3c with the exception of the I2_Meth structure which is that of the entire protein.

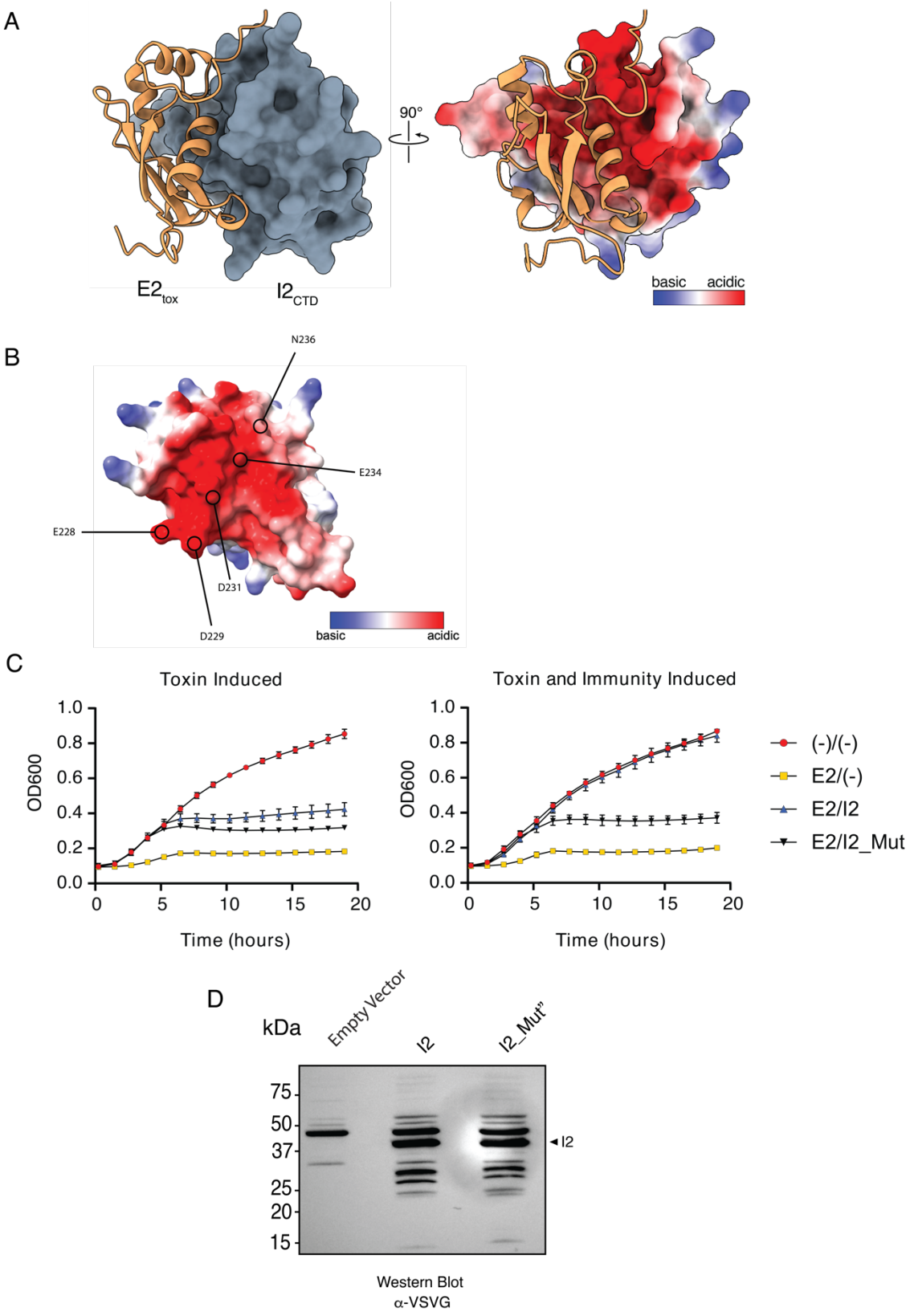

A
90°
basic
acidic
E2tox
I2CTD
B
N236
E234
E228
D231
D229
basic
acidic
C
Toxin Induced
Toxin and Immunity Induced
OD600
Time (hours)
(-)/(-)
E2/(-)
E2/I2
E2/I2_Mut
D
kDa
Empty Vector
I2
I2_Mut"
75
50
37
25
20
15
I2
Western Blot
α-VSVG

Figure 7: Mutagenesis experiments suggest the physical contact between $E2_{tox}$ and I2 is mediated through electrostatic interactions between the two. A. The ribbon diagrams and surface representations of the AlphaFold-predicted model of the protein complex and shows the electrostatic features of residues on $I2_{CTD}$ in the predicted binding site of the two proteins. **B.** The five residues for mutagenesis is shown on the predicted surface representation of $I2_{CTD}$ **C.** The overnight growth curve of the four strains in two conditions where toxin alone or toxin and immunity both were induced. I2_Mut refers to the quintuple mutant of I2 mentioned in the text. (E228R, D229R, E231R, E234R, N236K) **D.** Western blot assaying for the expression of the mutated version of I2 used in part **C.** Here, I2_Mut", the mutated variant of I2 also contains an additional E328R mutation close to its C-terminus in addition to the five mutations in I2_Mut.

