Development of an AMP-SECreting platform in E. coli for simpler AMP development (AMPSEC).

By

Kyle Tomaro

Masters' thesis submitted to the Faculty of Medicine Department of Biochemistry, Microbiology and Immunology in partial fulfillment of the requirement of the MSc. Degree

Supervisor: François-Xavier Campbell-Valois
Co-Supervisor: Mathieu Lavallée-Adam

Department of Biochemistry, Microbiology and Immunology
University of Ottawa
Ottawa, Canada

Abstract:

In the global fight against antibiotic resistance, the need for alternatives is more pressing than ever. Antimicrobial peptides (AMPs), short oligopeptides usually produced as part of the immune system of a host, have shown great promise against resistant bacteria, biofilms and even cancer cells. Engineering AMPs that are both specific to a set of bacteria and stable is among the main challenges of the field. Herein, we propose two separate tools to support these efforts.

The first one is an AMP SECretion system based in E. coli, dubbed AMPSEC, that can be used to produce active AMPs with specific targets (i.e., gram-positive bacteria or any specific specie)**.** This recombinant protein system uses surface display technologies coupled with specific protease activity to express, export, and release functional AMPs that could readily affect neighbouring target bacteria. The AMPSEC would be ideal to screen AMP libraries, removing the need for purification or chemical synthesis in order to observe toxicity. It could also be used for AMP production, where the secreted AMPs would be purified from the growth medium by HPLC. Finally, if the recombinant system is inserted in a probiotic host, it might even be useful to deliver AMPs in the gut to treat dysbiosis. Herein, we explored six surface display apparatuses for their applicability for AMPSEC and found three out of the six being fully functional in transporting the cargo although the cleaving activity needs to be coordinated better with its localization at the outer membrane. A robust proof-of-concept workflow has also been established and used to evaluate the performance of those six display apparatuses.

The second is a bioinformatics approach to highlighting the relationship between the primary structure and the microbial target specificity of AMPs. Our method first clusters

sequences from the DRAMP AMP repository using the Linclust algorithm. *De novo* motif discovery tools can then extract AMP sequence motifs relating to target specificity. These motifs could guide randomized sequence AMP library creation and decrease the number of inactive sequences generated. Clustering AMPs, however, proved to be rather challenging due to the large sequence length variation in the databases, the small sample size and their overall short lengths. It would then be necessary to design an algorithm suited to handle this specific kind of proteomics dataset. A library eventually created using the discovered motifs could then be used with AMPSEC. Combined, these two tools will further improve our ability to design stable AMPs targeting specific bacteria.

Acknowledgements:

I would like to thank Dr. François-Xavier Campbell-Valois for the amazing opportunity I have researching as part of his team and Dr. Mathieu Lavallée-Adam for taking me in as well and permitting me to learn a completely new discipline in bioinformatics. I'm thankful for the amazing support and work dynamic I have with both PIs and I am looking forward to contributing more work to their labs. I would also like to thank all the members from both laboratories for all their assistance with special thanks to Rachel for guiding me through using various bioinformatics resources.

Table of content:

List of abbreviations:

AMP	Antimicrobial peptide
AT	Autotransporter
IM	Inner membrane
INP	Ice nucleation protein
MB	Membrane
MSA	Multiple sequence alignment
OM	Outer membrane
OMP	Outer membrane protein
PSA	Pairwise sequence alignment
SDA	Surface display apparatus
SP	Signal peptide
TCA	Trichloroacetic acid
TMD	Transmembrane domain
WB	Western blot

List of tables:

List of figures:

Statement of contribution:

Protocols were designed by Kyle Tomaro, Dr. Campbell-Valois and Dr. Lavallée-Adam. Laboratory work, and implementation and execution of the computational methods were done by Kyle Tomaro. All figures were designed and drawn by Kyle Tomaro unless cited.

1 Introduction

1.1 Part I: Importance

1.1.1 The antimicrobial resistance crisis

Antimicrobial resistance is a consistent issue currently listed amongst the top 10 threats to global health[1,2]. In 2019, it was determined that at least 700,000 people worldwide die from drug-resistant bacteria each year. This number is predicted to skyrocket to 10 million deceased per year by 2050, which would significantly impact the world economy along with global health[3]. Alternative therapies are desperately needed. There are efforts on the discovery of new chemicals with different bacterial targets and belonging to novel chemical classes[4]. Phage therapy has also come a long way and shows great potential although it faces hurdles on its way to being approved for clinical use[5]. Modulating the microbiome using fecal transplant[6], probiotics[7] or antagonistic bacteria[8] is yet another emerging alternative. Herein, we explore the use of antimicrobial peptides (AMPs) as a potential contender in combating antimicrobial resistance.

1.1.2 The properties of antimicrobial peptides

AMPs are composed of about 5-100 residues and act against a wide range of microbes[9–11]. In nature, these are found in species from all kingdoms of life including humans[12–14]. They function as part of the innate immune system with their broad-spectrum activities against various invasive pathogens[15]. AMPs also encompass nonribosomal peptides from bacteria which, in contrast to ribosomal peptides, are not genetically encoded, but are instead generally created by large enzyme complexes known as nonribosomal peptide synthetases[16,17]. Moreover, thionins, defensins, maganins, temporins, cecropins and cathelicidins are also considered AMPs[18,19]. Their potential for clinical applications was

demonstrated by their efficacy against antibiotic resistant bacteria[20], biofilms[21–24] and even cancer cells[25,26]. These peptides traditionally function by disrupting the lipid bilayer[27–32] of their targets, however, an ever-increasing amount of AMPs show more complex mechanisms[26,29,33–37]. Indeed, it was believed that, in order to be effective, AMPs needed to be strongly positively charged and hydrophobic like most of the ones found in nature[26,28]. However, recent research efforts in peptide engineering demonstrated the emergence of a new family of effective AMPs that are neither particularly hydrophobic nor cationic[33]. These new synthetic AMPs diverge strongly from natural AMPs but bring clinical potential for the use of this antibiotic alternative. In general, AMPs have been divided into four groups according to their secondary structures; the α-helix and β-sheet classes are the main ones, but there is a large quantity of AMPs that belong to neither. These belong to the extended class, which is made of intrinsically disordered peptides that often possess an over-representation of certain amino acids (proline, arginine, cysteine, glycine and/or aromatic amino acids)[38,39]. AMPs can form their active secondary structures spontaneously or upon interaction with their target[40]. Some need post-translational modifications in order to function properly such as methylation[41], D-amino acids addition[42,43], phosphorylation[44], amidation[45], disulphide bond formation[46], glycosylation[47], cyclisation[17] and cleaving[48]. This is often the case for nonribosomal peptides whose synthetases can have the ability to include these features. However attractive the use of such convenient machinery would be, our proposed AMPSEC system depends on the recombinant construction of an AMP-secretion system fusion, which demands for genetic encoding. We chose to prioritize the simplicity in designing and ease of library creation over the chemical diversity and stability of nonribosomal peptides.

Antibiotics are small compounds extracted from natural sources or synthetically derived from them[49]. Unlike these conventional antibiotics, AMPs are short amino acid chains that makes them relatively easy to modify and engineer. They can be chemically synthesized as a more expensive option[50], or produced in a recombinant expression system[51,52], which is cheaper but can have limitations when using unconventional amino acids. As for the designing of the AMP sequence, unfortunately, the relationship between the activity and structure of AMPs is poorly understood. In fact, peptides showing similar structures can display widely different killing mechanisms[53,54]. In contrast, properties were identified whose modulation produces somewhat predictable changes in the activity of a given AMP; it was determined that modifications of the net charge[55], the helicity[30], the hydrophobicity[30,56], the C-terminal amino acid[57–59], and modifications like 'end-capping'[60], cyclisation[61–64] or the use of unnatural amino acids[65–67] can be used to tune the activity of an AMP. In the end, they are still far from being a commonly administered pharmaceutical due to their instability in internal environments comprising of certain pH and salt conditions and of proteases. Their hemolytic properties as well as their high cost of production are also major concerns for their therapeutic use[68]. Finally, their novel yet generic target being the lipid bilayer, some AMPs can target pathogens of a specific gram type but most have broad-spectrum effects.

1.1.3 The current state of AMPs in medicine

As of 2020, 6 out of the 10 commercially available AMP medications are for skin infections such as bacitracin and gramicidin D, while the rest, such as colistin and teicoplanin, are explicitly used as a last resort for serious gram-positive or multi-drug resistant gram-negative infections where antibiotics have failed[69]. Those 10 AMPs were also

either directly extracted from bacteria or are derivatives of each other. As their origin suggest, these specific AMPs come from nonribosomal pathways. Moreover, most of the AMPs currently in clinical trials are being developed for topical administration[69]. That said, engineering AMPs that are both specific and stable is the most difficult challenge of the field. To support these efforts, an increasing number of laboratories are working on the development of AMP research tools (Table 1).

Table 1. An overview of recent AMP research technologies. Note that the following techniques do not include methods necessitating chemical synthesis of AMPs. The list includes methods developed in the last decade (2012-2022).

Technology	Brief functionality	Purpose	Host
***lpo* knockout**[70]	Periplasmic-expressed AMP and leaky outer membrane	Production, Screening	*E. coli*
SLAY[33]	Surface displayed AMPs acting directly against the host in a self-screening tactic	Screening	Various
CL7 tag[71]	Traditional recombinant expression and purification of AMPs	Production	*E. coli, P. pastoris*
iLOV reporter[72]	Finding high AMP expression clones using an expression level fluorescent reporter	Production	*P. pastoris*
Growth assay[73]	Expressed AMPs affect the growth of a host in high-throughput self-screening parallel growth assays	Screening	*E. coli*
Double emulsion droplets[74]	Cell-free expression of AMPs in droplets containing bacterial and mammalian liposomes for screening membrane-targeting AMPs	Production, Screening	*Cell-free*
Expression screening platform[75]	Traditional recombinant expression and purification of AMPs using ELISA to detect high expression levels	Production	*E. coli, P. pastoris*

TEV-mediated secretion in gram-positive bacteria[76]	Two-cistron secretion system encoding an AMP and TEV where the release of TEV mediates the cleavage of the AMP	Production	*B. subtilis*
Secretion in Yeast[77,78]	Sec-mediated AMP secretion induced by methanol in yeast	Production	*P. pastoris*
Npro fusion[79]	Recombinant expression in inclusion bodies. Release of active AMP by self-cleaving activity after purification	Production	*E. coli*

Notably, a *lpo* knockout *E. coli* was designed for AMP screening[70]. This genetic knockout causes defects in the integrity of the outer membrane (OM), which lets the periplasmic contents leak out. This is combined with a signal peptide-AMP fusion to export the AMP to the periplasm, which is then passively secreted by the leaky membrane. Another study used surface display technology to express AMPs in the target host and have it exposed at the OM surface attached to a flexible tether. This library screening technique, dubbed SLAY, depends on the autotoxicity of the AMP against the host and identifies active sequences by finding the depleted clones using next generation sequencing[33]. Alternatively, a gram-positive system allows for the secretion of AMPs by co-secreting TEV protease[76]. The AMP is expressed as a fusion with a signal peptide in N-terminal and TEV in C-terminal to prevent toxicity against the host. After secretion, TEV is detached from the C-terminal and cleaves off the N-terminal to release the AMP.

Herein, we aim to contribute two novel tools that will simplify the AMP engineering process. As our first tool, we aimed to develop a secretion system that could be used in order to produce functional AMPs (AMPSEC)**.** This recombinant protein system based in *E. coli*

couples surface display technologies with reliable high-specificity protease activity to express, export, and release functional antimicrobials which could readily act on neighbouring targets (Fig.1). This tool would be a great screening method for AMP libraries, removing the need for purification or chemical synthesis in order to observe toxicity. However, as mentioned above, some AMP screening methods have already been established. Where their system was built for a very specific application, in contrast, the one presented here proves more versatile. Both the proposed secretion system, the gram-positive system and the *lpo* knockout system could also be used for AMP production. The secreted AMPs could be purified from the growth medium by HPLC in contrast to the SLAY system that can only serve for screening. Moreover, as an advantage over the gram-positive and *lpo* systems, *in vivo* screening assay could be possible since the AMPSEC system aims to not only maintain the integrity of the OM of the host bacterium, but also aims to function independently. The *lpo* system compromises the integrity of the OM, while the gram-positive secretion system using TEV relies on stepwise manual induction of the various functional steps of their system using glucose and IPTG supplementation to the media. In fact, if the proposed recombinant system is inserted in a probiotic host, it could be useful as a form of AMP delivery system to, for example, treat various types of dysbiosis.

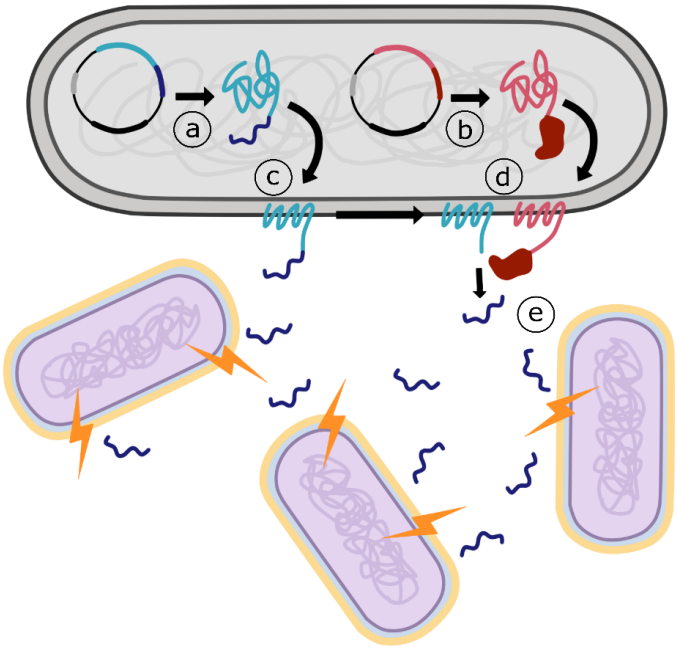


Figure 1. AMPSEC System. a) Expression of AMP surface display recombinant. **b)** Expression of protease surface display recombinant. **c)** AMP surface display. **d)** AMP release by surface expressed proteases. **e)** AMP toxicity towards the target.

Even with such promising screening tools, using random libraries for AMP discovery is not the most efficient protocol. Unfortunately, as mentioned, structure-function relationships in AMPs remains mostly a mystery[68]. That said, using the data from an AMP database, we hope to use sequence clustering approaches to discover novel peptide sequence motifs associated with specific targets. The discovery of AMP motifs associated with a specific target group would guide randomized sequence AMP library creation and lower the ratio of inactive sequences generated. Using such motifs would also reduce the number of AMPs generated that show toxicity against the screening host as opposed to the desired target bacterium. Subsequently, a peptide library relevant to this project could be generated and serve as a first application of the proposed AMPSEC.

1.2 Part II: AMPSEC

1.2.1 Surface display technologies

Surface display technologies have been used countless times to study protein activity. They have had numerous applications in vaccine development, catalytic or enzymatic studies, whole-cell biocatalysis and in various types of library screening[80]. One of the advantages of surface display apparatuses (SDAs) over normal protein expression is how extraction and purification steps become unnecessary when studying the function of a protein[81]. Although SDAs exist for yeast[82] and gram-positive bacteria[83], our project focuses on the ones designed for use in gram-negative species[84]. For the purpose of this study, these can be separated into outer membrane proteins (OMPs) and autotransporters (ATs). A third category of SDAs are the surface appendages such as fimbriae and flagellum, but since these necessitate insertional cargos these were not considered in this study[85]. As releasing the cargo is part of the functionality of AMPSEC, we need said cargo to be terminally attached to the SDA such that it only needs cleaving at either the C-terminal or N-terminal. OMP display systems are based on the natural localization of the protein to the outer membrane (OM) of the bacterium. A cargo can be placed at either the N or C terminal of these proteins given that the chosen end of the carrier OMP would face the exterior surface of the cell once folded inside the outer membrane. In contrast, ATs make up the type V secretion system (TVSS) of gram-negative bacteria. They work by inserting their C-terminal transmembrane domain (TMD) into the OM, while simultaneously exporting their N-terminally attached cargo to the extracellular surface with the help of the barrel assembly machine (Bam) complex, which acts as the catalyst for the insertion of the β-barrel TMD[86,87].

For AT-based display systems, the properties of the cargo as well as the flexibility and length of the linking amino acids connecting it to the TMD all play major roles in whether the system will be successful. Some properties of the cargo like its size, and the presence of oxidizable cysteine residues can affect whether the AT is able to translocate it effectively. Although cargo preference is often a case-by-case issue with ATs, a bigger size and the pre-translocation folding of the cargo, especially with disulfide bonds, were all shown to impair function[80,88,89]. It is recognized that one of the most important parts of the translocating unit of ATs is the linker sequence[80]. Although the minimal length of this varies from AT to AT, cargos generally benefit from a more generous and flexible length of linker[90–92]. Another important feature to consider is the promoter used for the expression of the system. Whereas ATs are naturally expressed in low levels, using a more powerful promoter is often desired to achieve a maximum of exposed proteins of interest. Unfortunately, an overexpression of display systems can be toxic to the host cell due to the failure of compartmentalization of the hydrophobic protein, making the system lose its function from degradation and inclusion body formation[93–95]. Finally, the signal peptides (SPs) used to target the protein to the periplasm can be another way to modulate the effectiveness of a system. Although no SP has been shown to cause issues, some systems have been optimized using non-native SPs[80,96,97]. With the emergence of the field, some papers were seen engineering optimized SP for this specific application[98,99]. A paper demonstrating the use of the MATE AT system, showed a boost in efficiency when using a SP native to the host bacterium in contrast to the original SP in the original design of the construct[99]. Although this being the only evidence of this SP-host compatibility, it is possible that the optimization of the SP may improve the targeting to the secretion machinery of the host cell more than it contributes to the function of the AT system.

When using OMPs as scaffolds for surface display, the same challenges apply as for the ATs. However, while switching the passenger domain of an AT is often the only step required to create a functional display system, OMPs are not carriers. The transmembrane domains of these proteins need to be optimized to accept a protein cargo either on one of its terminal ends or as an insertion within its sequence.

Such systems can form the foundation on which AMPSEC could be built. There are many which could have potential in that regard and the seven most promising SDAs were considered in the study (Table 2). Firstly, the Autotransporter Involved in Diffuse Adherence (AIDA-I) of *E. coli* is one example of a successfully used AT for practically all known applications of SDAs[80]. In *E. coli*, in more recent applications, it has been used for biocatalysis experiments[100–102], drug and vaccine development including antibody design[103–106]. In the same category, we find the EspP AT, also from *E. coli*. This autotransporter has been used a very limited number of times[97,107] but it promised being more efficient than its well-known counterpart AIDA[97]. We can also consider the MATE AT, an optimized version of the EhaA AT[96]. This system is relatively new and has not been used in many papers[99,108] but, as with EspP, it promised greater usability and efficiency compared to AIDA[96].

As for OMP-based systems, the most well-used system is the LPP-OmpA surface display apparatus, where the membrane domain of OmpA from *E. coli* was fused to a lipoprotein anchor[109]. This system takes a C-terminal cargo and has been used most recently for the display of enzymes[110–112], binding proteins[113] and virulence factors[114], and for the SLAY antimicrobial peptide screening method mentioned earlier[33]. Although this system is always successful in displaying its cargo, its efficiency is very poor and often harms the viability of the host cell[115–118]. Another highly optimized but rarely used system is the eCPX

system. This one stems from the OmpX protein from *E. coli*. This system was distinctively used to engineer living nanoparticle hybrids by displaying inorganic compound-binding peptides to the surface of *E. coli*[119,120]. It has also been used for establishing intercellular supramolecular interactions[121] and screening of affibody molecules[122]. Finally, we have the lipoproteins from *P. syringae* known as ice nucleation proteins (INPs). They have been used in whole cell biocatalysis[115,123,124] and a biochemical study of displayed proteins[125]. These small outer membrane proteins have shown great success since their engineering to become carrier proteins and were shown to keep better cell viability and have better efficiency than AIDA and Lpp-OmpA amongst others. For this reason, both the InaQN and InaVN INPs are used in this study.

Table 2. Overview of the surface display technologies selected in this study and their past applications.