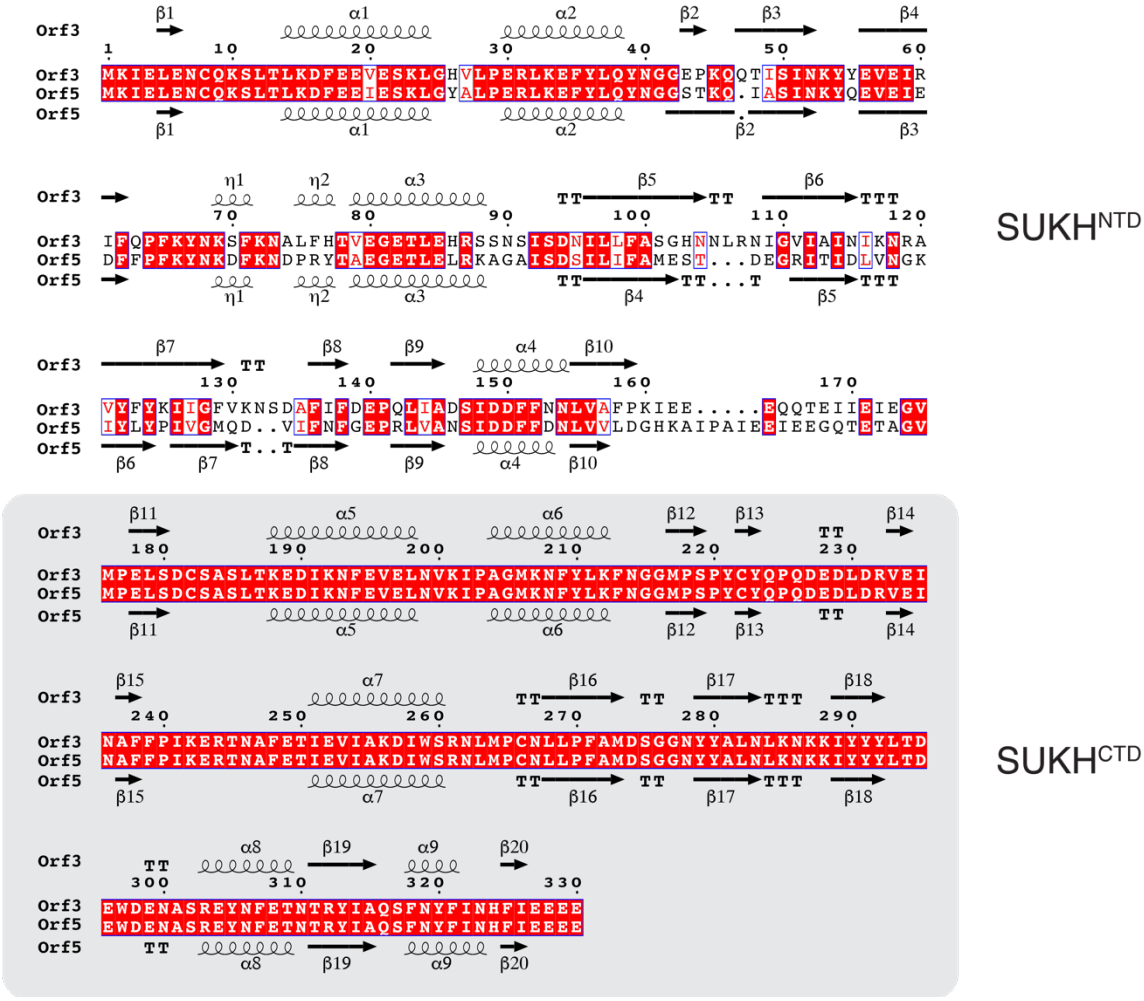


Figure 8: Protein sequence alignment of I2 homologues Orf3 and Orf5. The gray background represents the region of the two proteins sequences corresponding to the C-terminal SUKH domain.