Type	System	Origin	Host	Applications
AT	AIDA	*E. coli*	*E. coli*	Biocatalysis of human cytochrome P450 1A2 and cytochrome reductase[100]; Library screening and produciton of affibody molecules[103,104]; Display of Mx antiviral proteins[106]; Biocatalysis of *B. cepacia* lipase[101]; Display of threhalase for antiviral antibody discovery[105]; Biocatalysis of xylan hydrolysis by *C. cellulovorans* endo-β-1,4-xylanase[102]
	EspP	*E. coli*	*E. coli*	Secretion of CtxB[107]; Screening of affibody anticalin libraries[97]
	MATE	*E. coli*	*E. coli*	Display and secretion of test cargos (multiple cloning site, EstA, mCherry)[96]

			P. putida	Biocatalysis of cellulose hydrolysis[108]; Display of *B. gladioli* Esterase[99]
	YfaL*	*E. coli*	*E. coli*	Recombinant protein prification[126]; Secretion of agarolytic enzyme[127]
OMP	eCPX	*E. coli*	*E. coli*	Production of living metallic biohybrids[119]; Conjugation of fluorescent quantum dots on *E. coli*[120]; Building a supramolecular interaction net between E. coli cells[121]; Peptide screening for binding to Protective Atigen of *B. anthracis*[122]
	Lpp-OmpA	*E. coli*	*A. baumannii*, *P. aeruginosa*	AMP screening[33]
			E. coli	AMP screening[33]; Ligand immobilization[114]; Lead bioabsorption[113]; Biocatalysis of xylan hydrolysis[112]; Lipid biocatalysis by *B. licheniformis* lipase[111]; SpyCatcher-SpyTag display system[110]
	InaQN	*P. syringae*	*E. coli*	Study of the VP8* protein of the rotavirus[125]; Biocatalytic hydrolysis of organophosphates[115]; Phosphate biosorption[124]
			P. putida	Phosphate biosorption[124]
	InaVN	*P. syringae*	*P. putida*	Biocatalytic degradation of organophosphates[123]
			V. anguillarum	Display of GFP[128]

*Apparatus used for the display of the proteases

1.2.2 Protease-mediated cargo release

To achieve a functional secretion system, not only must the AMP be exported to the outer surface of the bacterium, but it must also be released. Many AT naturally liberate their native cargo after export; however, this auto-proteolytic activity is highly dependent on the conformations of its target cut site and may not be compatible with non-native cargos[129–131]. In addition, they lead to the release of their cargo fused to a considerable length of the linking sequence located between the cargo and the TMD. For the secretion of AMPs, this left over amino acid chain is of non trivial length as it can be over 100 amino acids long[132].

Alternatively, using outer membrane proteases endogenous to the host bacterium chosen could work by using their target cut site in the sequence of the AMPSEC. *E. coli* possesses an outer membrane targeted family of proteins called omptins, which includes proteases such as OmpT, OmpP and IcsP (SopA) in enteroinvasive strains. These proteases, however convenient, are not incredibly specific and possess broad roles in cleaving various proteins and peptides with the exception of IcsP, which only has IcsA as a target[133,134]. They share, however, a similar cleavage site located between two positive residues[133].

The previous two protease options would only necessitate a single expression vector for the SDA as the proteolytic activity used for the release of the cargo would either be catalysed by the SDA itself or completed by endogenous proteases. However, by having a second expression system for the protease, we would have the opportunity to look into non-endogenous alternatives that can present a higher cleavage specificity. Unfortunately, highly specific outer membrane proteases have not yet been discovered as most of them are part of the omptin family and thus tend to possess a broader range of target[135,136]. Therefore, we would need to target soluble proteases to the OM. For this method, it would be possible to

choose any protease we would like as long as it is compatible to be expressed as a recombinant with a SDA anchor. In this case, it would be preferable to find highly specific and reliable proteases to prevent any potential use with non-specific cleavage. It would also be preferred if the selected protease did not have any involvement in human pathogenesis. Tag-cleaving proteases used in protein purification represent a group of often synthetically optimized proteases with very specific cleaving sites and high efficiency. In this list, we can find the enterokinase, factor Xa, Genenase™ I, Thrombin, Tobacco Etch Virus (TEV) protease, inteins and SUMO proteases. While most of the above work with specific sequence site recognition, the inteins and SUMO proteases work with a self -cleaving mechanism and conformational site recognition respectively[137]. Within this list, the most reliable seem to be TEV, inteins and SUMO proteases, while the others have all been reported causing unspecific cleaving in some situations[138,139].

1.2.3 Protein secretion systems

With the combination of a functional SDA and a protease, an assembled AMPSEC would be able to express, export and release a given cargo protein. In literature, a few attempts at creating such a protein secreting system using this method can be found. In 2012, Jong et al.[140] produced an independent secretion system using both the Hbp AT and the EspC AT from *E. coli* and their respective autoproteolytic activities under the *lac*UV5 promoter. They replaced the side domains of Hbp and the cargo domain of EspC with the *Mycobacterium tuberculosis* antigen and vaccine target ESAT6 and achieved successful secretion. Similarly, the serine protease AT Pet of *Enterobacteriaceae* was used in the same way by replacing its original cargo with proteins of sizes ranging from 100 kDa to 10 kDa[141]. The new cargo proteins were effectively released by the AT itself.

While these previous papers used the natural secretion mechanism of the AT, Fleetwood et al.[103], used the AIDA AT coupled with endogenous OmpT to secrete an IgG-specific affibody molecule. They used an inducible araBAD promoter to produce their SDA containing an OmpT recognition site for their affibody molecule and transformed their system in an OmpT-expressing DH5α strain. The affibody molecule was purified from the supernatant using column chromatography. Their system afforded good cell viability and was shown to be versatile in the secretion of various other affibody molecules. Similarly, OmpT-positive *E. coli* strains were used for the secretion of mCherry using a novel engineered AT dubbed MATE under the control of an inducible T5 promoter[96]. mCherry was also successfully secreted by Zhou et al.[126] using a two-strain secretion system. The research group made use of two separate SDA: one for the surface display of the cargo protein and another for the surface display of the protease. The system used a SUMO protease-mediated release with the Ulp1 SUMO protease. mCherry was displayed using the Lpp-OmpA OMP-based SDA, while Ulp1 was carried to the surface by the YfaL AT. Both systems were transformed in separate strains and the two were co-incubated to allow them to interact at the surface of the bacterial cells.

While the above summarizes the independent secretion systems utilizing SDAs created thus far, some have alternatively achieved secretion by exogenously treating their cargo-presenting cells with a specific protease like TEV[105,127]. Although the existence of functional secretion systems shows promises for the creation of AMPSEC, many aspects need to be taken into consideration for the goals of this project. Firstly, AMPs are quite small and cargo compatibility with the SDA chosen can be an issue. Secondly, the activity of an AMP can be very easily altered and leftover amino acids from a proteolytic release may

impede its activity. Finally, host cell viability as well as the independence of the system is crucial for *in vivo* applications where the proteins cannot be induced, and the bacteria cannot be treated.

1.2.4 Hypothesis and objectives

By using surface display technologies, coupled with proteolytic activity, a secretion system can be designed for the production of functional AMPs (Fig.1):

- Design a proof-of-concept workflow to study the detailed functionality of the secretion systems.
- Design recombinant surface display apparatuses coupled with protease activity to achieve secretion through type V-like secretion pathways in *E. coli*.

1.3 Part III: Using bioinformatics to understand AMPs

1.3.1 Protein clustering

With the emergence of artificial intelligence, machine learning has become a common tool in proteomics data analysis[142]. It can be divided into three approaches: supervised, unsupervised and reinforcement learning. Supervised learning uses labelled data to train the model for use on classifying unlabelled data[143]. Reinforcement learning trains the model against a specific environment with a set goal to which it needs to find the optimal solution[144]. An unsupervised learning model is trained on unlabelled data in which it discovers features and patterns to train itself in discovering classifications[145]. This last method is used to bring forth previously unknown patterns and characteristics in the data. For example, it is in this third family of algorithms that we find clustering algorithms. These algorithms create order and categorization of unlabelled datasets by grouping the data according to similar features[146]. There are four major kinds of clustering methods:

partitioning, hierarchical, overlapping and probabilistic. In partitioning clustering, a piece of data can only belong to one group. K-means is one of the most popular partitioning clustering algorithms. It creates a predefined number of clusters (k clusters) by dividing the data according to the closest cluster mean (centroid). In hierarchical clustering, every piece of data is a cluster, and it is the incremental agglomeration of these clusters that creates increasingly larger groups[147]. In overlapping clustering methods, the same data point can belong to multiple groups with varying degrees of closeness to them. This method includes the fuzzy c-means algorithm[148], which works similarly to the k-means algorithm with the exception that the membership parameter of a data point to a cluster is non-binary. A data point then belongs to every cluster with a varying degree of membership. Finally, probabilistic clustering associates the data by the probability value of each data point belonging in each cluster[149,150]. A probabilistic approach can be used in combination with the previous two methods of hierarchical and overlapping clustering.

1.3.2 Protein sequence alignments

Often necessary to the protein clustering pipeline comes sequence alignment algorithms[151]. These algorithms access the similarity between the sequences contained in the dataset. One of the most common types of sequence alignment methods is the pairwise sequence alignment (PSA) where sequences are compared two at a time. EMBOSS[152], BLAST[153], CD-HIT[154], UCLUST[155], Needleman-Wunsh[156], and Smith-Waterman[157] are all examples of pairwise sequence alignment algorithms. A second just as common type of alignment algorithm is the multiple sequence alignment (MSA) methods that assumes evolutionary homology to arrange sequences into arrays, where the residues of each column are the same or functionally similar[158]. MUSCLE[159,160], CLUSTALW[161], Clustal omega[162],

T-Coffee[163] and K-align[164,165] are all examples of multiple sequence alignment algorithms. In both cases the alignment can be done either globally (whole sequence alignment) or locally (partial sequence alignment). While PSA seems to be particularly useful in measuring sequence similarity in their structure, in their function or in evolution[156,157], MSA serves in finding sequence homology in conserved non-primary structures and in building phylogenetic trees.

1.3.3 Protein motif discovery

The protein clustering algorithm will yield groups of peptides that the algorithm has found to be related in some way. Next, it is often left to *de novo* motif discovery algorithms to extract what exactly is the familiar sequence characteristics that a group of peptides may share. In the context of this thesis, after generating AMP clusters by applying previously defined clustering algorithms to a database of antimicrobial sequences, clusters could be found enriched in a certain preferred target organism. Using motif discovery tools on those clusters could extract sequence motifs that give the AMPs contained in the cluster their specific target activity.

Although motif discovery algorithms in DNA and RNA are quite well established[166], it is a much more complicated issue for proteins given their larger alphabet and the complexity of the amino acid properties[167]. At the moment, we can find the MEME Suite[168] and the Teiresias[169] algorithms for protein *de novo* motif discovery. Within the MEME Suite, the MEME[170], XSTREME[171] and GLAM2[172] algorithms allow for different approaches to motif discovery. MEME uses a finite mixture model extending expectation maximization to find motifs that can occur in all or some sequences as well as once or multiple times per sequence. This algorithm outputs a pre-set number of motifs. XSTREME combines both

MEME and STREAM[173] motif discovery algorithms along with the SEA[174] enrichment analysis algorithm to find enriched motifs that can be found anywhere within the sequences. GLAM2 specializes in finding short functional motifs while considering gaps and insertions and deletions. It behaves similarly to a hidden Markov model that considers the interdependence of the residues in the motif sequences and is based on a Gibbs sampling algorithm[175]. Finally, the Teiresias algorithm enumerates through the given sequences to find repeated patterns before combining them to output motifs with maximized length.

Alternatively, motif discovery algorithms could be developed using clonal selection algorithms, which are one of the types of artificial immune system algorithms[176]. These are based on the antigen recognition process of the immune system and algorithms simulating this can be used for pattern recognition.

From there, we find two main ways of representing motifs: deterministic or probabilistic. In the deterministic representations, regular expressions are used to showcase the diversity of potential residues at each position in the motif sequence (Fig 2A). However, this method only shows the possibilities without retaining which is the most likely. This is where the probabilistic methods come in. These methods score the probability of each residue appearing at each position in the motif and can be represented by a sequence logo (Fig 2B). These discovered motifs could then be used for guided AMP library creation against the given target to find more specific and stable AMPs with therapeutic potential. This created library could then be screened using the AMPSEC system and the effective AMPs could further be characterized also using AMPSEC in *in vitro* and *in vivo* assays.

A

G[^DEFHKPRWY][A-Z][ACGNST][^P]

B

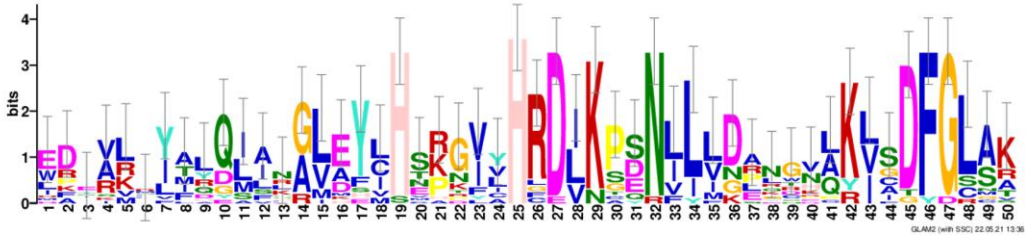


Figure 2. Representations of protein sequence motifs. A) Using regular expressions to describe a sequence motif. Brackets indicate which amino acids can be in the given position and where the brackets with "^" indicates which amino acids cannot be at the given position. This example was taken from Mulder and Nelzek, 2006.[177] **B)** Using probability logos to illustrate which amino acids can most likely be found at each position. The height of the amino acids is proportional to their information content. This example was taken from the sample output of MEME Suite's GLAM 2 algorithm[172] which can be found here: GLAM2 Results (meme-suite.org).

1.3.4 The use of bioinformatics in AMP research

In general, in the field of AMPS, bioinformatics has been mainly used to predict functional sequences[178]. In other words, the algorithms developed for AMP research are aimed at identifying novel active antimicrobial sequences. These include both conventional machine learning approaches[178,179] as well as deep learning approaches in more recent years[180,181]. These techniques are used for preliminary *in silico* screening of sequence libraries for potentially functional sequences that could be tested *in vitro*. Alternatively, some efforts have also been made in the classification of AMPs, however, those were still all done with the aim to predict functional sequences[182]. Moreover, any sub-classifications of the AMPs have to do with their activity (i.e., cellular targets)[179,183–185]. Herein, we propose to develop an approach designed to understand AMPs on the basis of their target species or cell type. To avoid unhealthy broad-spectrum activities in AMPs, designing sequences with the specificity placed as a priority over the activity would be an interesting new angle to

approach AMP sequence design for clinical use. To achieve this, we tried MMSeq2's Linclust[186] algorithm which falls in the category of partitioning clustering algorithms. This algorithm makes use of gapped and ungapped local sequence alignments to score the similarities between sequences. These similarity scores are then used by their greedy clustering algorithm to form the final groups.

1.3.5 Hypothesis and objectives

By using MMSeq2's Linclust algorithm, we can cluster AMPs in a way that will let us explore the correlation between the AMPs' primary structures and their targets:

- Extract AMP data from the DRAMP database.
- Run Sequence data through the Linclust algorithm to output large clusters.
- Explore the target specificity of each cluster in relation to the nature of the sequences contained in it.

2 Chapter I – Development of a peptide secretion system

2.1 Materials & Methods:

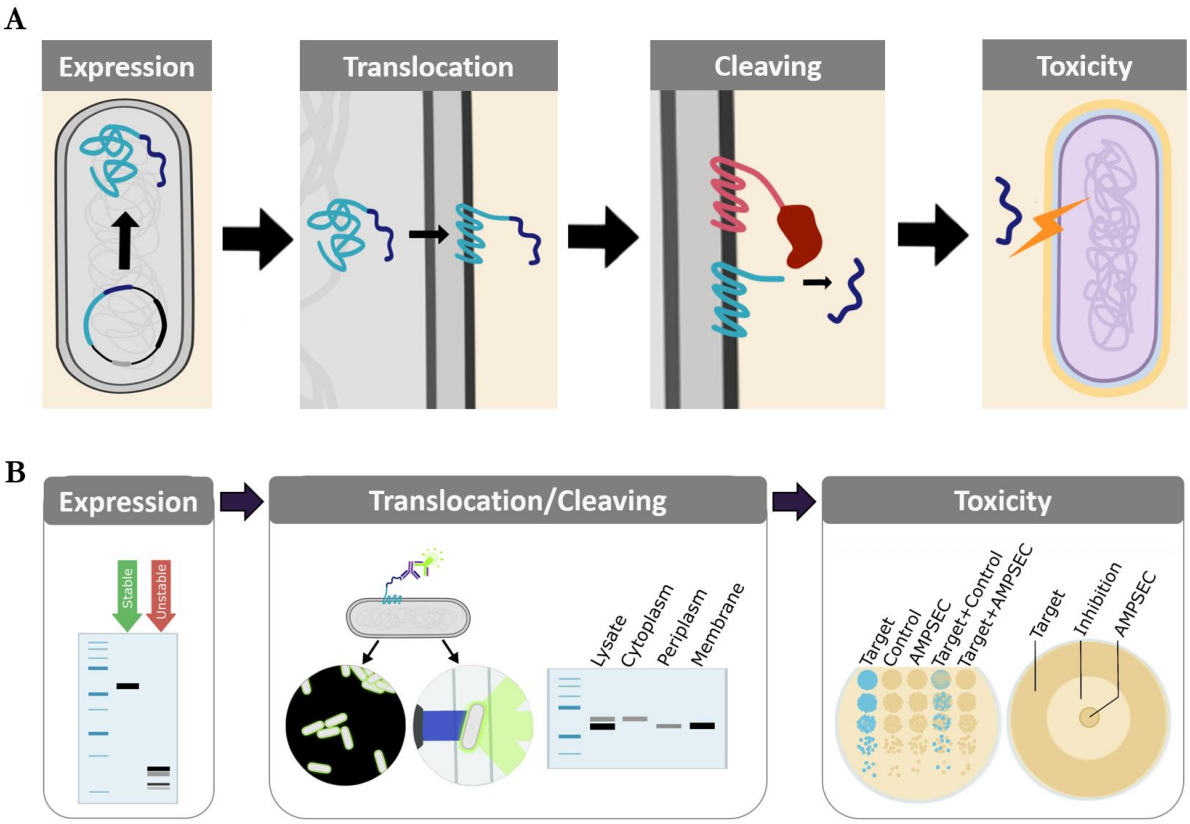


Figure 3. Proof-of-concept workflow for the development of AMPSEC. A) Theoretical representation of key milestones in the functionality of AMPSEC. **B)** Experimental pipeline designed to assess the achievement of each milestone illustrated in A.

2.1.1 Bacteria, DNA and Chemicals

The bacterial strains used in this study include *Escherichia coli* BL21 derived from an M15 BL21 and *Listeria innocua* (ATCC 33090). *Escherichia coli* DH5α was also used for cloning and to amplify the gene for OmpT. The *E. coli* strains were grown using Luria Bertany (LB) medium while the *L. innocua* strain is grown with Brain Heart Infusion (BHI) medium. In cases where *E. coli* and *L. innocua* are grown together, BHI medium is favoured. Plasmids used are the pMK90 vector (GeneBank U13854.1) originally obtained from Dr.

Inga Benz and the pSU2.1 vector[187] for the first theoretical constructs. The vectors pUdO1s and pUdO3c were used for the rest of the study. The pUdO plasmid series were created and designed by France Ourida Manigat and Dr. Campbell-Valois as part of the bioGARAGE project (Unpublished). All primers and genes were ordered from Integrated DNA Technologies except for CtxB containing genes that were ordered from GeneWiz. Genes were ordered by synthesis into a vector. All primer and gene sequences are listed in Appendix I. All chemicals, buffers, media, enzymes and reagents were obtained from Thermo Fisher Scientific unless specified otherwise.

2.1.2 Gene Assembly and Mutagenesis PCR

All PCRs were done using the Q5 High-Fidelity DNA Polymerase (NEB #M0491). The reaction mix consisted of 1X of Q5 Polymerase Buffer (NEB), 1X High GC Enhancer (NEB), 200 μM of dNTPs, 0.02 U/μL of Q5 High-Fidelity DNA Polymerase and 0.5 μM of forward and reverse primers. Tubes were loaded into a T100 Thermal cycler (Bio-Rad #1861096) set to the following program: 98 °C (30 s) ; 30 cycles of 98 °C (10 s) for denaturation, 30 s at the given annealing temperature and 72 °C for the given elongation time; 72 °C (2 mins) and a final hold temperature at 4 °C. The amplified DNA was treated with DpnI and purified using the Gel and PCR purification kit (BioBasic #9K-006-0008) following the manufacturer's protocol.