SUPPLEMENTAL

Table S1: Oligonucleotides used in this study. Fwd and rev refers to the forward and reverse primers respectively.

Oligonucleotide	Sequence (5'- 3')	Restriction Enzyme
GA1_E2_longCT _fwd	TGTTAAGCTACATATGGGGCGCATTTTAAACCT GTATGG	NdeI
GA1_E2_rev	TCAATCAGTATCTAGAACTCGTTTTCTGGTAGCT ATTTTTTTTTC	XbaI
GA1_I2_fwd	TGTTAAGCTAGAGCTCACGGGAGGAAAGAT GGAGTTGAATAAGATTAAGGATTCTTTAATCC	NdeI
GA1_I2_rev	TCAATCAGTATCTAGATTCCTCTGGGATAAGGT GCG	XbaI
GA1_E2_Expression_ Short_fwd	TCAATCAGTAGAATTCGGGCATTTCCCGCAAGT CGGCG	EcoRI
GA1_E2_Expression_ Medium_ rev	TCAATCAGTAGAATTCGGGGAATCAGCAGAAC ACCG	EcoRI
GA1_E2_Expression_ Long_fwd	TCAATCAGTAGAATTCGGGGCGCATTTTAAACC TGTATGGCTACG	EcoRI
GA1_E2_Expression_ rev	TCAATCAGTAAAGCTTTTAACTCGTTTTCTGGTA GCTATTTTTTTTTTCTGACGG	HindIII
GA1_I2_Expression _fwd	TCAATCAGTACATATGATGGAGTTGAATAAGAT TAAGGATTCTTTAATCCATATCG	NdeI
GA1_I2_Expression _rev	TCAATCAGTACTCGAGTTATTCCTCTGGGATAA GGTGCGTAATAAAATAACG	XhoI
I2_PV_fwd	TGTTAAGCTAGAATTCACGGGAGGAAAGATGG AACTGAATGAACTGAAGAACATTCTGATCG	EcoI
I2_PV_rev	TCAATCAGTATCTAGATTCCTCCGGGATCAGAT GGG	XbaI
I2_BU _fwd	TGTTAAGCTAGAGCTCACGGGAGGAAAGATGA AGCTGCTCGAACTGAAGCACATCC	SacI
I2_BU_rev	TCAATCAGTATCTAGAGTTTTCCGGGATCAGAT GGG	XbaI
I2_PV_fwd	TGTTAAGCTACATATGGAACTGAATGAACTGAA GAAC	NdeI
I2_PV _rev	TCAATCAGTACTCGAGTTATTCCTCCGGG	XhoI
I2_BU_fwd	TGTTAAGCTACATATGAAGCTGCTCGAACTGAA GC	NdeI
I2_BU _rev	TCAATCAGTACTCGAGTTAGTTTTCCGGG	XhoI

GA1_I2_fwd_ $SUKH^{NTD}$	TGTTAAGCTAGAGCTCACGGGAGGAAAGATGC AACTGAGTGAAGACG	SacI
GA1_I2_rev_ $SUKH^{NTD}$	TCAATCAGTATCTAGACTTGGGTTCATATTTCA GC	XbaI
GA1_I2_fwd_ $SUKH^{CTD}$	TGTTAAGCTAGAGCTCACGGGAGGAAAGATGT CAGCAAAGAAATTGTCGAGCG	SacI
GA1_I2_rev_ $SUKH^{CTD}$	TCAATCAGTATCTAGATGGGATAAGGTGCGTAA TAAAATAACGG	XbaI
Orf3_fwd	TGTTAAGCTAGAGCTCACGGGAGGAAAGATGA AGATCGAGCTGG	SacI
Orf3_rev	TGTTAAGCTAGAGCTCACGGGAGGAAAGATGA AGATCGAGCTGG	XbaI
Orf5_fwd	TGTTAAGCTAGAGCTCACGGGAGGAAAGATGA AGATCGAGC	SacI
Orf5_rev	TCAATCAGTATCTAGATTCTTCTTCTTCG	XbaI
I2_Meth_fwd	TGTTAAGCTAGAGCTCACGGGAGGAAAGATGG TGTCAAGAATAC	SacI
I2_Meth_rev	TCAATCAGTATCTAGAATCTTCAAATTCG	XbaI
GA1_I2_Expression_ 161_fwd	TCAATCAGTACATATGGAATCAGACGACAACCT GATCCGC	NdeI
GA1_I2_Expression_ 175_fwd	TCAATCAGTACATATGGAAAAGCTGAAAATTGA GACATCAGC	NdeI
GA1_I2_Expression_ 182_fwd	TCAATCAGTACATATGTCAGCAAAGAAATTGTC GAGCGAGG	NdeI
E2_Meth_ longCT_ fwd	TGTTAAGCTACATATGAAAGGGGGGATAAACA TATATTGTTATCTTGTGAATCCG	NdeI
E2_Meth_ mediumCT_fwd	TGTTAAGCTACATATGACTGGGCAGCAAAGCAT TGTGGAAATAACG	NdeI
E2_Meth_ shortCT _fwd	TGTTAAGCTACATATGGGAATAAGCCCTGTAGA TGCCGAAGGG	NdeI
E2_Meth_XbaI_rev	TCAATCAGTATCTAGAACCACATTTAGGTTTGC GGCCC	XbaI
I2_Meth_Expression_ fwd	TCAATCAGTACATATGGTGTCAAGAATACTTTT CTCAAAGGGGGG	NdeI
I2_Meth_Expression_ rev	TCAATCAGTACTCGAGCTAATCTTCAAATTCGT CAGTTTCGACAAGCTGG	XhoI
E2_Meth_Expression longCT_ fwd	TCAATCAGTAGAATTCGGAATAAGCCCTGTAGA TGCCGAAGGG	EcoI
E2_Meth_Expression mediumCT_fwd	TCAATCAGTAGAATTCACTGGGCAGCAAAGCAT TGTGG	EcoI
E2_Meth_Expression shortCT _fwd	TCAATCAGTAGAATTCAAAGGGGGGATAAACA TATATTGTTATCTTGTGAATCC	EcoI
E2_Meth_Expression rev	TCAATCAGTAAAGCTTTTAACCACATTTAGGTT TGCGGCCC	HindIII

GA1_I2_fwd_ITC	TCAATCAGTACATATGATGGAGTTGAATAAGATTAAGGATTCTTTAATCCATATCG	NdeI
GA1_I2_rev_ITC	TCAATCAGTACTCGAGTTCCTCTGGGATAAGGTGCGTAATAAAATAACGG	XhoI
I2_PV_fwd_ITC	TCAATCAGTACATATGATGGAACTGAATGAACTGAAGAACATTCTGATCGACTGCG	NdeI
I2_PV_rev_ITC	TCAATCAGTACTCGAGTTCCTCCGGGATCAGATGGGTGATGAAGTAGCGG	XhoI
I2_BU_fwd_ITC	TCAATCAGTACATATGATGAAGCTGCTCGAACTGAAGCACATCCTCATCG	NdeI
I2_BU_rev_ITC	TCAATCAGTACTCGAGGTTTTCCGGGATCAGATGGGTAACG	XhoI
Orf3_fwd_ITC	TCAATCAGTACATATGATGAAGATCGAGCTGGAGAACTGCCAGAAGAGTCTGACACTGAAAGATTTTGAGG	NdeI
Orf3_rev_ITC	TCAATCAGTACTCGAGTTCTTCTTCCTCAATAAAATGATTAATGAAGTAGTTAAAAGACTGGGCG	XhoI
Orf5_fwd_ITC	TCAATCAGTACATATGATGAAGATCGAGCTGGAGAACTGCCAGAAGAGTCTGACACTGAAAGATTTTGAGG	NdeI
Orf5_rev_ITC	TCAATCAGTACTCGAGTTCTTCTTCTTCGATGAAATGGTTAATGAAGTAGTTGAAACTCTGGGCG	XhoI
I2_Meth_fwd_ITC	TCAATCAGTACATATGATGGTGTCAAGAATACTTTTCTCAAAGGGGGG	NdeI
I2_Meth_rev_ITC	TCAATCAGTACTCGAGATCTTCAAATTCGTCAGTTTCGACAAGCTGG	XhoI
GA1_I2_D231R_fwd	AATTTCTTCTCACCGCAAGATGAAGATATGAGAGAGGTTGAAATTAACATCTTCTTGCCG	N/A
GA1_I2_D231R_rev	AATTTCAACCTCTCTCATATCTTCATCTTGCGGTGAGAAGAAATTCAAGTTGGG	N/A
GA1_I2_E228R_fwd	TTCTCACCGCAAGATAGGGATATGGATGAGGTTGAAATTAACATCTTCTTGCCG	N/A
GA1_I2_E228R_rev	AACCTCATCCATATCCCTATCTTGCGGTGAGAAGAAATTCAAGTTGGGCATGCCGCCG	N/A
GA1_I2_E234R_N236K_fwd	ATGGATGAGGTTAGGATTAAGATCTTCTTGCCGATCAAGTATCCGC	N/A
GA1_I2_E234R_N236K_rev	CGGCAAGAAGATCTTAATCCTAACCTCATCCATATCTTCATCTTGCGG	N/A
GA1_I2_E234R_N236K_E228R_D229R_DD231R_fwd	AGGAGGATGAGGGAGGTTAGGATTAAGATCTTCTTGCCGATCAAGTATCCGC	N/A