The mutagenesis of AIDA's cleaving site was done using the primers HMIO292-HMIO293 at an annealing temperature of 60.5 °C and an elongation time of 2 mins. The purified linear plasmids were phosphorylated using 10 U of PNK in 1X PNK Buffer A with 1 mM ATP. The plasmid was then circularized by ligation using 5 U of T4 ligase in 0.9X T4 Ligase Buffer. The mix was incubated 15 mins at RT then ligated plasmids were transformed

in DH5α for screening. The linker removals were done using the primers HMIO399-HMIO401, an annealing temperature of 60.5 °C and an elongation time of 2 mins 30 secs. The same process as described above was used for recircularization of the linear plasmid.

For the OmpT-using constructs, the myc tag and the OmpT cleaving site was inserted into the surface display apparatus in the pMK90 plasmid using PCR. The same method described above for AIDA's cleaving site mutation was used. The primers HMIO425-HMIO426 were used with an annealing temperature of 63 °C and an elongation time of 3 mins. The sequence of OmpT was amplified by PCR from the genome of DH5α using primers HMIO431-HMIO432, an annealing temperature of 61 °C and an elongation time of 30 secs. OmpT was inserted into pSU2.1 between the BamHI and XbaI restriction sites. 1 µg of the amplified OmpT and the destination vector was digested with 10 U each of BamHI and XbaI in 1X Fast Digest Green buffer. The mix was incubated 20 mins at 37 °C and the enzymes were subsequently deactivated at 80 °C for 5 mins. The desired digested DNA was extracted from an agarose gel using the BioBasic gel and PCR purification kit. Insert and vector DNA were ligated with a 1:3 vector:insert ratio using 5 U of T4 Ligase in 1X T4 Ligase Buffer. The mix was incubated 2 hrs at RT then ligated plasmids were transformed in DH5α for screening.

For the toxicity assay, the OTII cargo was replaced by Plantaricin-423wt (Pln-423wt) and Plantaricin-423 mutant (S27D + K36N) (Pln-423m) using Gibson Assembly. The Pln-423 sequences were taken from Guralp et al.[70] and synthesised in plasmids by BioBasic. Primers HMIO463-HMIO465 were used with an annealing temperature of 60 °C and an elongation time of 4 mins to amplify the vector while primers HMIO480-HMIO482 were used with an annealing temperature of 67 °C and an elongation time of 10 secs to amplify the

Pln-423 genes. The NEBBuilder® Hifi DNA Assembly Master Mix (NEB #E2621) was used according to the manufacturer's protocol using a 1:5 vector:insert ratio. The FLAG tag was also removed using mutagenesis PCR to prevent it disturbing the activity of Pln-423 using primers HMIO485 and HMIO487-HMIO488 with an annealing temperature of 67 °C and an elongation time of 4 mins.

For the CtxB-carrying constructs, the AIDA sequence was amplified was amplified by PCR from the pMK90 vector using primers HMIO507-HMIO508, an annealing temperature of 60 °C and an elongation time of 45 secs. All the other necessary genes were synthesised. Genes were assembled into the pUdO1s and pUdO3c vectors using the Golden Gate Assembly BsaI-HFv® Master Mix (NEB #E1601) according to the manifacturer's protocol. The list of genes used to build the constructs used are listed in Appendix I.

2.1.3 Induction of Recombinant Secretion Systems

BL21 *E. coli* was transformed by electroporation with every construct and the empty plasmid version for the protease vector when applicable. The constructs using a separately expressed protease were also transformed with the protease-expressing vector as a separate strain. Overnight cultures in LB with antibiotics were made at 37 °C with shaking. On the day of the experiment LB broth with antibiotics was inoculated with fresh ON cultures and incubated at 37 °C with shaking up to the specified OD_{600} (~4-6 hrs).

2.1.4 Cell Fractionation

20 mL induced cultures to OD_{600} = 2.0 were kept at 4 °C for the duration of this experiment unless specified otherwise. This protocol is an adjusted combination of the Malherbe et al.[188] and Sandrini et al.[189] protocols. Cells were pelleted at 4000xg 10 mins and resuspended in 800 µL PBS (Wisent Inc. #311-425-CL). The next few centrifugations are all

at 6000xg 2 mins until specified otherwise. Cells were resuspended in 900 µL spheroplast buffer (0.1 M Tris pH 8.0, 500 mM sucrose, 0.5 mM EDTA pH 8.0) for a 5 mins incubation. They were then resuspended in 400 µL of hypotonic solution (1 mM $MgCl_2$) for 15 s before adding 20 µL of 20 mM $MgSO_4$. Cells were pelleted and the supernatant was kept as the periplasmic fraction. Pellet was resuspended in 1mL lysis buffer (50 mM Tris-Hcl, 150 mM NaCl, pH 8.0) with 1 mM of PMSF (Sigma-Aldrich P7626), 150 µg/mL of lysozyme (BioShop #LYS702), 10 µg/mL of DNAse I (Sigma-Aldrich #DN25) and 10 mM of $MgCl_2$ then incubated at RT for 10 mins. The cells were sonicated using the UPH50 Ultrasonic Processor (Heilscher) set to 80% intensity with a pulse frequency of 0.8 s. The lysate was left on ice for 5 mins before centrifuging for 10 mins at 6000xg to remove whole cells and large debris. The supernatant was ultracentrifuged at 186 000xg for 30 mins. The supernatant is the cytoplasmic fraction, and the pellet was resuspended in lysis buffer as the total membrane fraction. All fractions were corrected to the same volume before western blot analysis.

2.1.5 Protease Accessibility Assay

5mL induced cultures to OD_{600} = 1.0 were pelleted 4000xg 15 mins and cells were washed with 1 mL PBS. The cells were resuspended in 250 uL of 0.5X lysis buffer with or without 50 U of TEV (NEB #P8112) or 50 U of Ulp1 (Cedarlane #NU01015) then incubated for 1h at 30 °C. The whole cells were then separated from the supernatant by pelleting. The proteins in the supernatant were precipitated by adding ¼ volume of 100% trichloroacetic acid (TCA) and incubating on ice for 30mins. Precipitated proteins were pelleted by a 10 mins spin at 18 000xg then washed twice with ice cold acetone using 5 mins spins. The supernatant proteins and whole cells were analysed by western blot.

2.1.6 Western Blots

Culture samples or fraction samples were combined with Laemli (4% SDS, 20%$_{v/v}$ glycerol, 120 mM Tris pH 6.8, 0.02% Bromophenol Blue, 10%$_{v/v}$ β-mercaptoethanol) and heated at 95 °C for five minutes. The samples were run on 12% SDS-PAGE gels for the SDA and protease western blots and on 15% gels for cargo western blots using the Precision Plus Protein™ All Blue protein standard (Bio-Rad #1610373) as reference. Proteins were transferred onto PVDF membranes (Bio-Rad #1620177) and indirectly immunostained. At room temperature, membranes were blocked with TBST (20 mM Tris pH 7.5, 150 mM NaCl, 0.2% Tween) with 5% skim milk powder for 20 minutes then incubated 1h with the primary antibody diluted in TBS (20 mM Tris pH 7.5, 150 mM NaCl) with 1% Tween and 4% BSA (GE Healthcare Life Science #SH30574.02) and, finally 45 mins with the secondary antibody diluted in TBST-5% Milk. Membranes were washed three times 5 mins with TBST between each step then three times 10 mins after the last step. The primary antibodies used were mouse anti-myc (GeneScript #A00704), mouse anti-FLAG (Sigma-Aldrich #F3165), rat anti-ollas (Gifted from Dr. Ryan C. Russell, UOttawa). The secondary antibodies used were goat anti-mouse HRP (Jackson Immunoreseach ##115-035-003) and goat anti-rat HRP (Jackson Immunoreseach #112-035-167).

2.1.7 Fluorescence Microscopy

Coverslips were first sterilised then incubated in 10 μg/mL of polylysine in PBS ON at 4 °C. The coverslips were rinsed three times with PBS after that. 5 mL cultures were induced 3 hrs (until OD_{600nm} = ~0.8) then pelleted at 4000xg 5 mins. Cells were resuspended in 1mL 4% PFA in PBS for 15 mins at RT. All the following centrifugations are at 4000xg 2 mins, and the incubations are all at RT. Cells were washed twice with PBS then separated

between permeabilized cells and non-permeabilized cells. The permeabilized cells were incubated 1h in 70% EtOH while the others were resuspended in PBS during the incubation. All cells were washed twice with PBS then incubated 10 mins at room temperature in 10 μg/mL of polylysine in PBS. Cells were washed twice with PBS then resuspended in PBS. The samples were diluted 2:25 in PBS and the whole diluted mixture was placed on the coverslip and incubated 15 mins. Coverslips were rinsed once with PBS and then, the permeabilized samples coverslips were incubated in 25 μg/mL of lysozyme for 30 mins while the other samples were incubated in PBS. Coverslips were rinsed twice with 4% BSA in PBS then incubated 30 mins with the primary antibody diluted in PBS-1% BSA. Coverslips were washed twice with PBS and incubated 30 mins in the dark with the secondary antibody diluted in PBS-1% BSA. Coverslips were kept in the dark from then on. Coverslips were finally rinsed two more times with PBS before being mounted on the slides using Mowiol (24% glycerol, 9.6% Mowiol 4-88 (Sigma-Aldrich #81381), 96 mM Tris pH 8.5, 2.5% DABCO) mounting medium. The primary antibodies used were the mouse anti-myc and mouse anti-FLAG as described in the 'Western Blot' section (2.1.6) while the secondary antibody used was anti-mouse AF488. Cells were imaged using the Zeiss LSM 880 Confocal Microscope.

2.1.8 Flow Cytometry

4mL cultures were induced to OD_{600nm} of 1 (3.5 hrs). All centrifugation steps are at 4000xg for 2 mins and all incubations are at RT. The culture was resuspended in 500 μL 4% paraformaldehyde 10 mins. The cells were then resuspended in 100 uL ice-cold PBS with 0.1% sodium azide and 3% BSA, and pelleted. The pellet was resuspended in primary antibody diluted in PBS-1% BSA and incubated for 30 mins. Cells were washed twice with

the ice-cold PBS solution described previously. Cells were then incubated in the dark 30 mins with the secondary antibody diluted in PBS-1% BSA. Cells were kept in the dark from then on. Cells were finally washed twice then resuspended in 1mL of the wash solution. The antibody used are the same as described for the "Fluorescence Microscopy" methods (2.1.7). The Gallios flow Cytometer from Beckman was used to measure fluorescence.

2.1.9 Agar Plug Diffusion Assay

This assay was done as described in Balouiri et al.[190]. Briefly, the AMPSEC strain was spread on an LB plate a day before with antibiotics for induction. 150 µL of an ON culture of the target strain was seeded in 2 mL of 0.5% agar and poured on top of a BHI agar plate. A 10 mm diameter plug from the AMPSEC plate was placed on top of the target bed. Plates are placed in the refrigerator for 2 hrs before being places in the 37 °C incubator for ON growth to allow diffusion of the secreted AMP.

2.1.10 Chromogenic Toxicity Assay

This assay was done as described in Tucker et al.[33]. Briefly, 5mL cultures of the target and AMPSEC strains are induced for 2.5 hrs to an OD_{600nm} of about 0.6-0.8. Cultures are then diluted to a final OD_{600nm} of 0.01 in fresh media with no antibiotics. The AMPSEC and target strains were diluted individually and together as a 1:1 mix. The diluted cultures are then incubated 3 hrs. BHI agar plate with 50 µg/mL of X-glucoside were dotted with the diluted cultures before and after the 3 hrs incubation. Serial dilutions were made in PBS ($X10^{-2}$, $X10^{-3}$, $X10^{-4}$, $X10^{-5}$, $X10^{-6}$).

2.2 Results:

2.2.1 Achieving peptide secretion with the AIDA surface display apparatus and its autoproteolytic activity fails due to the instability of the construct.

Our first attempt at this project was to use a readily available surface display apparatus that had previously been used by our lab. The autotransporter (AT) AIDA-I, found in *E. coli*, acts as a diffuse adherence adhesin and is primordial for the pathogenicity of EPEC strains[191]. It has been used a significant number of times in recombinant expression experiments to display functional proteins on the OM of a gram-negative bacterium[103,192–197]. It uses the Sec pathway to traverse the inner membrane (IM), after which the 49 residues-long signal peptide (SP) of AIDA is cleaved. Then BamA complex facilitates the insertion of the mature protein in the OM while it simultaneously translocates its cargo[198]. To release this cargo, AIDA possesses an intramolecular protease activity that cleaves between the S846 and A847 residues[199]. This simplistic auto-proteolytic property of AIDA and its countless previous uses make it particularly attractive for this project. It would also remove the need for an associated protease expression.

The pMK90 plasmid was designed for peptide presentation[194]. This plasmid encodes the AIDA-I autotransporter under its constitutive promoter; *aidAp*[193] and allows the user to edit or change the cargo for recombinant studies. This AIDA-I construct harbours the A847I mutation at the cleaving site of the AT, which inactivates its auto-processing activity; this mutation allowed for the system to be used for surface display. By reversing its cleaving site mutation back to its wild genotype, we hoped to restore its original auto-proteolytic phenotype. In theory, with its cleaving site made functional again, we'll be able to see the secretion of our chosen cargo. Another feature of this AIDA construct is the size of the

linking sequence, which has been minimized to its shortest functional length (~54 out of the original 162 residues). This linker region connects the cargo domain to the TMD and was shown to be essential for the functionality of the AT with its native cargo[132].

Since we are hoping to use this system for the secretion of AMPs, which are short peptides, we plan on using an equally small but inactive peptide for its design. Thanks to Dr. Campbell-Valois's previous studies, a pMK90 construct with the OT-II chicken ovalbumin antigen fused to a FLAG tag placed as cargo was previously made by mutagenesis PCR (unpublished, personal communications) and can be used for this project (Fig 4A). With this in hand, the first step was to restore the protease activity of AIDA by mutagenesis PCR of I847A (Fig 4A). In principle, that would result in the entire minimal linker region of 54 amino acids being left attached to our passenger after cleaving. Considering the small size of AMPs and their sequence-sensitive nature, such a relatively large string of residues left attached could severely impair its activity. In this case, it would be in our best interest to get rid of this region while preserving the cleaving site so that the cut is made as close to the cargo as possible with a minimal residue leftover at the C-terminal of the cargo. For that purpose, further mutagenesis PCRs were made to remove this region in both the cleaving and inactive AIDA constructs (Fig 4A).

The expression of this first design was evaluated in the *E. coli* BL21 strain using the FLAG tag for Western blot (WB) immunodetection of the OTII cargo (Fig 4B). The full-sized construct with the linker can be seen at ~64 kDa while the linker-less version is seen at ~58 kDa. Successful reactivation of the cleaving site can be seen by the appearance of the very faint ~12 kDa band in the cleaving construct with the linker. The faintness of the band may be due to the uptake and degradation of the secreted inactive peptide by the bacteria.

Removal of the linker sequence, however, caused the construct to become unstable and be degraded whether the cleaving site was active or not. This is shown by the unspecific degradation bands indicated by the arrow in Fig 4B. The deletion of the essential linker sequence impairs translocation, and the use of a separate protease was deemed necessary. As explained previously, keeping the linker sequence that would stay attached to our cargo is not ideal and AIDA does not function properly without its linker[132].

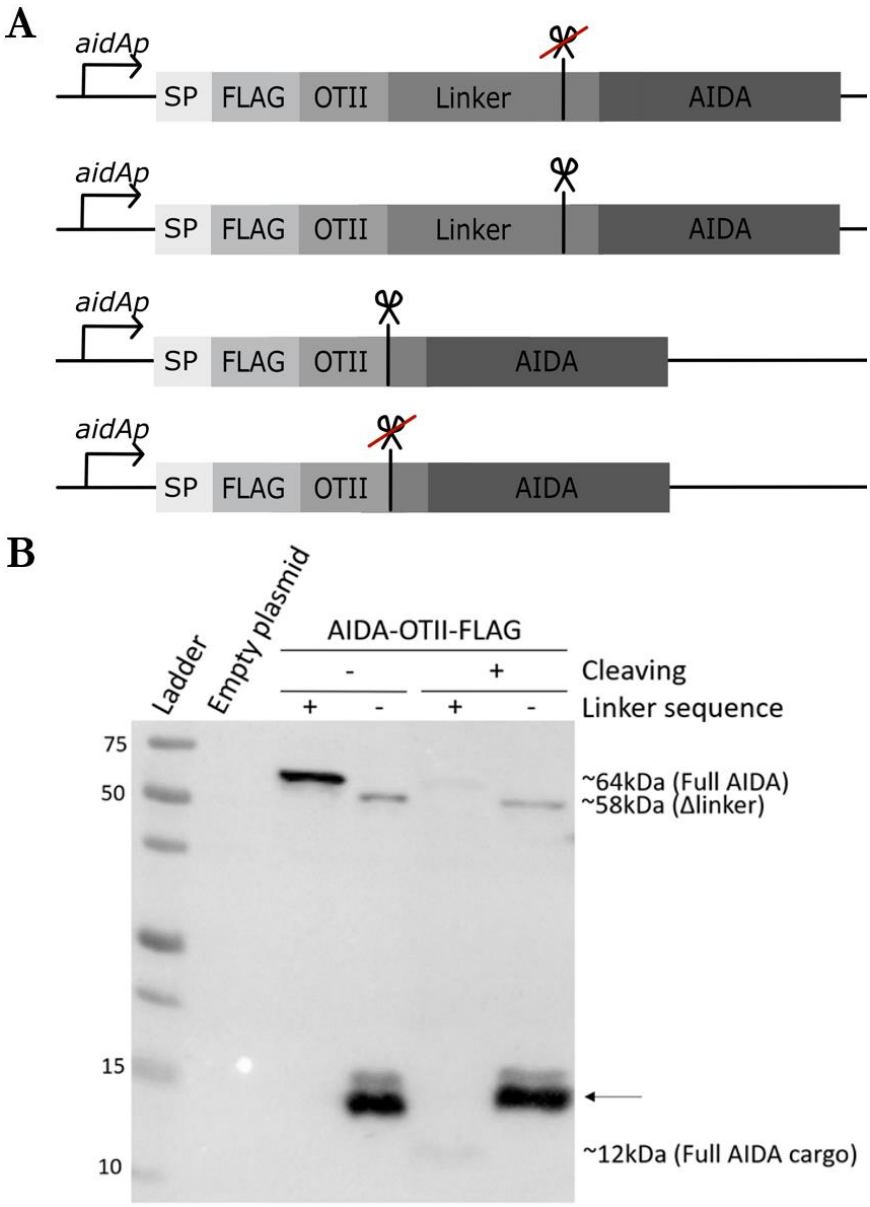


Figure 4. OTII-secreting system expression using the autoproteolytic function of AIDA. **A)** OTII secreting constructs under the constitutive *aidAp*. Constructs with the crossed-out scissors harbour the inactivating A847I mutation. SP = AIDA's SP; FLAG = Tag following the cargo; OTII = Inactive test cargo. The DNA sequences of all the parts can be found in Annexe I. Note that the AIDA sequence listed in Annexe I includes the linker. **B)** BL21 strains were induced by a 4-hour incubation up to OD_{600}s of ~2. Culture samples ran on a 15% SDS gel before their transfer onto the PVDF membrane. Cargo was immunodetected using a mouse anti-FLAG primary antibody and an anti-mouse HRP-conjugated secondary antibody. The empty vector expresses AIDA without any cargo. Masses on the right are the approximate expected sizes of the described protein bands. The arrow points to unstable degradation product. The masses of relevant ladder bands are labelled on the left in kDa.

2.2.2 Using the endogenous protease (OmpT) to cleave the cargo is disadvantageous due to its lack of specificity, but the overall secretion system showed promise.

As our first protease-coupled system, we thought of using OmpT, an OM protein of *E. coli*. Since OmpT is endogenous to most *E. coli* strains and since we're developing this tool for an *E. coli* host, it would allow us to keep this system simple with only one expression vector. OmpT generally cleaves after positively charged residues with preferences to Lys and Arg residues but it does possess a preferred cleaving site of A R/K ↓ R/K A/V[200–203]. In this case, the AR↓KV cleaving site was placed right between the linker and cargo to achieve cleaving (Fig 5A). A myc tag was also added before the cleaving site (Fig 5A). With this we could follow the cargo using the FLAG tag and the TMD of the surface display apparatus with the myc tag separately.

As done for the previous constructs, the expression of this new design was first assessed using WBs in Fig 5B. This construct was compared to the earlier one with the full inactive linker. The new construct can be seen expressing in three distinct bands. These were hypothesized to represent the full-length immature protein around 65 kDa, the translocated protein after its signal peptide is cleaved at ~60 kDa, and, finally, the processed version where the cargo has been cleaved at around 54 kDa. In the presence of the OmpT protease, the favoured form of the display system seems to be the fully processed version. This can be further confirmed by looking at the WB that followed the cargo. Without OmpT, only two bands can be seen: one for the full length at ~65 kDa and another for the translocated version without its signal peptide around 60 kDa. The third-lowest band is not observed as the cargo would be lost at this point. When OmpT is present, a very faint amount of the full-sized protein can be seen but as the construct is translocated, it is quickly processed and loses its

cargo. We attempted to visualize the cleaved version of the cargo using a 20% SDS gel but failed (results not shown). This may be due to the small size of the cargo (Fig 5A), or it might be due to the fast degradation of the peptide once secreted.

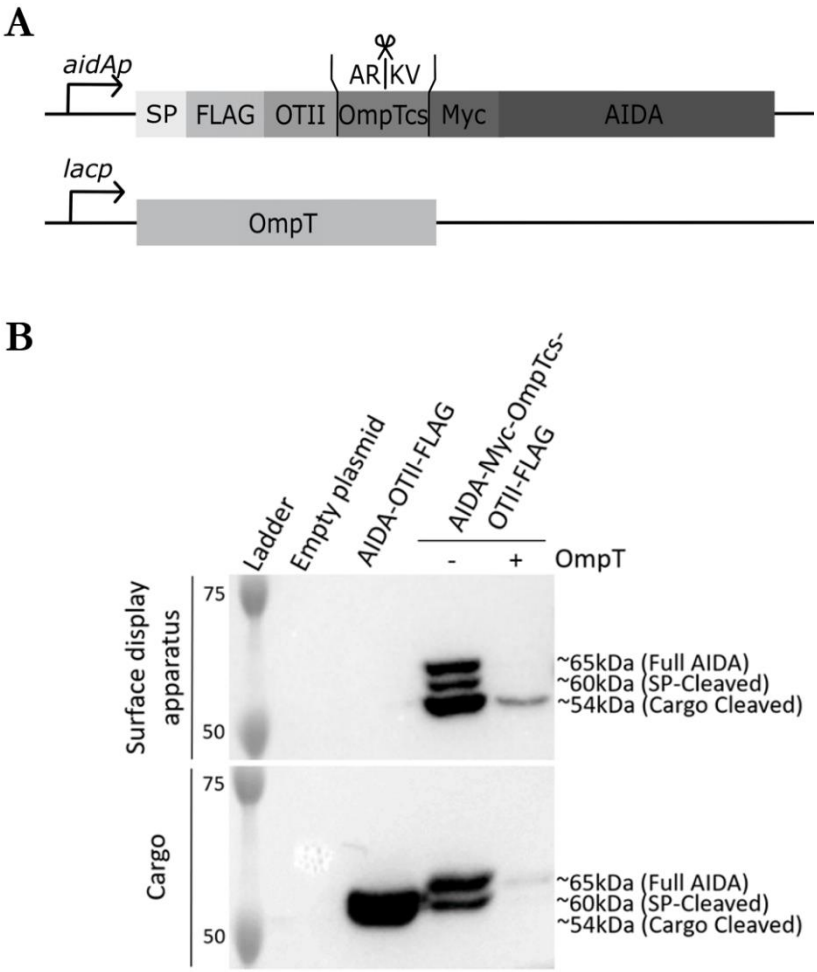


Figure 5. OTII-secreting system expression using OmpT for cargo cleaving. A) The OTII construct is under the constitutive *aidAp*. The expression of OmpT is under an endogenous *lacp*. SP = AIDA's SP; FLAG = Tag that follows the cargo; OTII = Inactive test cargo; OmpTcs = OmpT's cleaving sequence; Myc = Tag following the SDA; AIDA = AIDA AT; OmpT = OmpT protease. The DNA sequences of all the parts can be found in Annexe I. **B)** After induction of the SDA and OmpT up to OD_{600} of ~2. Culture samples ran on an 8% SDS gel before their transfer onto the PVDF membrane. The surface display apparatus and cargo were immunodetected using a mouse anti-myc primary antibody and a mouse anti-FLAG primary antibody respectively. An anti-mouse HRP-conjugated secondary antibody was then used for both. The empty vector expresses AIDA without any cargo. Masses on the right are the expected sizes of the protein species indicated. The masses of relevant ladder bands are labelled on the left in kDa.

From this blot (Fig 5B), we can also observe that cleaving seems to occur even without the protease. This may be due to the presence of OmpP, another outer membrane protease of *E. coli* that is present in the BL21 lab strains. This protease, like OmpT, favours positive residues and shares an almost identical preferred cleaving site of L S A R/K ↓ R/K G A/R G[201]. Moreover, these proteases are not known to be specific so it would be possible for OmpP to target the OmpT cleaving site used. It would also be important to note the large difference in the expression level of the system between the strains with and without OmpT. The over expression of a second protein (OmpT) might place a burden on the bacteria, causing a lower expression of the recombinant display apparatus. With respect to this, it might be relevant to compare the expression of this cleavable construct to that of the AIDA construct without the OmpT cleaving site with OmpT being expressed in both cases. This experiment would allow us to visualize the impact of the expression of OmpT on that of the SDA.

Moving on to functionality, the proper export of the cargo using this system was viewed using three different techniques: microscopy, flow cytometry and cell fractionation (Fig 6). In Fig 6A, we look at the localization of our expressed system. In the absence of the protease, we can see the three distinct bands observed previously (Fig 5B). We can also note that almost the entirety of the proteins produced, reaches the membrane fraction. Although these results demonstrate a successful functionality, it also disproves our earlier assumption that the top band was associated with a version of the construct that still had its SP (i.e., the cytoplasmic specie which has not gone across the IM). It would be more accurate to say that this top band is associated with the translocated AIDA which crossed the IM and whose SP was cleaved while the last band should still be the cleaved version. For the middle band, it is

likely a truncated version of the protein. Similarly to the display apparatus, the cargo can be seen properly following it to the membrane fraction. In the presence of the protease, however, other than the reduction in overall expression that was noted previously (Fig 5B), we see only the presence of the fully cleaved version of our display system per full cleaving of all the produced recombinant display apparatuses. Moreover, this is accompanied by a complete loss of the cargo signal following the loss of said cargo by its release into the surrounding media. Unfortunately, no signal was observed in the lower molecular weight region of this WB, thus failing to visualize the cleaved cargo in the total fraction. Again, this may be due to the cargo being of such a small size that the SDS-Page gel used lacked the necessary resolution or due to its quick degradation once released. In Fig 6B-C, we look at the outside surface of our cells to observe what is exposed. In accordance with the cell fractionation results, in Fig 6B, we can observe the display of the AIDA system at the outer surface with a diminished fluorescence in the presence of OmpT in agreement with the lower expression levels. Fig 6C is also in agreement with the results of Fig 6A: we can observe the exposed cargo in the absence of OmpT, while a complete loss of signal is noted when the protease is being expressed. Finally, those same results are corroborated by Fig 6D where signal is observed for the expression of the display system with and without the protease. As for the cargo (Fig 6E), a significant loss in fluorescence is observed when OmpT is expressed in the cells. Taken together, these results demonstrate a functional secretion system with AIDA successfully displaying the peptide cargo on the surface of *E. coli* and said cargo subsequentially getting cleaved. This prompted us to move forward to using an active AMP cargo with our system.

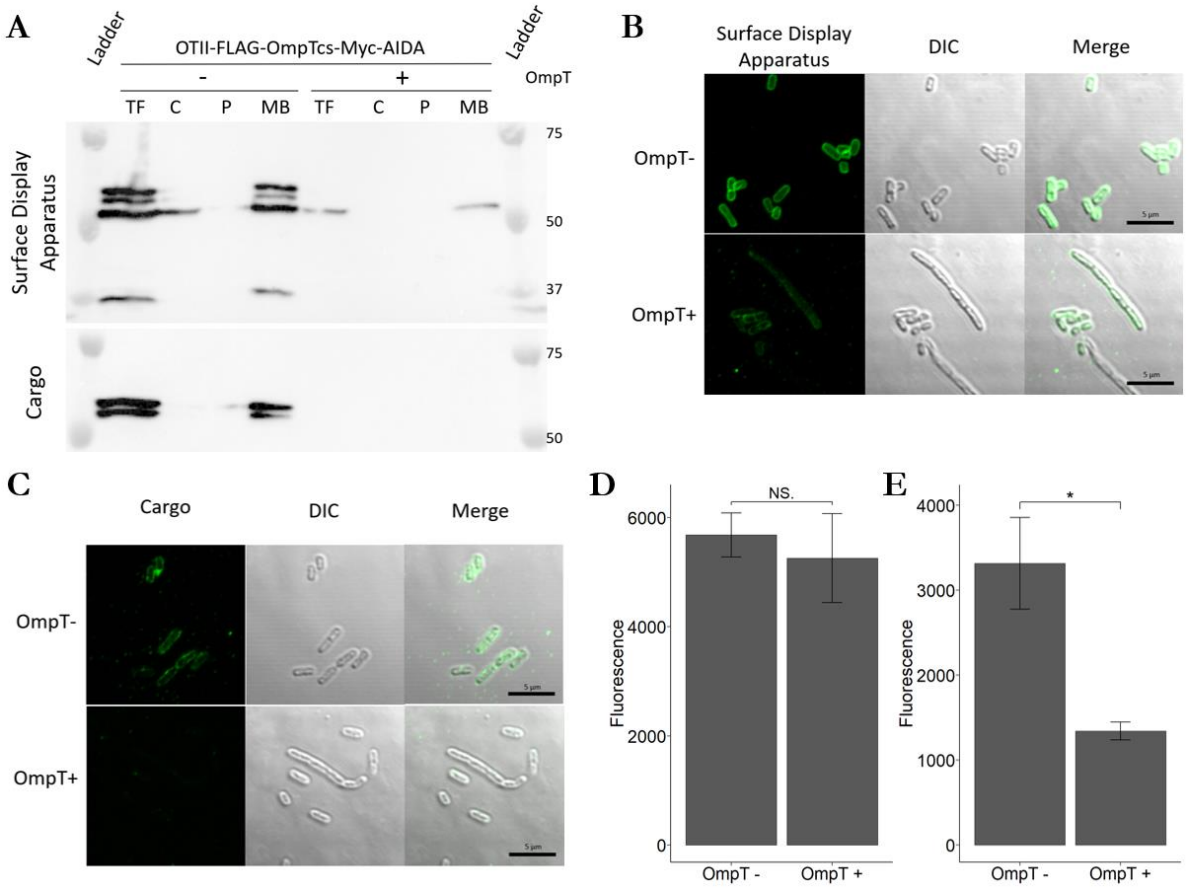


Figure 6. Functional study of the AIDA/OmpT prototype secretion system. A) Two separate WBs of cell fractions. TF = Total cell lysate with media, C = Cytoplasm, P = Periplasm, MB = Membranes. Bacteria were grown at 37 °C up to an OD_{600} of ~2.0. The periplasmic fraction was extracted by outer membrane permeabilization before lysis by sonication. Cytoplasmic and membrane fractions were separated by ultracentrifugation. An HRP-coupled secondary antibody was used for imaging. Numbers on the right side show the weights of the ladder in kDa. **B-C)** Fluorescence microscopy imaging of immunostained fixed bacteria initially grown to an OD_{600} of ~1. **D-E)** Flow cytometry on immunostained fixed cells initially grown to an OD_{600} ~1. **D)** Surface display apparatus, **E)** Cargo. Results presented are representative of the median fluorescence values of 3 separate experiments. A paired t-test was used for statistical significance. An AF488 secondary antibody was used for fluorescence in both the microscopy and flow cytometry experiments. In all the above, mouse anti-myc (surface display apparatus) and mouse anti-FLAG (cargo) primary antibodies were used. P-value notation: * $0.05 > p > 0.01$ and N.S.: $p > 0.05$.

The next step was to view toxicity by replacing the inactive cargo by an AMP and exposing our secreting strain to the target of the AMP (Fig 7). The researchers who developed an AMP secreting strain using an *lpo* knockout *E. coli* combined with a signal peptide-AMP fusion used plantaricin-423 (Pln-423) to target the gram-positive bacteria *L. innocua*[70]. We took the functional AMP sequence from their paper and placed it in our system, removing the FLAG tag in case it caused issues with its function (Fig 7A). Two toxicity assays were tested at this step: a colorimetric assay (Fig 7B) and an agar plug assay (results not shown). The latter being more quantitative could be used eventually to measure the effect of the secretion system used on the minimum inhibitory concentration (MIC) of the AMP compared to a purified extract. Unfortunately, no toxicity was observed in either case. No loss of the blue *L. innocua* strain was seen in the colorimetric assay (Fig 7B) while no inhibition ring could be seen in the agar plug assay (results not shown). As mentioned earlier, OmpT is a protease that targets positive residues. Moreover, OmpT also has a role in protecting the host from antimicrobial peptides by degrading them[133]. Pln-423 is a cationic AMP and is thus made mostly of positive residues. It would be possible that the overabundance of OmpT at the surface of the cell due to overexpression may be causing not only the cleaving of the cargo AMP but also its complete degradation. Alternatively, it may also be a possibility that the few left-over amino acids from the OmpT cleaving site may have abolished its activity. However, since we have also not been able to observe the cleaved OTII-FLAG cargo used previously, this could support a degradation hypothesis.

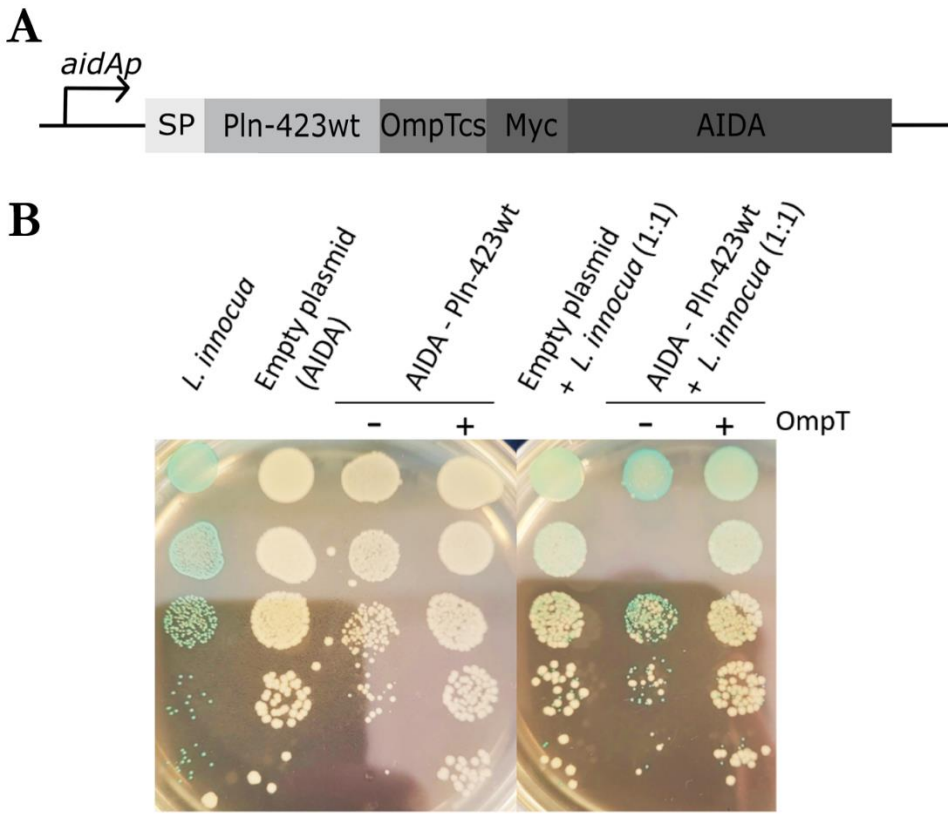


Figure 7. Colorimetric toxicity assay of the AIDA/OmpT prototype secretion system. A) The Pln-423 construct is under the constitutive aidAp. SP = AIDA's SP; Pln-423wt = Plantaricin-423; OmpTcs = OmpT's cleaving sequence; Myc = Tag following the SDA; AIDA = AIDA AT. The DNA sequences of all the parts can be found in Annexe I. **B)** Bacteria were first individually grown to an OD_{600} of ~0.6-0.8 with their respective antibiotics before being inoculated in 1:1 mixes in fresh media without antibiotics. Bacteria incubated 3 hrs before being dotted on agar supplemented with X-glucoside. PBS was used for the serial dilutions. Dilution factors from top to bottom: 10^2, 10^3, 10^4, 10^5, 10^6.

2.2.3 Comparing various display systems for their applicability in a secretion system design.

Despite the lack of toxicity observed with the OmpT-mediated plantaricin secretion, its potential was clearly demonstrated with its successful export and cleaving of the cargo. To explore thoroughly this concept, we tested seven of the most substantial surface display apparatuses including AIDA for the display and protease-mediated secretion of a model cargo (Fig 8 and Table 2).

For these new systems, we also changed the test cargo we used. To avoid any ambiguity in trying to visualize the cleaved cargo, we replaced the 2.5 kDa OTII protein with a 13 kDa CtxB protein corresponding to the B subunit of the cholerae toxin that we are confident will be detectable with more ease on our gels. This new cargo is not novel and has been used previously to test out display apparatuses such as AIDA[192,204,205] and EspP[107].

As the last improvement to the secretion system, we picked more reliable proteases to mediate the cargo release instead of OmpT: TEV and Ulp1. The AT systems have their cargo in N-terminal and the ultra-specific protease which would leave the least amount of amino acid overhang was found to be TEV[206] (Fig 8A). Its cleaving sequence is ENLYFQ↓(G/S) which will leave a short 6 amino acid overhang to the C-terminal of our AMP. Whether this will be enough to disturb the AMPs function is unknown and would need to be assessed on a case-by-case basis. The OMPs have their cargo in C-terminal which allows the use of the SUMO-Ulp1 cleaving system (Fig 8B). Ulp1 is an ultra-specific ubiquitin-like protease that recognizes the structure of the SUMO peptide and cleaves at its C-terminal regardless of the following amino acid with the exception of a proline[207], leaving no amino acid overhang. Along with the SDA itself, the protease will be targeted to the outer membrane using their own AT: YfaL (Fig 8C). The SUMO-Ulp1 cleaving system described here has been successfully demonstrated although the cargo-presenting SDA and the protease-presenting SDA were in different strains. The two strains were co-incubated to observe the activity of Ulp1 in cleaving the displayed cargo[126].

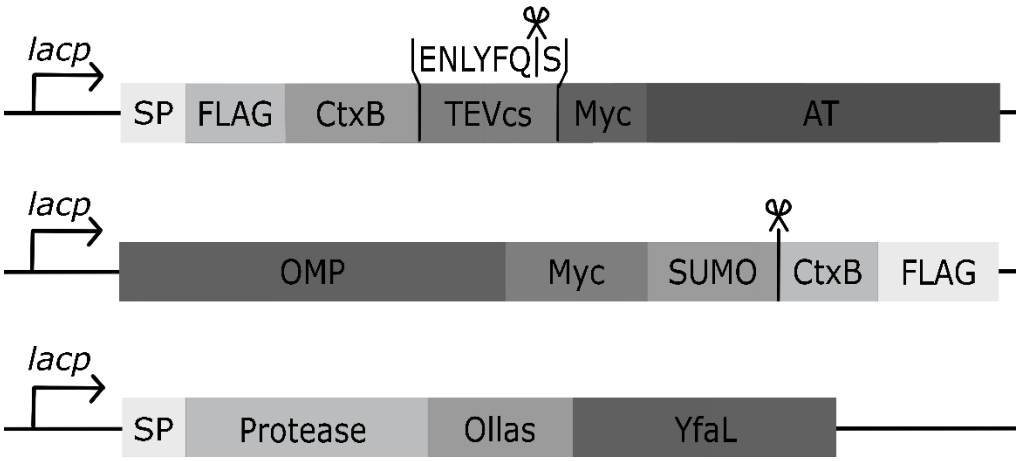


Figure 8. Secretion system constructs and their co-expressed protease system. All above systems use the endogenous *lacp*. From top to bottom; AT-mediated secretion systems, OMP-mediated secretion systems and protease SDAs. SP = AIDA's (top) or YfaL's (bottom) signal peptide; FLAG = Tag to follow the cargo; CtxB = 13kDa test cargo; TEVcs = TEV's cleaving site; Myc = Tag to follow the SDA; AT = Linker and TMD of a given AT (AIDA, EspP or MATE); OMP = SP and TMD of a given OMP (LPP-OmpA, InaQN or InaVN, eCPX); SUMO = Peptide cleaved in C-terminal by Ulp1; Protease = TEV or Ulp1, Ollas = Tag to follow the protease construct, YfaL = Linker and TMD of the YfaL AT. All the DNA sequences of the parts can be found in Annexe I.