GA1_I2_E234R_N236K_E228R_D229R_DD231R_rev	CCTCATCCTCCTATCTTGCGGTGAGAAGAAATTCAAGTTGGGC	N/A
GA1_I2_E328R	TCAATCAGTATCTAGATTCCCTTGGGATAAGGTGCGTAATAAAATAACGG	XbaI

Table S2: Plasmids Used in this study

Plasmids	Description	Source
pSCrhaB-CV		Cardona & Valvano 2005
pSCrhaB2::GA1_E2_193_CTD	contains the toxic E2 construct starting at the 192th residues	This study
pPSV39-CV		(Rietsch et al., 2005)
pPSV39-CV::GA1_I2	contains *GA1_I2*	This study
pETDuet-1::GA1_E2_193_CTD::GA1_I2	pETDuet vector containing the long E2 construct in MCS1 and GA1_I2 in MCS2	This study
pETDuet-1::GA1_E2_229_CTD::GA1_I2	pETDuet vector containing the medium E2 construct in MCS1 and GA1_I2 in MCS2	This study
pETDuet-1::GA1_E2_255_CTD::GA1_I2	pETDuet vector containing the short E2 construct in MCS1 and GA1_I2 in MCS2	This study
pPSV39-CV::I2_PV	Contains I2_PV	This study
pPSV39-CV::I2_BU	Contains I2_BU	This study
pETDuet-1::GA1_E2_255_CTD::I2_PV	pETDuet vector containing the medium E2 construct in MCS1 and I2_PV in MCS2	This study
pETDuet-1::GA1_E2_255_CTD::I2_BU	pETDuet vector containing the medium E2 construct in MCS1 and I2_PV in MCS2	This study
pPSV39-CV::GA1_I2_SUKH1_46_480	Contains the N-terminal SUKH domain of I2	This study
pPSV39-CV::GA1_I2_SUKH2_544_978	Contains the C-terminal SUKH domain of I2	This study
pPSV39-CV::Orf3	Contains Orf3	This study
pPSV39-CV::Orf5	Contains Orf5	This study
pPSV39-CV::I2_Meth	Contains I2_Meth	This study
pETDuet-1::GA1_E2_229_CTD::GA1_I2_161_CTD	pETDuet vector containing the medium E2 construct in MCS1 and GA1_I2 from the 162th residue to the C-terminus in MCS2	This study

pETDuet-1::GA1_E2_229_CTD::GA1_I2_175_CTD	pETDuet vector containing the medium E2 construct in MCS1 and GA1_I2 from the 175th residue to the C-terminus in MCS2	This study
pETDuet-1::GA1_E2_229_CTD::GA1_I2_182_CTD	pETDuet vector containing the medium E2 construct in MCS1 and GA1_I2 from the 182th residue to the C-terminus in MCS2	This study
pSCrhaB2::E2_Meth_1327_CTD	Contains the long E2_Meth construct starting at the 1327th residue of E2_Meth	This study
pETDuet-1::LW13_Eff_1327_CTD::LW13_Imm	pETDuet vector containing the Long E2_meth construct in MCS1 and I2_Meth in MCS2	This study
pETDuet-1::LW13_Eff_1367_CTD::LW13_Imm	pETDuet vector containing the medium E2_meth construct in MCS1 and I2_Meth in MCS2	This study
pETDuet-1::LW13_Eff_1393_CTD::LW13_Imm	pETDuet vector containing the Short E2_meth construct in MCS1 and I2_Meth in MCS2	This study
pET29b::GA1_I2	Contains GA1_I2	This study
pET29b::I2_PV	Contains I2_PV	This study
pET29b::I2_BU	Contains I2_BU	This study
pET29b::Orf3	Contains Orf3	This study
pET29b::Orf5	Contains Orf5	This study
pET29b::I2_Meth	Contains I2_Meth	This study
pPSV39-CV::GA1_I5	Contains GA1_I5	This study
pPSV39-CV::GA1_I2_E228R_D229R_D231R_E234R_N236K	Contains GA1_I2 with mutations at the specified residues	This study
pPSV39-CV::GA1_I2_E228R_D229R_D231R_E234R_N236K_E328R	Contains GA1_I2 with mutations at the specified residues	This study