All the genes used were synthesized except for the AIDA sequence, which was amplified from our earlier construct. The pieces were assembled by Golden Gate reaction and the constructs were transformed in BL21 on their own with an empty plasmid of the other resistance, or together in SDA/Protease pairs. Unfortunately, it is at this stage that we encountered issues with one of our constructs: eCPX. Due to the use of a constitutive promotor, the expressed construct might be toxic and, as of now, no successful assembly was observed.

Nevertheless, the remaining 6 SDAs were put through the proof-of-concept (POC) pipeline described in Materials and Methods (Fig 3) and tested in the previous results (Fig 5-7). The first step of the POC workflow is to look at the expression of the freshly designed constructs to make sure they are stable (Fig 9). Most systems showed expression of what

could be associated with the proper construct with some exceptions (Table 3). AIDA showed absolutely no bands in both that SDA and cargo WBs hinting at an issue pertaining with expression. The expression of the EspP construct (~48 kDa) was also accompanied by a certain amount of unspecific degradation products. This wasn't unexpected as this system was described as producing unidentifiable secondary bands[97]. For now, we can see that those degradations happened at the N-terminal as these lower molecular weight bands are not seen in the WB following the cargo suggesting that it was lost. Supporting this, in the cargo WB, we can see a cargo-sized band (~18 kDa) suggesting the spontaneous cleaving of the cargo without the presence of the protease. In that case, one of the two degradation bands observed would most likely correspond to a protein close to the processed SDA (~34 kDa). With the protease expressed, however, we have a complete loss of SDA signal, and we only observe the smaller cargo-sized band on the cargo WB. For now, this is hinting that something unexpected might be happening with the cleaving as we would expect to still be able to see a band on the SDA WB representing the cleaved version of the AT (~34 kDa). For MATE, without the protease, we can see the full SDA band at the approximate expected height (~60 kDa) on the SDA WB and on the cargo WB proving that it corresponds to the uncleaved version. The presence of the protease, surprisingly, does not seem to affect it with its bands staying identical in both WBs. This could hint at an accessibility issue of the protease. For InaQN and InaVN a band representing the full-sized construct can be seen in both the SDA and cargo WBs (~55 kDa) without Ulp1. With the protease, on the other hand, there is obvious cleaving happening with the appearance of a prominent band of slightly lower molecular weight in the SDA WB representing the cleaved SDA (~42 kDa). We could almost say that the protease is successful in cleaving near 100% of the proteins. This is accompanied by the appearance of bands around 18 kDa in the cargo WB representing the

cleaved cargo. Finally, for Lpp-OmpA, we can also see a full-size band in both the cargo and SDA WBs (~60 kDa) without Ulp1, although they are faint. The protease expression in this case seem to cause the SDA to become unstable and get further degraded beyond the cleaving of the cargo. Unfortunately, the cleaved cargo could not be observed in this case. Now, looking at the expression of the proteases themselves, we can see their expression (i.e., ~79 kDa for TEV and ~77 kDa for Ulp1) with unknown secondary bands in all cases. The degradation pattern was also observed to be different for both proteases which could mean that this instability may be caused in part by the recombinant proteases and that each protease affects the stability differently. This degradation, however, does not seem to impact the function of the SDA and the activity of the protease it carries. Of course, the protease system bands are only observed in the strains in which the protease-expressing plasmid is present. It was noted that the expression of the Ulp1-carrying systems had a much fainter expression than the TEV constructs which lead to needing a longer exposition time during imaging to observe the Ulp1-YfaL bands. Note that all uncropped WBs images can be found in Annexe III for transparency. In the end, all systems seem to present their full-sized bands showing proper expression. The 6 SDAs could then move on to a more functional evaluation of their performance using the next steps of the POC workflow (i.e., the cell fractionation, microscopy, and flow cytometry experiments) (Fig 10-15).

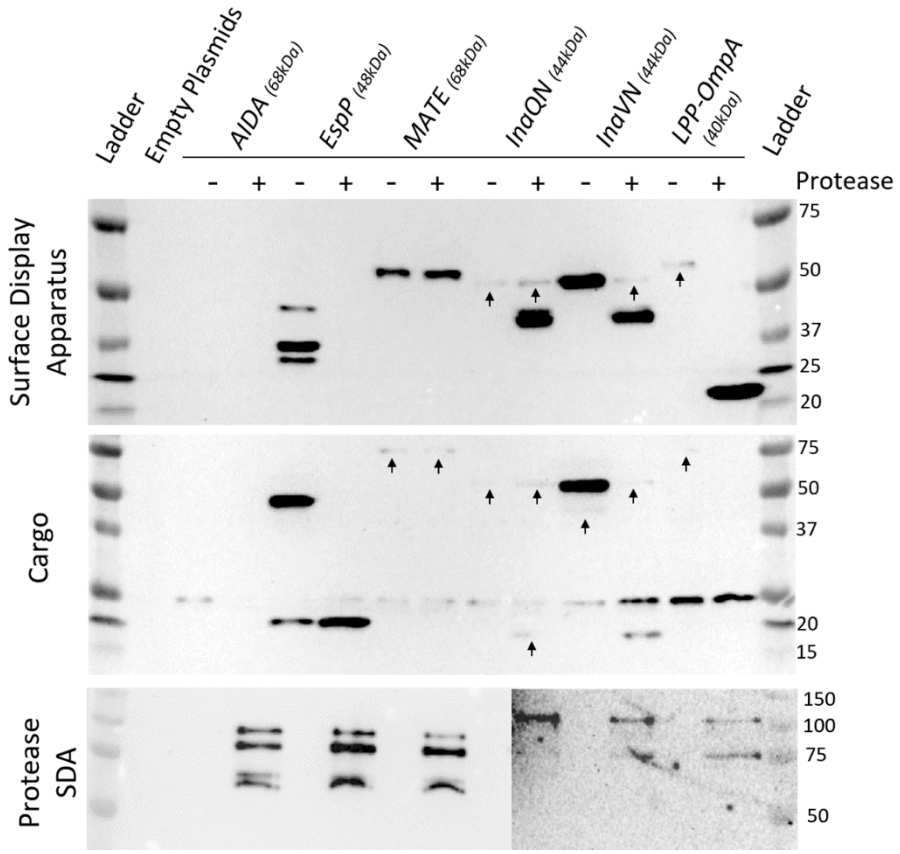


Figure 9. Expression of tested secretion systems. Three separate WBs of whole cell lysates with media of bacteria incubated at 37 °C up to an OD_{600} of ~2.0. Mouse anti-myc and anti-FLAG antibodies were used to follow the surface display apparatuses and the cargos respectively while a rat anti-ollas antibody was used to follow the proteases. Anti-mouse and anti-rat HRP-coupled secondary antibodies were used for imaging. Arrows point to bands of interest that only appeared faintly. Numbers on the right show the weights of the ladder in kDa. Molecular weights written by each SDA labels represent the theoretical size of their full lengths.

Table 3. Molecular weights of relevant full and partial constructs. The theoretical and observed molecular weights for each protein are listed. The full-size, translocated SDAs correspond to the full construct sizes without the signal peptide and are indicated by black arrow heads in WB figures. The cleaved, translocated SDAs correspond to the size of the TMDs of the constructs without the cargo and are indicated by blue arrow heads in WB figures. The cleaved cargos correspond to the sizes of the released cargos and are indicated by red arrow heads in WB figures. The full-size, protease SDAs correspond to the full constructs carrying the proteases without the signal peptide and are indicated by green arrow heads in WB figures.

Band	SDA	Theoretical MW (kDa)	Observed MW (kDa)
Full-size, translocated SDA	AIDA	68	N/A
	EspP	48	~48
	MATE	68	~60
	InaQN	44	~55
	InaVN	44	~55
	Lpp-OmpA	40	~60
Cleaved, translocated SDA	AIDA	54	N/A
	EspP	34	~34
	MATE	54	N/A
	InaQN	31	~42
	InaVN	31	~42
	Lpp-OmpA	27	~47
Cleaved Cargo	AIDA	14	N/A
	EspP	14	~18-21
	MATE	14	~18-21
	InaQN	13	~13-19
	InaVN	13	~13-19
	Lpp-OmpA	13	N/A
Full-size, protease SDA	YfaL (TEV)	79	~79
	YfaL (Ulp1)	77	~77

Starting with the AIDA system (Fig 10), although no expression could be detected for the SDA, faint bands can be seen on the cargo WB (Fig 10A). These band were, however, observed on other WBs such as the Lpp-OmpA cargo WB results (Fig 12A) and were identified as being background. This further suggests an issue with the stability of the expressed system. Note that the thick band at 25 kDa is also background. The microscopy imaging seems to suggest that the weak amount of expressed protein seem to stay stuck inside the cells and never reach the outer surface. Indeed, barely any signal is detected on the outer surface of the bacteria (Fig 10B, C) while, a noticeable amount of signal can be found when looking past the outer membrane (Fig 10D, E). We can conclude that an expression and stability issue is most likely causing the results observed for AIDA. The TEV SDA, however, can be seen successfully exported to the outer membrane (Fig 10A). All degradation bands are also reaching the membrane.

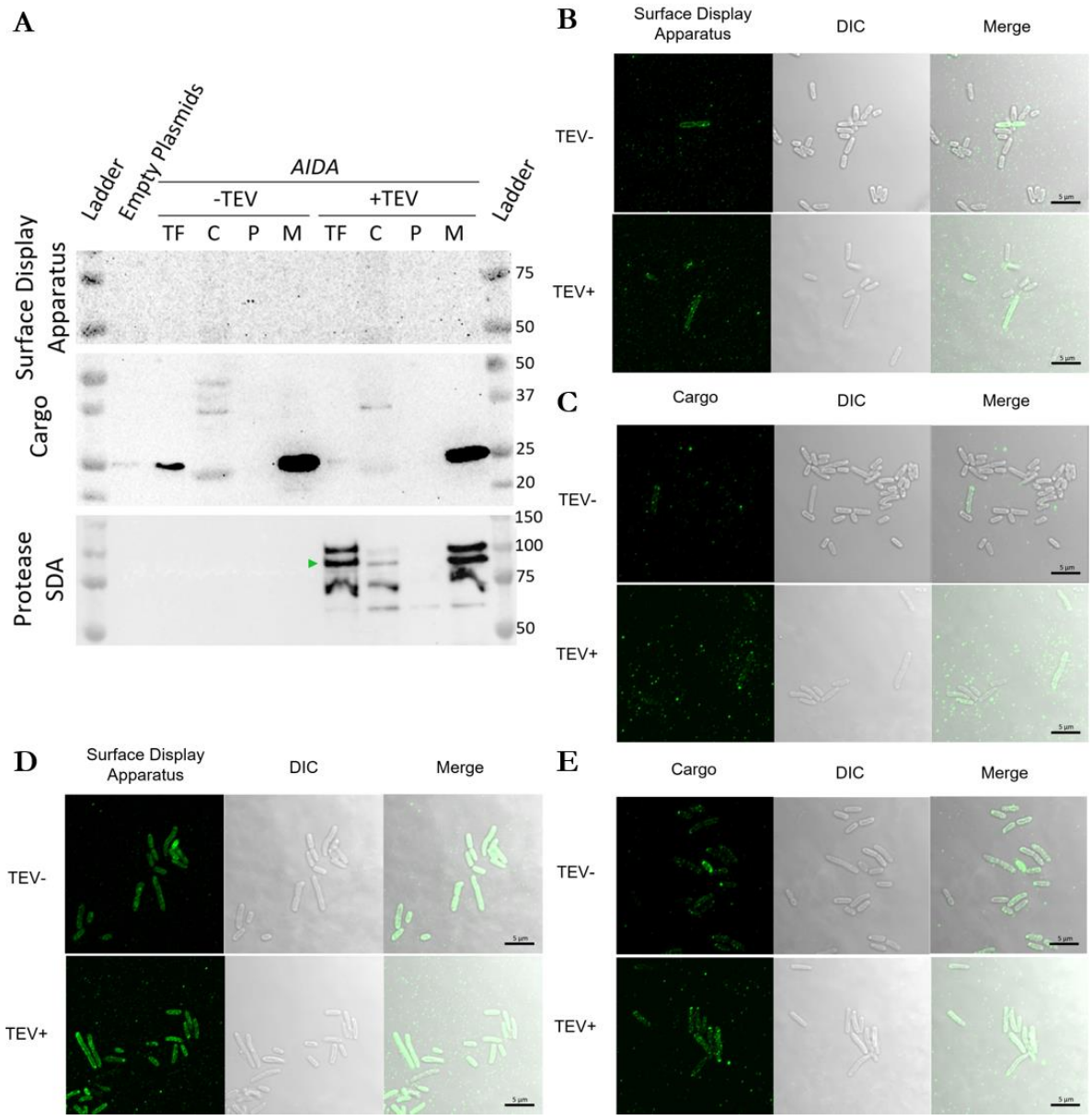


Figure 10. Functional study of the AIDA/TEV secretion system. A) Three separate WBs of cell fractions. TF = Whole cell lysate with media, C = Cytoplasm, P = Periplasm, M = Membranes. Bacteria were grown at 37 °C up to an OD_{600} of ~2.0. The periplasmic fraction was extracted by outer membrane permeabilization before lysis by sonication. Cytoplasmic and membrane fractions were separated by ultracentrifugation. A rat and a mouse HRP-coupled secondary antibody were used for imaging. The green arrowhead points at the full-size protease SDA (~79 kDa). Numbers on the right side show the weights of the ladder in kDa. **B-E)** Fluorescence microscopy imaging of indirectly stained fixed bacteria initially grown to an OD_{600} of ~1. In **D)** and **E)**, the bacteria were permeabilized by lysozyme and EtOH treatment to allow penetration of the antibodies. An AF488 secondary antibody was used for fluorescence. In all the above, mouse anti-myc (surface display apparatus), mouse anti-FLAG (cargo) and rat anti-ollas (protease) primary antibodies were used when

applicable. N.B.: This figure description is valid for figures 10-14 and supplementary figure 2.

Next, we evaluate the EspP system (Fig 11). The cell fractionation (Fig 11A) shows that, without TEV expression, the display apparatus and its cargo reach the OM with little remaining in the other compartments. As with the expression WB, we can see the full-sized band (~48 kDa, black arrowhead) reaching to MB fraction with an almost 100% efficiency. The uncleaved cargo also follows this same trend (~48 kDa, black arrowhead). With TEV present, however, none of our construct reaches the MB and the SDA seems to get completely degraded. There is also a periplasmic accumulation of cleaved cargo as seen in the cargo WB (~18 kDa, red arrowhead). This result continued to suggest something going wrong with the cleaving mechanism of our proposed system. The microscopy results are consistent with the aforementioned observations (Fig 11B, C): without the protease, we see that both the display system and its cargo is well exposed on the outer surface while, with the expression of TEV, there is an almost complete loss of signal in both cases. The permeabilized cells (Fig 11D, E) show us this periplasmic cleaved cargo in the presence of TEV that is concentrated at the poles of the bacterium with the SDA seeming to be spread within the periplasm and cytoplasm. In this case we can infer that the cleaving is occurring prematurely in the cytoplasm and maybe even in the periplasm. The cleaved cargo would possess the SP needed to get exported in the periplasm where it would accumulate as observed. In Fig 11A we can also notice spontaneous cleaving happening without TEV as described in the expression WB (Fig 9). This is shown by the presence of unidentified secondary bands in the SDA WB among which a band seeming to belong to the cleaved SDA is observed (~34 kDa, blue arrowhead). Although capable of translocating and

displaying the cargo on *E. coli*, this system seems to suffer from premature cleaving as well as alternate spontaneous cleaving events. Finally, the protease system in Fig 11A behaved exactly the same as observed in the previous AIDA-containing strains (Fig 10A) and can be seen successfully exporting TEV to the outer membrane.

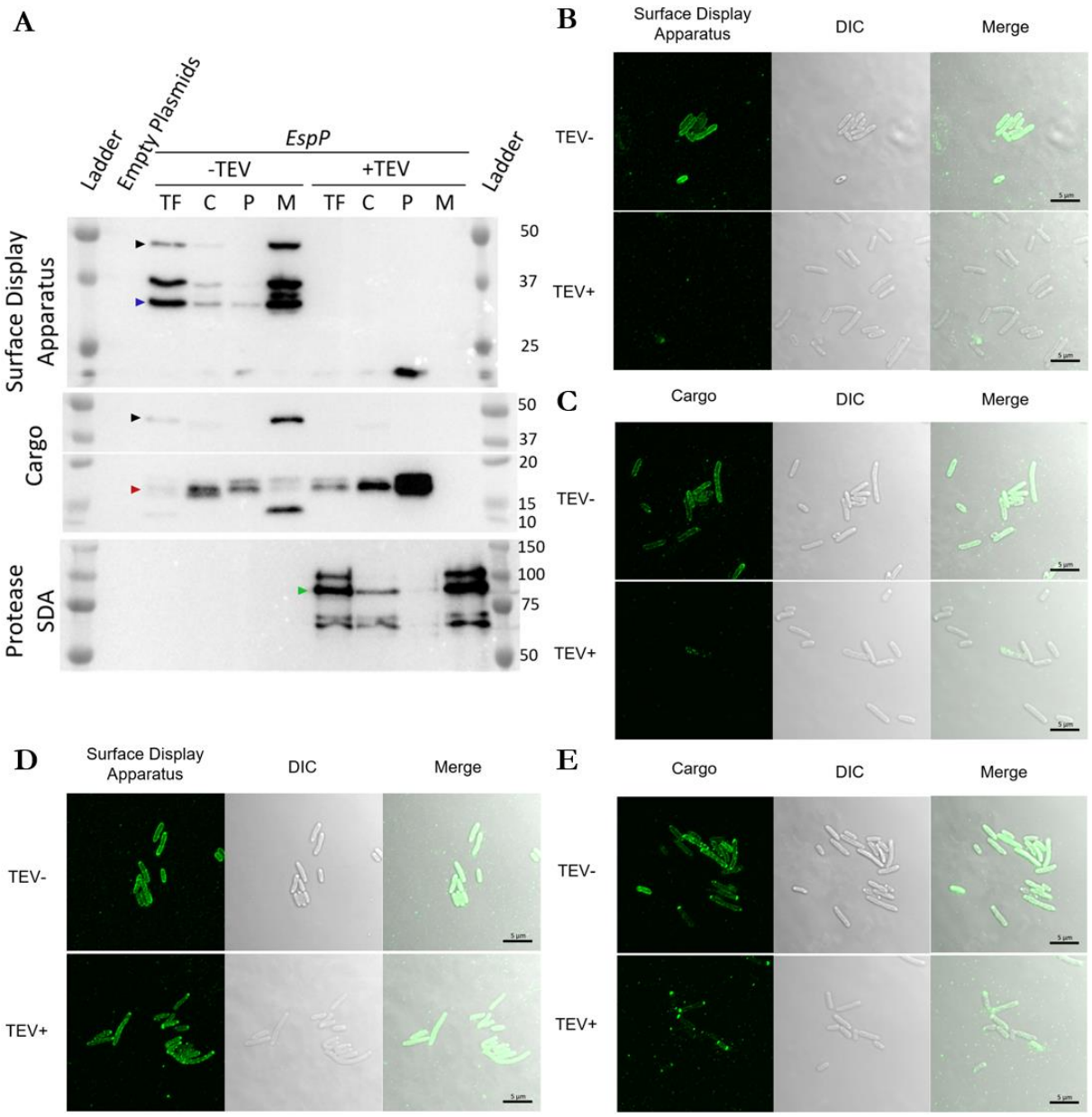


Figure 11. Functional study of the EspP/TEV secretion system. A) Three separate WBs of cell fractions. TF = Whole cell lysate with media, C = Cytoplasm, P = Periplasm, M = Membranes. The black arrowhead points at the full-sized SDA (~48 kDa). The blue arrowhead points at the cleaved TMD of the SDA (~34 kDa). The red arrowhead points at the cleaved cargo (~18 kDa). The green arrowhead points at the full-size protease SDA (~79 kDa). Numbers on the right side show the weights of the ladder in kDa. **B-E)** Fluorescence microscopy imaging of immunostained fixed. In **D)** and **E)**, the bacteria were permeabilized to allow penetration of the antibodies. See Fig. 10 for more details.

Similar results are seen with the LPP-OmpA/Ulp1 secretion system (Fig 12). In the SDA WB (Fig 12A), we can see the export of the SDA to the OM with almost 100% efficacy as displayed by the unique band around 60 kDa (black arrowhead). Unfortunately, the WB following the cargo does not show any remarkable bands with the ones present seeming to be background as seen on the AIDA cargo western blot (Fig 10A). This WB was consistent in not detecting any bands of interest through repeated experiments. In the presence of the protease Ulp1, we see a very faint band that appears in the MB fraction at the right height to be the cleaved TMD of the SDA (~47 kDa, blue arrowhead) hinting at a proper cleaving function. Unfortunately, this band is faint and a membrane-bound degradation product is favoured following cleaving. This product is visible at the very bottom of the SDA WB. The microscopy results seem to corroborate the results described above, as well as shed light on another issue. The SDA and cargo are being exposed at the surface without Ulp1, but this signal is lost in both cases once the protease is expressed (Fig 12B, C). Although we would expect to see the MB-associated degraded SDA still exposed at the OM when the protease is expressed, it is possible that the degraded product is not functional and is unable to translocate the cargo across the OM. Looking inside the cells (Fig 12D, E) in the presence of the protease, we can see that the degraded product seems more cytoplasmic than membrane associated as the cell fractionation suggested. That said, these results further the narrative that the proteases do not function as expected. As suggested previously, it might be that the protease is cleaving the cargo in the cytosol, causing the SDA to become unstable and get further degraded by cytoplasmic proteases. Since the cargo, in this case, would be C-terminal, the SDA would still have its signal peptide. Even degraded, it could reach the OM and anchor itself by its associated lipoprotein domain. However, the degraded OmpA may fail at integrating the membrane. Furthermore, in those images we observed the unnatural

elongation of the cells. This is the result of the stress caused by this particular SDA. This is in accordance with the earlier mentioned observations in literature where Lpp-OmpA often causes poor cell viability[109,110,115,116]. In the end, this system seems fully functional in *E. coli* but suffers from premature cleaving and seems to cause cell viability issues. To finish, the protease WB (Fig 12A) shows successful export of Ulp1 to the OM and behaves similarly to the TEV-carrying system (Fig 10A & 11A). This is not surprising as they both share the same AT (YfaL).

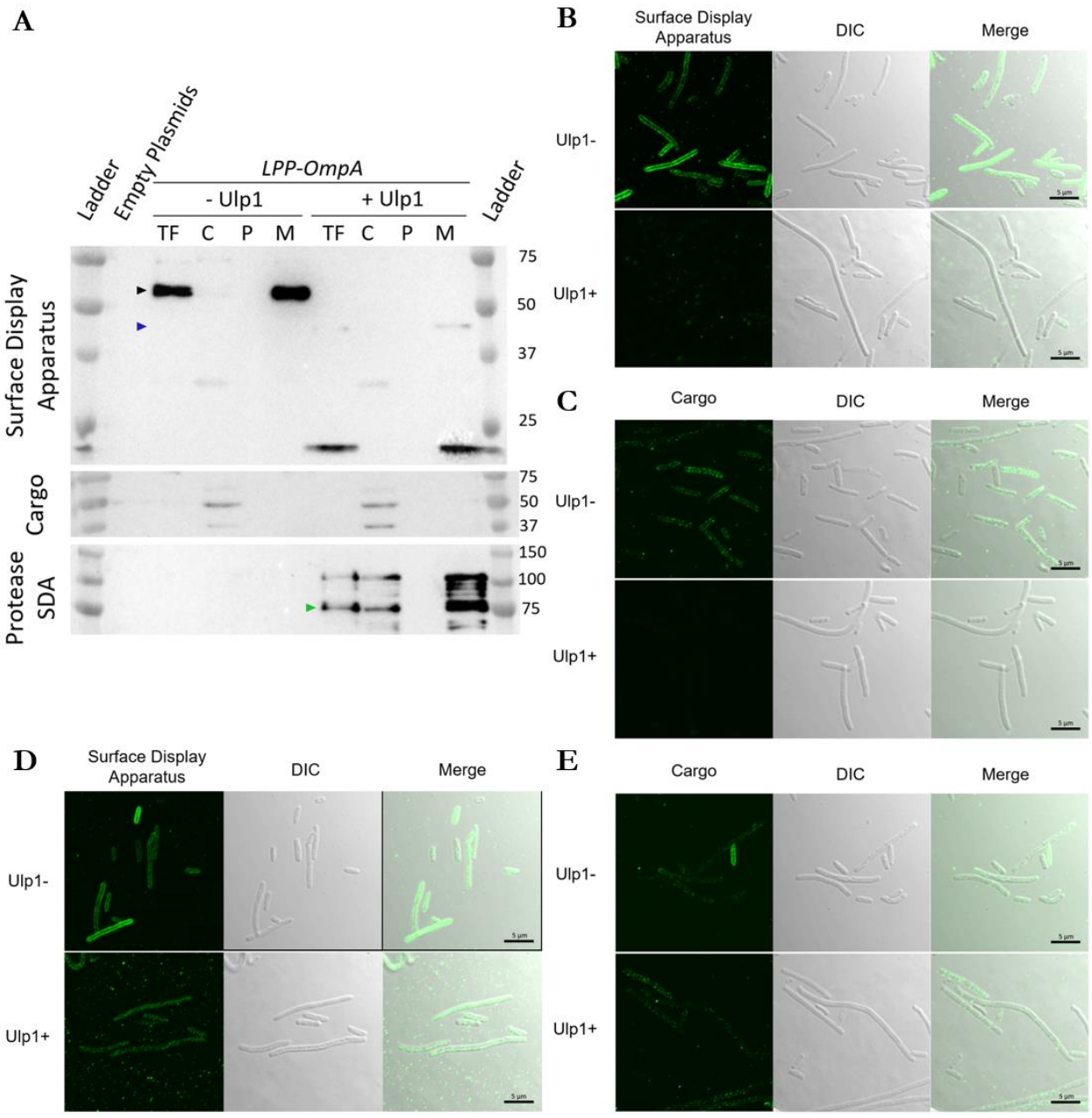


Figure 12. Functional study of the Lpp-OmpA/Ulp1 secretion system. A) Three separate WBs of cell fractions. TF = Whole cell lysate with media, C = Cytoplasm, P = Periplasm, M = Membranes. The black arrowhead points at the full-sized SDA (~60 kDa). The blue arrowhead points at the cleaved TMD of the SDA (~47 kDa). The green arrowhead points at the full-size protease SDA (~77 kDa). Numbers on the right side show the weights of the ladder in kDa. **B-E)** Fluorescence microscopy imaging of immunostained fixed. In **D)** and **E)**, the bacteria were permeabilized to allow penetration of the antibodies. See Fig. 10 for more details.

As mentioned in the previous cases, the INP systems also seem to present issues with their protease. Since the results of both INPs are functionally identical, only the InaVN results are shown here (Fig 13). The InaQN results can be found in Supplementary Figure 1. The WB (Fig 13A and Supp fig 1A) initially suggests proper functionality with the display apparatus and the cargo both reaching the outer membrane. We can see the band associated with the full-sized SDA reaching the membrane with a 60% efficiency in the SDA WB (~55 kDa, black arrowhead). This is mirrored in the cargo WB (~55 kDa, black arrowhead). Some faint unidentified degradation bands can be seen in the SDA WB but as they are also present in the cargo WB, they are more likely due to instability in the INP domain and not due to spontaneous cleaving of the cargo. However, with the expression of Ulp1, a new band appear at a lower kDa representing the cleaved SDA on its WB (~42 kDa, blue arrowhead). The cargo is also seen cleaved and remains in the cytoplasm (~13 kDa, red arrowhead). It is noted that the full-size band is still visible in the presence of Ulp1 although it is much fainter. This faint full-sized band is also mirrored in the cargo WB which further demonstrates that the cleaving is not 100% efficient. Moreover, the WB observations seem to suggest a premature cleaving of the cargo happening in the cytoplasm as seemed to be the case for both EspP and Lpp-OmpA (Fig 11-12). Indeed, the localization of the cleaved cargo in the cytoplasm suggests premature cleaving of the cargo. The proteases used, both TEV and Ulp1, are soluble and it is entirely possible that these could be functional in the cytoplasm as they are being expressed. This could lead to the premature processing of the SDAs before they can begin their journey to transport their cargo. The premature cleaving of their cargo can cause the now unstable SDAs to disappear in inclusion bodies which are discarded during the cell fractionation protocol. This is especially true for EspP where cleaving the cargo removes the signal peptide, leaving EspP stranded in the cytoplasm which

is unfavourable for a hydrophobic membrane protein. Now looking at the microscopy results (Fig 13B-E and Supp fig 1B-E), a unique issue seems to be plaguing the INPs. Although the blot suggests a good amount of the SDAs are reaching the membrane, absolutely no fluorescence is seen on the exterior surface of the cells (Fig 13B, C and Supp fig 1B, C). In the contrary, some signal can be seen when looking inside of those cells (Fig 13D, E and Supp fig 1D, E). There can be a rather simple explanation for this; it might be that, while the SDA reaches the outer membrane successfully, the cargo itself is not being translocated out of the periplasm. This could be due to the incompatibility of the cargo with the chosen SDA or issues with the sequence of the SDA and its compatibility with the host. Since the mechanism of translocation and the pathway by which INPs reach the outer membrane is mostly unknown[208], it is hard to tell what could be wrong in this case. Continuing, the microscopy of the INPs show elongation of the cells as observed in Lpp-OmpA but much less intensely. Briefly, these two INP SDAs present functional issues with the presentation of the cargo on the outer surface of *E. coli*, suffer from premature cleaving and display some level of toxicity. Finally, as with Lpp-OmpA (Fig 12A), the protease SDA in Fig 13A seems to function the same and reaches the MB successfully.

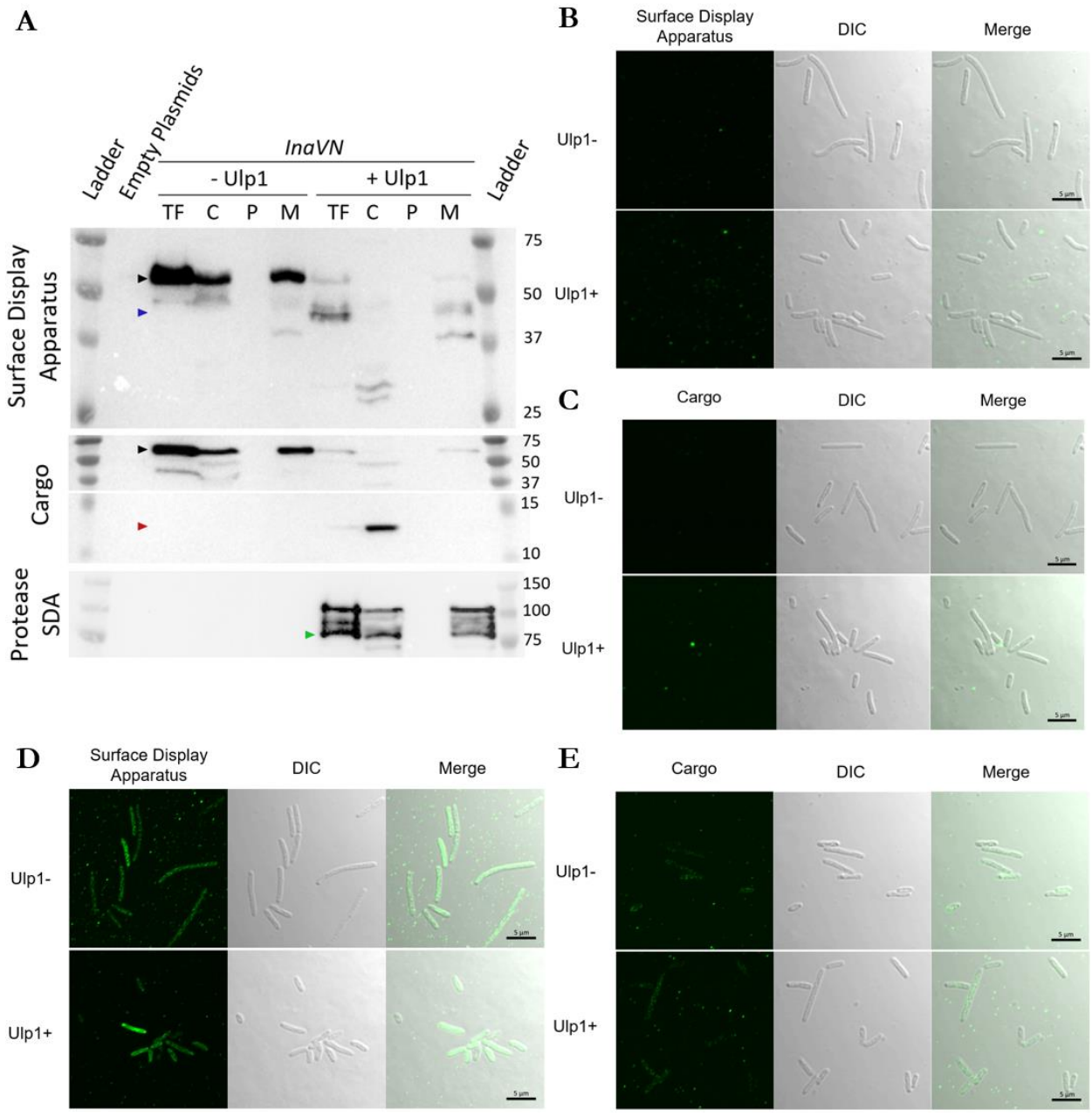


Figure 13. Functional study of the InaVN/Ulp1 secretion system. A) Three separate WBs of cell fractions. TF = Whole cell lysate with media, C = Cytoplasm, P = Periplasm, M = Membranes. The black arrowhead points at the full-sized SDA (~55 kDa). The blue arrowhead points at the cleaved TMD of the SDA (~42 kDa). The red arrowhead points at the cleaved cargo (~13 kDa). The green arrowhead points at the full-size protease SDA (~77 kDa). Numbers on the right side show the weights of the ladder in kDa. **B-E)** Fluorescence microscopy imaging of immunostained fixed. In **D)** and **E)**, the bacteria were permeabilized to allow penetration of the antibodies. See Fig. 10 for more details.

Moving on to the last of the secretion systems tested, MATE also presents its own unique issue. Unlike the others, there seems to be no functional difference with and without the expression of the protease. In both cases, the SDA reaches the outer membrane very efficiently without any degradation happening as shown in the SDA WB by the clear bands around 60 kDa (Fig 14A, black arrowhead). The same can be said for the cargo; there seem to be no difference in the presence or absence of the protease and no cleavage product is observed with the cargo following the SDA to the outer membrane properly in the WB (~79 kDa, black arrowhead). The cytoplasmic band in the cargo WB was identified as background. The protease WB shows the same functional export of TEV as seen in the EspP and AIDA WBs (Fig 10A and 11A). Consistent with the cell fractionation observations, both the SDA and its cargo are successfully displayed on the surface of the bacterium regardless of TEV's expression (Fig 14B, C). Even when looking inside the bacterium, we can see that most of the signal comes from the outer membrane (Fig 14D, E). The problem here is the clear lack of protease activity. This could be explained if the target site of TEV were to be inaccessible. It is possible that the spacing sequence linking the TMD of MATE and its cargo is too short, causing the cut site to be tucked in too tight between the hydrophobic domain and the globular CtxB cargo. In the end, MATE was proven to be functional although the cleaving site seems inaccessible to TEV.

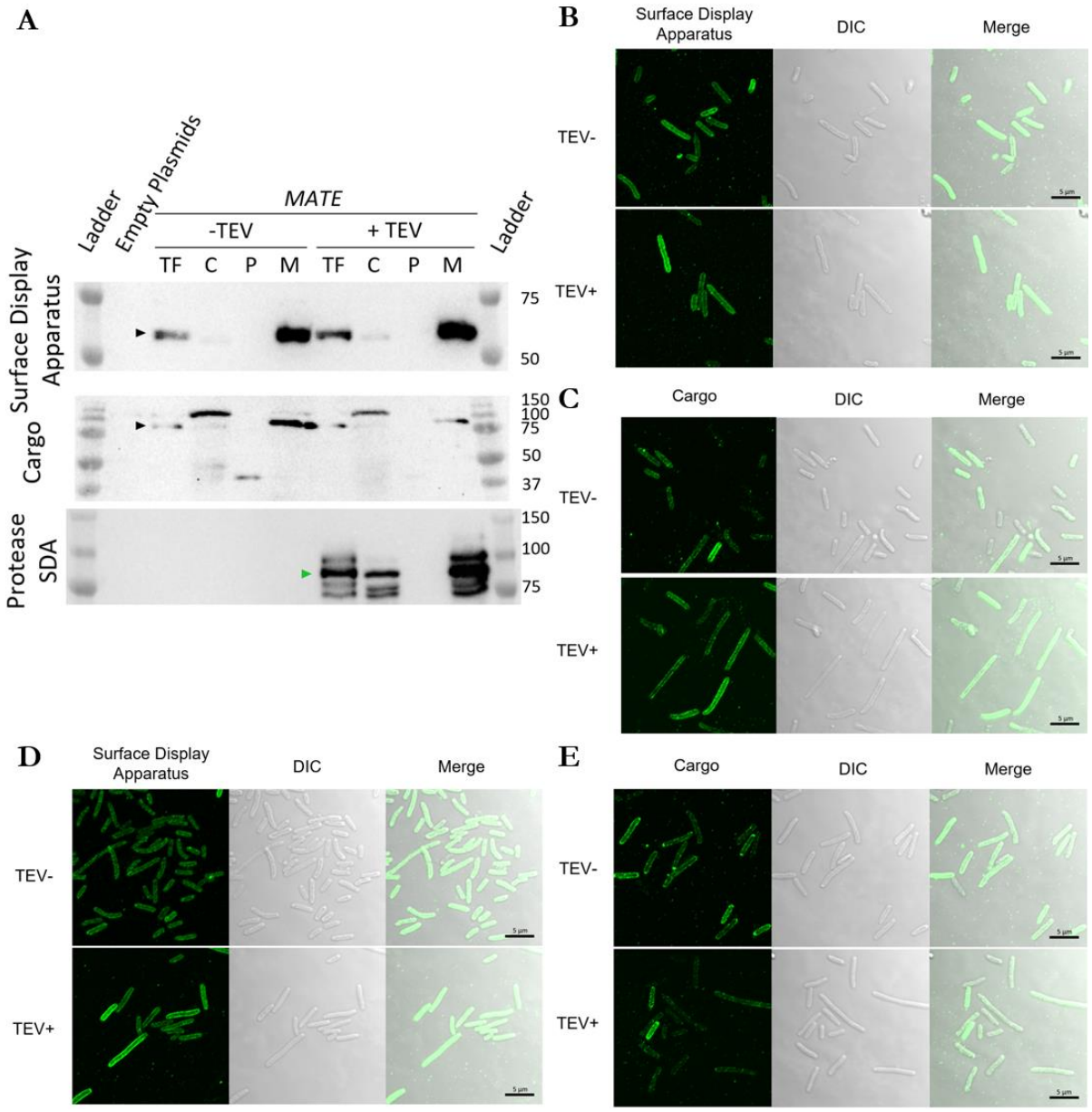


Figure 14. Functional study of the MATE/TEV secretion system. A) Three separate WBs of cell fractions. TF = Whole cell lysate with media, C = Cytoplasm, P = Periplasm, M = Membranes. The black arrowhead points at the full-sized SDA (~60 kDa). The green arrowhead points at the full-size protease SDA (~79 kDa). Numbers on the right side show the weights of the ladder in kDa. **B-E)** Fluorescence microscopy imaging of immunostained fixed. In **D)** and **E)**, the bacteria were permeabilized to allow penetration of the antibodies. See Fig. 10 for more details.

Finally, the flow cytometry results are consistent with the observed fluorescence microscopy images (Fig 10-14B-E), and they were also able to compare the different surface display levels of each apparatus (Fig 15). AIDA and the two INPs show no signal due to the expression problem and the translocation problem respectively. EspP and Lpp-OmpA show the same pattern with successful surface display without the protease but complete loss of signal for both the SDA and cargo when the protease is expressed. MATE shows the same level of surface exposition with and without the protease as described previously. For the three observable system, the display efficiency of EspP was seen to outmatch both MATE and Lpp-OmpA. The latter has the lowest display efficiency. The poor performance of Lpp-OmpA has been tied to the promoter strength, where higher expression triggers the formation of inclusion bodies and failure to reach the outer membrane[115–118].

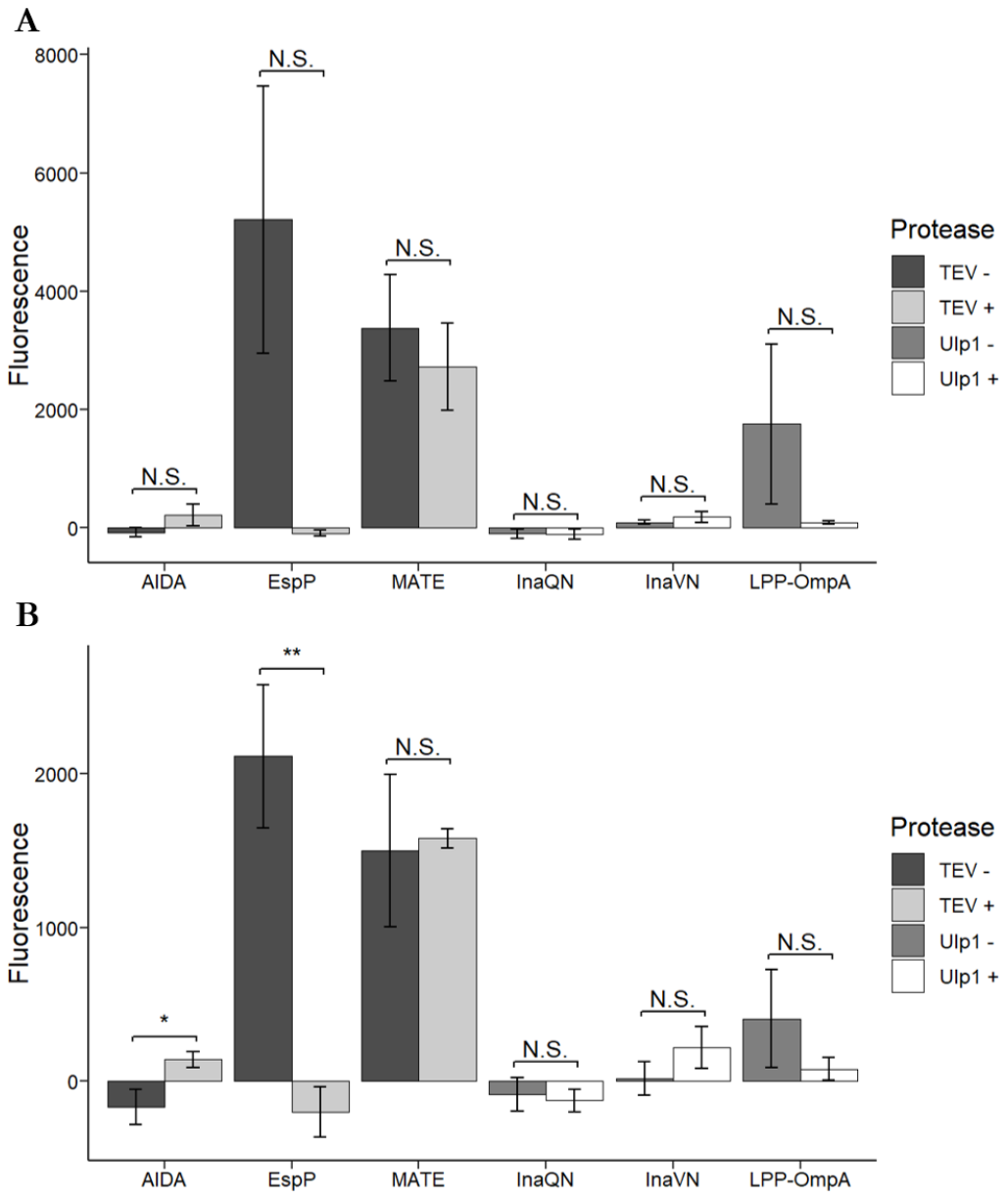


Figure 15. Comparison of the surface exposition levels of the tested constructs. Flow cytometry on indirectly stained fixed cells initially grown to an OD_{600} ~1. A mouse anti-myc (surface display apparatus) and a mouse anti-FLAG (cargo) primary antibody were used along with an AF488 fluorescent secondary antibody. **A)** Surface display apparatus exposition **B)** Cargo exposition. Results presented are representative of the median fluorescence values of 3 separate experiments. A paired t-test was used for statistical significance. P-value notation: * $0.05 > p > 0.01$, ** $0.01 > p > 0.001$ and N.S.: $p > 0.05$.

As a final test, we looked at the accessibility of the proteases to our surface displayed constructs to see if cleaving at the OM was possible if the protease expression and localization were to be coordinated properly with that of the SDA. We treated the three constructs that showed proper surface display with their respective exogenous protease (Fig 16). Looking at the SDA WB, for EspP we can see clear signs of cleaving by TEV. However, due to the spontaneous cleaving described previously for this construct, it seems that the pre-cleaved SDAs are favoured by TEV over the full-sized constructs. Indeed, in the presence of TEV, we can see the disappearance of the partially cleaved unidentified bands that were around 37 kDa and the remaining faint presence of the full-sized band (~48 kDa, black arrowhead). We also see the appearance of the fully cleaved TMD of EspP around 34 kDa (blue arrowhead). These results allow us to further confirm the localization of the spontaneous cleaving: since the TEV-cleaved SDA is of lower molecular weight than all the other bands, we can infer that EspP is cleaving upstream of the TEV cut site, somewhere within the CtxB sequence. When looking at the cargo WBs for this construct, the spontaneous cleaving of the cargo seems to happen throughout its journey. Since we can see cleaved cargo bands (~18 kDa, red arrowhead) in the pellet, this means this cargo is contained inside the cells and was cleaved either in the cytoplasm or periplasm. However, we can also see successfully released cargo in the supernatant WB (~18 kDa, red arrowhead). Some of this cargo has accumulated due to EspP's spontaneous cleaving as shown in the lane without TEV but a clear increase of cleaved cargo is seen with the TEV-treated cells. This shows that cleaving of the cargo at the outer membrane is possible for this construct. It is notable that uncleaved cargo in the presence of TEV is also seen but the band seem slightly fainter than without TEV (~48 kDa, black arrowhead). This results further demonstrates the preference of TEV for the already partially cleaved SDA. This particularity of TEV observed

here can support the results seen for MATE. In fact, the bands on the SDA WB are of the same intensity with no lower molecular weight band observed (~60 kDa, black arrowhead). No cleaving seems to occur to the MATE construct which is consistent with the accessibility issue hypothesized earlier. Accessibility issues may also be involved for the EspP SDA although those are more aggravated in the MATE SDA to the point where cleaving is not at all observable for the latter. In the cargo WB, we unfortunately fail to see the band corresponding to the full-sized construct in the non-treated lane. However, this band was successfully visualised in a previous attempt (results not shown). Looking at the supernatant, for MATE in the presence of TEV, we can see a very small amount of cargo being cleaved (~18 kDa, red arrowhead). Again, this furthers the conclusion that accessibility is not optimized but cleaving at the OM is possible. For Lpp-OmpA, however, no cleaving activity by Ulp1 can be observed at the SDA level with the bands staying at the same height with no lower molecular weight bands in the Ulp1-treated lane (~60 kDa, black arrowhead). For its cargo WB, as with TEV, there was failure in observing any bands. For this construct, with these results alone, we could say that cleaving of the cargo at the OM for Lpp-OmpA might present accessibility issues. It is important to mention that the bands associated with the EspP SDA were much brighter than the others in previous experiments and the sample here have been somewhat over-diluted by a factor of 3X which is why the bands for EspP are faint. Again, all uncropped WBs can be found in Annexe III. With this, we can tentatively conclude that TEV has access to the surface displayed cargo and that MATE does suffer from an accessibility issue. This experiment will still be repeated and optimized in the hopes of obtaining more explicit results, especially for Lpp-OmpA.

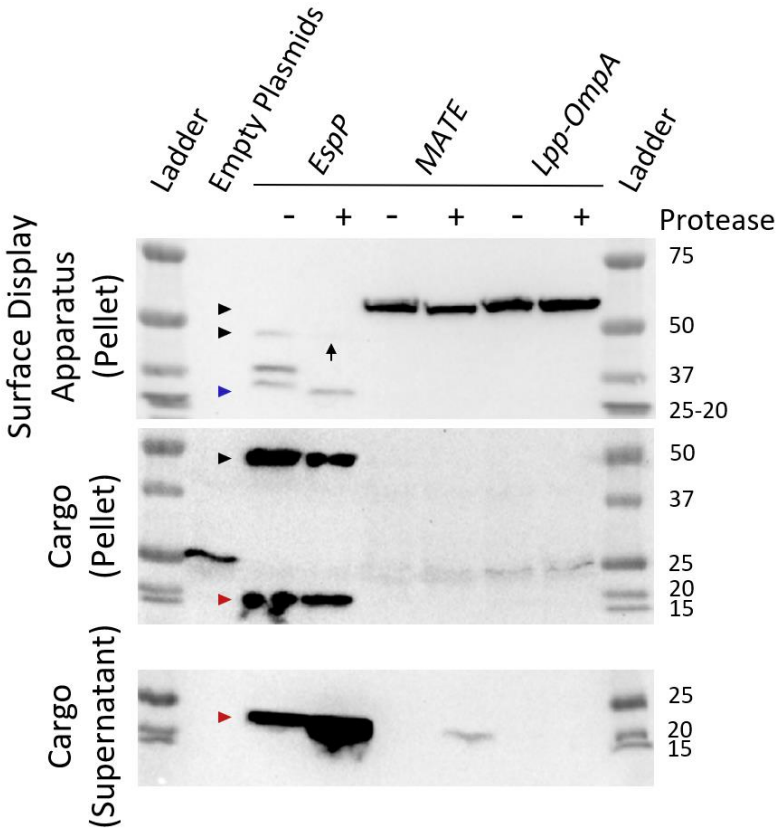


Figure 16. Accessibility of the proteases to the surface displayed cargo. Cells induced to an OD_{600} = ~1.0 were concentrated 50X and treated with commercial TEV or Ulp1 for 1 h at 30 °C. Whole cells were separated from the supernatant and the supernatant proteins were precipitated with TCA prior to analysis by WB. An anti-mouse HRP antibody was used for imaging. The arrow points to a band of interest that only appeared faintly. The black arrowheads points at the full-sized SDAs (MATE & Lpp-OmpA: ~60 kDa; EspP: ~48 kDa). The blue arrowhead points at the cleaved TMD of EspP (~34 kDa). The red arrowhead points at the cleaved cargo (~18 kDa). Numbers on the right show the weights of the ladder in kDa.

3 Chapter II – Antimicrobial peptide clustering

3.1 Methods:

3.1.1 AMP database

The database used in the project is the DRAMP 2.0 AMP database by Kang et al.[14]. It has since been updated to its 3.0 version[11] but the contents used in this study were from its 2.0 iteration. The database contents were directly downloaded in its .xlsx form as it was conveniently made available on their website: http://dramp.cpu-bioinfor.org/downloads/. The database was downloaded in parts separated by AMP activity. The separate files were combined into one single database file and any duplicates or entries with unknown sequences were removed. The sequences were then converted into FASTA format.

3.1.2 Linclust clustering algorithm

MMSeq2's Linclust clusters together short peptides as well as datasets with large size variations[186]. Linclust first extracts m number of k-mers from each sequence. Each k-mer forms a group where all sequences sharing said k-mer are included. The longest sequence in each of these groups is deemed a central sequence and groups sharing the same central sequence are merged together. All sequences in the groups are then aligned to their central sequences by both global Hamming distance and by local gapless alignment. The Hamming distance measures the minimum number of substitutions needed to transform one sequence into the other. The global alignment is made so that this number is as small as possible. This first alignment score is used to remove sequence associations below a certain cut off value. The ones above this value are realigned to the central sequence using gapped local alignment and the sequences that satisfy established criteria (E-value, sequence identity, sequence

coverage) are edge-linked with their central sequence. A greedy algorithm recruits the sequences into their final clusters with a given cluster representative.

The coverage mode 1 was used where coverage is measured by how much the query (central sequence) covers the target sequences as the query should be the longest sequence. The E-value uses gap-corrected Karlin-Altschul statistics to be computed and has a default value of 0.001. The minimum coverage according to the coverage mode chosen has a default setting of 80% and the minimum sequence identity has a default value of 90%. The number of k-mer per sequence generated by default is 20. In the following experiments, the tested settings are as follows:

Table 4. Algorithm parameters used for AMP clustering through Linclust. List of tested parameters to achieve the lowest number of final clusters.

Run	K-mer/sequence	Min. sequence identity	Min. coverage	E-value
1	Default	Default	Default	Default
2	101	Default	Default	Default
3	101	0.5	Default	Default
4	101	0.5	0.5	Default
5	101	0.5	0.3	Default
6	101	0.5	0.5	0.005
7	101	0.5	0.5	0.01

3.2 Results:

In the hopes of eventually applying AMPSEC for AMP development the sequences available on the DRAMP AMP database[14] were expected to be clustered to enable a characterization of the target specificity of each generated clusters. The contents of the database first ran through MMSeq2's Linclust algorithm with the default settings in coverage

mode 1 (Table 4, Run 1; Fig 17A). Using these settings, 3185 out of 4996 sequences were not clustered and remained as singletons. The parameters were then modified to be increasingly inclusive. First, to maximize the sensitivity of the algorithm we used a number of k-mers equal to the length of the longest sequence in our dataset (101). We also lowered the minimum sequence identity to the lowest possible setting of 50%. Unfortunately, no matter how inclusive the other parameters of the algorithm were set to, it was seemingly impossible to cluster sequences in large groups (Table 5 and Supp fig 2). Run 5 had the highest cluster mean and lowest number of final clusters (Fig. 17B and Table 5). Our 4996 peptides could only be clustered in over 2000 clusters with the mean cluster size being of only ~2 (Fig 17B, Table 5). Such small clusters would not be useful in trying to build target-sequence correlations within them. Other parameters were tested to no avail and those graphical distribution of those are shown in supplementary figure 2.

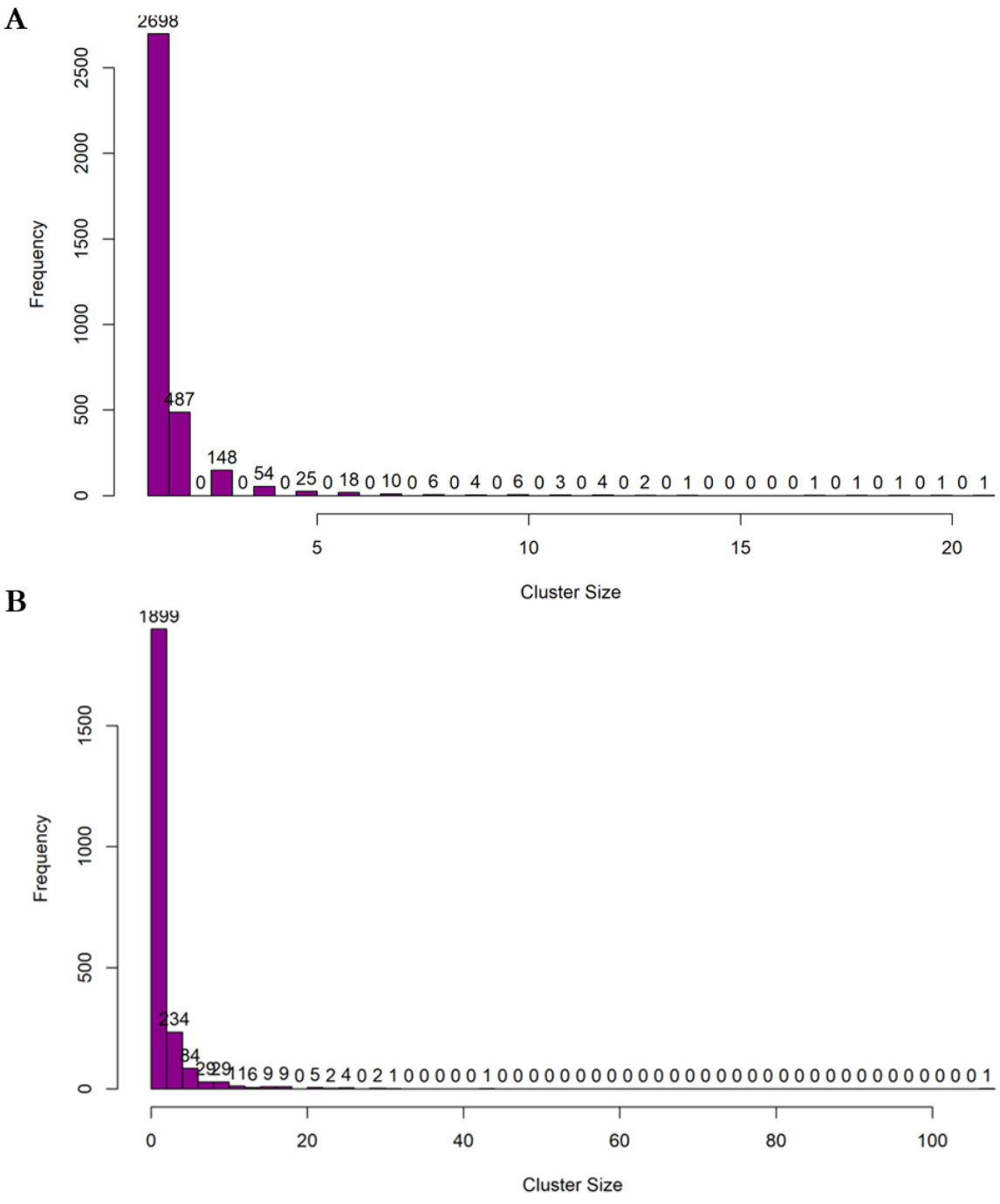


Figure 17. Outputted cluster size distribution using MMSeq2's Linclust peptide clustering algorithm. 4996 FASTA sequences were inputted using coverage mode 1. **A)** Run 1: The default parameters were used. **B)** Run 5: The parameters were adjusted to 101 k-mer per sequence, 50% minimum sequence identity, 50% minimum coverage, and an E-value threshold of 0.005.

Table 5. Statistics from the AMP clusters generated by Linclust. The 4996 sequences from the DRAMP database were run through MMSeq2's Linclust algorithm. This table lists the number of clusters outputted by the algorithm for each set of parameters used (Table 4) along with the mean number of sequences per cluster.

Run	Number of clusters	Mean cluster size
1	3472	1.439067
2	3457	1.445312
3	2364	2.113838
4	2334	2.14102
5	2327	2.147463
6	2334	2.14102
7	2334	2.14102

When investigating this, it was found that the likely cause was that a large part of the sequences contained in the database were lists of point mutated versions of the same parent peptide sequence. The algorithm must then be clustering these groups of almost identical sequences together and, because of the high sequence identity of these groups, it is likely that it is unable to combine them with peptides of similar properties but of vastly different sequences. Most clusters of over two members are often just those point mutated sequences grouped together, while unique peptides are singled out.

4 Discussion:

Herein, we made a first attempt at creating an independent peptide secretion system and explored a potential approach for the proteomics study of AMPs. Neither ended in the creation of a finalized product yet both paved the way to achieving that goal. Both the proposed secretion system and the study of AMP sequences bring complexities to the table that will be addressed in upcoming experiments. As for now, we can already begin to put together solutions for the issues brought to light in this study and explore alternatives.

4.1 Engineering a secretion system and its challenges

4.1.1 Solving the issues of our current systems

Modifying surface display technologies to create a secretion system is not a novel idea, however, making such a system in a way that is entirely independent on both the expression and cleaving levels is an entirely new challenge. The AIDA AT is a well-appreciated SDA and as it was readily available to our research group, it served as the first prototype. Including AIDA, most ATs possess autoproteolytic activity in order to cleave their native cargo[129]. The family of serine protease autotransporters of *Enterobacteriaceae* (SPATEs) all possess autoproteolytic activity in the same way as AIDA. The cleaving mechanism of other ATs sometimes involve a second partner. Hap is cleaved by other Hap AT, while IgaP, App and IcsA make use of a separate protease (NalP and IcsP). Our first idea was to harness this proteolytic activity to create an incredibly simple secretion system. The issue here lied in the cleaving site of the ATs, which are usually found within the linker sequence; the string of amino acids connecting the cargo to the TMD. This particular area was proven to be essential for the stability and functionality of the AT. For AIDA, for example, the minimum size of the linker region necessary for translocation is around 40

residues in contrast to the original linker size of 162 residues[132]. These 40 residues contain the cleaving site within them. For our system, however, the area of the linker that is found between the cargo and the cleaving site should be removed to limit the number of amino acids left attached to the cargo. This is especially important considering the nature of the cargos we are developing this technology for. AMPs are short and unstable so any extra amino acids added to its sequences may result in partial or complete loss of activity. With this, we attempted to create a functional secretion system with the partial removal of the linking sequence, however, this proved unsuccessful as the deletion rendered the SDA nonfunctional. This was not surprising as mentioned earlier, these deleted residues were shown to be essential by Maurer and Meyer[132].

Our next prototype increased the complexity of the system as we opted for using an endogenous OM protease for cleaving which allowed us to place the cut site right next to the cargo. Unfortunately, although this system proved functional in transporting and cleaving the cargo, it failed due to the nature of most OM proteases endogenous to gram-negative bacteria. Endogenous OM proteases are from the omptin family of proteases and they mostly target positive residues and do not possess a particular specific recognized cleaving site although some show preferences for particular amino acids[133,209]. Moreover, OmpT, the protease used in this study is known to play a role in the degradation of foreign AMPs. It is hypothesized that the secreted AMPs are being instantly degraded by the overexpressed OmpT. Despite this, it was proven that the creation of a fully functional AMP secretion system should be possible with perhaps a more careful selection of a protease for the cleaving step.

Following these results, we moved on to a more robust and thorough protocol for the

design of our secretion system. Six established SDAs for cargo secretion and a seventh for protease display were selected from literature. These systems were used as the transporters in our secretion system constructs and their performance was assessed using our POC workflow. A brief summary of the performance of each system can be found in Table 6.

Table 6. Summary of the new AMPSEC prototypes and at which stages they failed. A) The various steps of the secretion systems' functionality are horizontally listed. The table has been colour coded according to which step has been successfully executed by each system. Green means a successful observation of this step while red indicates a problem was met at the specified stage that stunted the progress to the next steps. **B)** Short description of why the systems failed at the red-coloured stage in Table 6A.

A

Type	SDS	DNA construct	Expression	Surface display	Cleaving
AT	AIDA				
	EspP				
	MATE				
OMP	eCPX				
	LPP-OmpA				
	InaQN				
	InaVN				

B

Type	SDS	Issues
AT	AIDA	Unsuccessful expression of the construct.
	EspP	Premature cleaving resulting in the loss of the signal peptide and spontaneous cleaving within the cargo.
	MATE	TEV was unable to access the cut site.
OMP	eCPX	Toxicity of the construct.
	LPP-OmpA	Premature cleaving leading to instability and degradation.
	InaQN	Issues with translocation of the cargo to the extracellular surface and premature cleaving.
	InaVN	Issues with translocation of the cargo to the extracellular surface and premature cleaving.

Premature cleaving of the cargo seems to be the biggest problem observed in the result section. To approach this problem, we first made sure that cleaving of the cargo at the surface of the cell is possible using a protease accessibility assay. As it was tentatively established that cleaving does occur on the bacterial surface, a future alternative for *in vitro* application of this project would be the use of an inducible promoter for the expression of the protease system. With this in place, both display systems could be contained inside the same bacterium and the cargo SDA can be given time to reach the outer membrane before induction of the protease SDA expression. If this proves successful, to render this dynamic system independent on induction supplements, an oscillatory promoter could be used for the expression of the protease. A good example of an oscillatory gene expression system would be the circadian rhythm of eukaryotes[210]. Such programs are also present in bacteria like the FtsZ and KaiABC gene regulation[211]. These, however, are very complex genetic networks involving a large number of metabolites and proteins making them difficult to exploit for biotechnological applications. Artificial oscillatory networks have thus been developed by research groups[212]. The Elowitz-Leibler oscillatory network is dubbed “repressilator” and consists of a negative feedback loop involving the lac repressor (LacI), the tet repressor (TetR) and the λ repressor (λ cI)[213]. In this system, a plasmid contains the tet promoter regulating the expression of the λcI protein, the λ promoter regulating the expression of LacI and the lac promoter regulating the expression of TetR. A separate reporter plasmid encodes the gene of interest regulated by the tet promoter. The half-life of the produced proteins was also shortened to be closer to that of mRNA by the addition of ssrA-based tag in C-terminal of the repressors, making them targets of intracellular proteases. This system was originally shown to produce oscillatory expression in only 40% of the cell population, however, this was disproven by more advanced cellular imaging technology where 100% of the cells were

seen demonstrating oscillating gene expression[214]. This same research group who re-evaluated the robustness of the repressilator also further optimized its functioning[214]. The improved system showed consistent oscillations for long periods in a variety of growth conditions. A different research group also refined the system by adding a quorum sensing component to help the cells synchronise[215]. A second oscillating expression system was dubbed the "metabolator" by Fung et al. as it uses the metabolism to create an oscillatory genetic expression of a reporter gene[216]. The acetate pathway was exploited where the production of acetyl-CoA from the metabolism of sugars and fatty acids is converted into acetyl phosphate by phosphate acetyltransferase (Pta) whose expression is repressed by the accumulation of acetyl phosphate. The accumulation of this metabolite also induces the production of acetyl-CoA synthetase to convert the acetyl phosphate back to acetyl-CoA. The acetyl phosphate is also exported out of the cells as acetate. As in the previous system, all the enzymes and repressors involved are fitted with a degradation tag to reduce their half-life. The oscillating gene of interest is expressed in a way that reflects the activation and inactivation cycle of Pta. This system proved more robust than the previous with 60% of the cells showing an oscillatory expression pattern with a cycle lasting about 45 minutes. This pattern is lost after 4 hours due to acetate accumulation in the media but oscillation could potentially be maintained longer by renewing the media periodically[216]. The above oscillators are the only fully independent oscillators that have been applied in bacteria. The latter having a more robust performance could be a candidate for regulating the expression of the protease display apparatus in this project.

For MATE, giving the construct a longer linker sequence may solve the accessibility issue of TEV. That said, once TEV gains access to the cut site, MATE would likely become

susceptible to the premature cleaving issue brought to light in the other constructs, suggesting that the solution above should also be applied to a cleavable MATE.

For the two INPs, it is difficult to tell what could be done to solve their lack of cargo translocation. This, however, isn't an unknown issue for the INPs as the lack of surface display has been noted in a few recent cases[90,217,218]. In the papers, the lack of surface exposition observed was hypothesized to be due to bad translocation, just as we suspect in this project[217,218]. Although the older articles seem to hint at the outstanding performance of INPs, none of those articles looked at surface exposition of the cargo but only looked at cellular localization and at the presence of the INP in the OM[115,123,128]. Considering the INP constructs are properly expressed and do localize at the outer membrane, it would not be surprising to assume a successful display is achieved with those results alone. Moreover, the only older article showing proper surface display evidence using INPs were expressing their systems in *Pseudomonas putidas*[124]. The INPs are native from *Pseudomonas spp.*, perhaps the expression of INPs in *E. coli* creates a functional incompatibility causing them to fail at translocating their cargo which could explain the difficulties observed. Although never used before, adding a spacer between the cargo and the INPs may help with compatibility. Other recent articles who seemed to successfully use INPs are using different INP genes[219,220]. Perhaps it would be worth exploring other INP options as well.

In EspP's case, the spontaneously cleaving phenotype observed may be worth fixing. EspP, like AIDA, does also possess autoproteolytic activity however, the catalytic residue in the former (D1120) was mutated from an Asp to Asn as described in Binder et al.[97], which was shown to inactivate the processing of its natural cargo[221]. If EspP is not responsible for this cleaving, OmpP could be responsible for this degradation effect. From the results, we

can approximate that the alternate cleaving sites are positioned near the TEV cut site as hinted by the presence of cleaved residues in the cargo WBs at around the same height as the expected cargo in Figures 9, 11A and 16. More concretely, however, this cleaving phenotype, was also observed by Binder et al.[97] and they hypothesised it being due to a XYF C-terminal motif that may interact with DegS, a stress-response chaperone protease of *E. coli*. This motif, unfortunately, was stated to be essential for the OM localization of the AT.

With AIDA, the observed outcome is surprising considering that this SDA was functioning in an exemplary fashion in the initial prototypes but is now completely dysfunctional. To figure which of the changes that were made to the construct suddenly caused these expression and functional troubles, we could try returning to the previous test cargo or even the previous promoter used in the initial prototypes. However, it is likely an incompatibility with the new promoter or the translation initiation sequence.

As for eCPX, troubleshooting at the level of the molecular biology will be needed to fix the issues encountered in the assembly and transformation of the plasmid.

Finally, the POC workflow established in this project thoroughly demonstrated its robustness in the consistency of the observed results. Across all experiments, the narratives were coherent and further supports the hypotheses put forward to explain the outcomes.

4.1.2 Alternatives

In the present thesis, only one of the many approaches was considered for the development of a secretion system[222]. For example, another type of protease we could have used is intervening proteins (inteins). These are self-excising precursor proteins that simultaneously ligate the sequences at their N and C terminals[223]. Inteins have been used for

protein purification, protein cyclization, protein tagging, and as biomarkers or biosensors[224]. Tomomi et al. ligated proteins at the surface of live CHO cells using the protein trans-splicing function of split inteins[225]. The C-terminal fragment of the intein was surface-expressed and the N-terminal fragment was exogenously added to the serum media. The two intein domains were reunited with the help of a small-molecule-ligand and its receptor. This assembly activated the splicing function of the intein which ligated the biotin protein that was located on the N-terminal end of the intein to the membrane-bound protein attached to the C-terminal end of the intein, effectively biotinylating the cell membrane. We can use a similar approach in this project; the desired AMP can be split in two and fused to the N and C terminal domains of a high affinity split intein[226]. These separate intein domain can be co-expressed and surface displayed using an N-terminal fusion transporter for the N-intein and a C-terminal transporter for the C-intein. The use of a flexible tether and the membrane localization of both intein domains would allow them to assemble and catalyze the ligation of our two AMP halves and the release of the functional AMP (Fig 18A). That said, this system could also be susceptible to cytoplasmic release of our AMP if the expression of both intein domains is not separated in time as observed with our EspP, Lpp-OmpA and INP constructs.

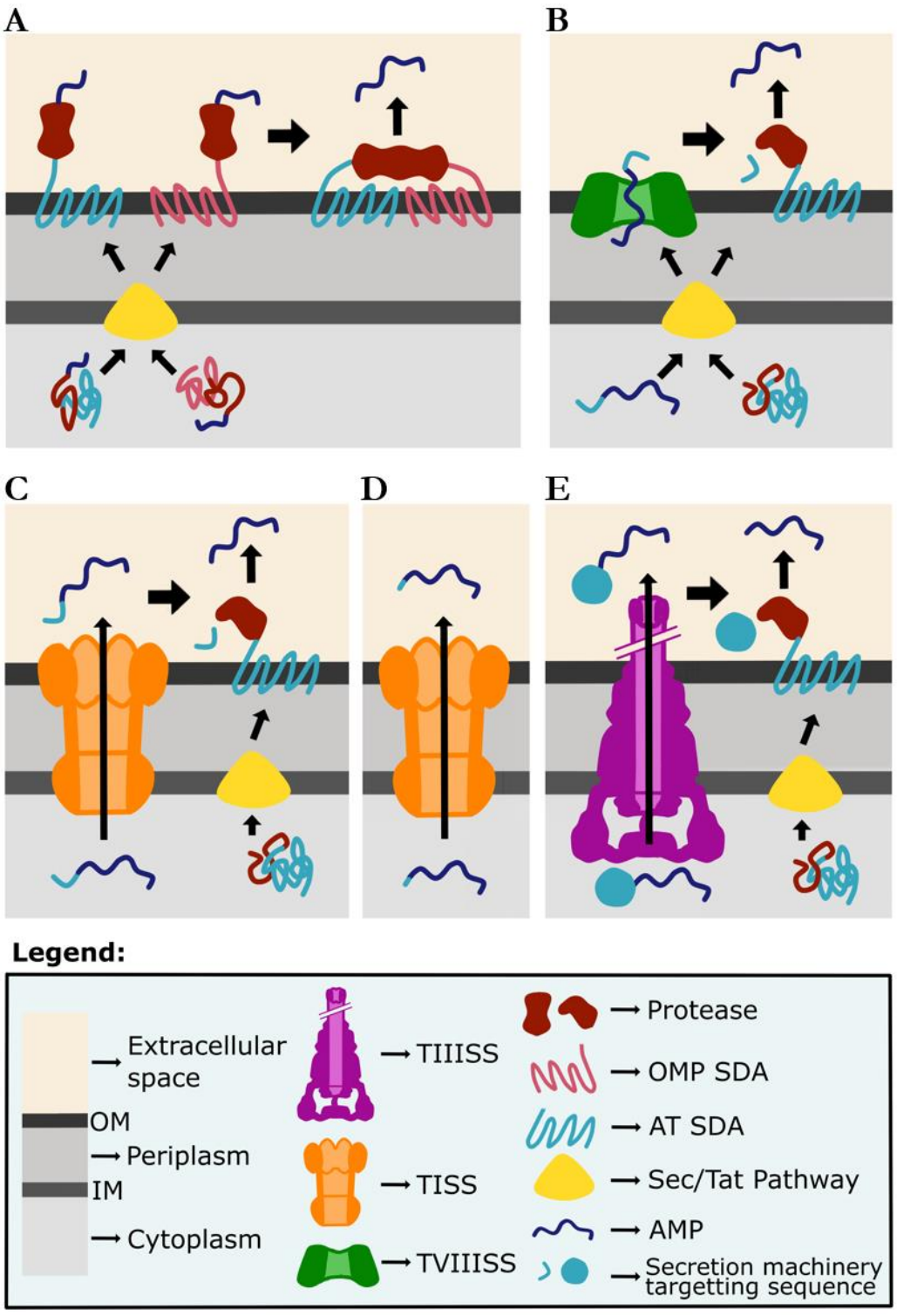


Figure 18. Alternative secretion pathways for the design of AMPSEC. A) Intein-mediated secretion using split inteins targeted to the OM by an AT and OMP SDA. **B)** TVIIISS-mediated secretion where the AMP must first use the SEC pathway before being secreted. The TVIIISS signal peptide then needs to be removed by a MB-associated protease

C) TISS-mediated secretion where the signal peptide necessitates cleaving by a MB-associated protease. **D)** TISS-mediated secretion where the signal peptide is of trivial length and does not demand for its removal. **E)** TIIISS-mediated secretion where the AMP is fused to a TIIISS substrate. This substrate would then be cleaved by a MB-associated protease.

Including the TVSS, gram-negative bacteria include secretion systems from type I to type IX[227,228]. Like the TVSS, the type II, type VIII and type IX secretion systems use the Sec/Tat pathways to pass through the IM. In contrast to the former, the TIISS is used to secrete folded proteins. Unfortunately, substrate recognition is poorly understood and no universal signal sequence has been identified[229] making this secretion pathway not ideal for its application in the present project. The TVIIISS, also known as the Curli secretion pathway, serves in the secretion of functional amyloids. The fibrous proteins are used in biofilm formation and to help with various interactions such as invasion, colonization, and immune response[230]. The secretion of fusion proteins through this pathway has been attempted and the system showed a clear preference for unfolded proteins no matter the length[231]. Secretion of an intrinsically unfolded protein ERD10 was successful while the secretion of FimC and mCherry, proteins of a similar size but which folds prior to secretion, was unsuccessful. For our application, large, folded AMPs will likely not be an issue without their target. This pathway, however, would require the addition of a secretion signal that would necessitate cleaving after secretion. A potential approach would be using the surface displayed Ulp1 protease along with the insertion of the SUMO sequence between our AMP and the N-terminal TVIIISS signal sequence (Fig 18B). Theoretically, upon secretion, Ulp1 would recognize the SUMO peptide and release the AMP.

Finally, the TIXSS found exclusively in the *Bacteroidetes* phylum serves for effector secretion including various toxins and motility proteins[228]. This system, however, was

established as a novel secretion system only in 2010[232]. Given its novelty, little is known about the nature of the substrate interactions with the system, much less about the recognition of its substrates.

In contrast to the above, the secretion systems I, III, IV, VI and VII transport their secretome directly from the cytoplasm through both the IM and OM in one step. The TISS is used by bacteria to secrete proteins associated with nutrient acquisition or virulence. For this system, the secretion signals responsible for substrate recognition have been well studied and are most often located within the last 60-50 residues in C-terminal and includes a glycine-rich sequence upstream (GGxGxDxUx)[233,234]. This signal, however, is not cleaved during or after secretion. The variety of TISS substrates is also an impressive factor, with them ranging from 19 kDa (HasA; *Serratia marcescens*)[235] to 520 kDa (LapA; *Pseudomonas fluorescens*)[236]. Work by Christian Hanke et al.[237] used the α-hemolysin (HlyA) signal sequence to secrete cholesterol esterase/lipase from *Pseudomonas spp.* into the growth media coupled with tag removal by the surface protease OmpT expressed by the *E. coli* host. HlyA is secreted by the unique mechanism of the HlyB/HlyD complex spanning both membranes. In order to use this technique for our purpose, however, we would need to change the OmpT protease for one of the specific proteases explored in this thesis. The TEV protease would be favoured in this case since the secretion signal that needs to be cleaved is in C-terminal (Fig 18C). Similarly, Asada Tomonori et al.[238] used the HasA signal sequence to secrete antibodies through the LipBCD TISS in BL21 *E. coli*. In this case, however, the signal sequence is so small, composed of 3 residues in C-terminal ("VTV"), that the necessity for a cleaving system may be avoided (Fig 18D). Even more convenient would be the TISS of gram-positive bacteria which can include the peptidase-containing ATP-binding cassette

transporter which cleaving the secretion signal[239]. This would, however, involve changing to a gram-positive host.

The TIIISS in found in pathogenic bacteria and serves to secrete effectors inside target eukaryotic cells. This system features a complex hierarchical secretion of substrates mediated by precise regulatory events and chaperones[240]. The native system also depends on target-cell membrane contact to activate the final steps of secretion. Moreover, the exact sequence that causes the recruitment of the substrate-carrying chaperones to the TIIISS is unknown. Although the fusion of a desired cargo peptide to one of the substrates of the TIIISS may allow for secretion, this fusion partner would then need to be cleaved (Fig 18E). It would also be necessary to use a constitutively secreting version of the TIIISS for this application using an IpaD or IpaB deficient TIIISS[241].

Similarly, to the above, the TVISS mediate toxin secretion inside target bacterial cells to compete. Secretion of effectors is dependent in the interaction of the substrate with core components of the spike and expelled tube[242]. Unfortunately, the residues responsible for these interactions are unknown. The activation of this system is based on quorum sensing and is triggered at a high population density[243]. Even with this, there are a lot of unknowns concerning the regulation of the TVISS and the selection of its substrates which makes this system unripe for the proposed application.

The TIVSS deals mainly with DNA secretion, but was also found to serve in toxin, transforming protein and intracellular life effector secretion[244]. So far, advances in the understanding of this secretion system have found that sequences on the C-proximal and C-terminal of the substrates are important for being directed through its machinery[244]. Although the nature of those signal domains have been studied, it is not know whether these sequences

are sufficient for recognition by the TIVSS[245–247]. To make use of this secretion pathway in this project, an effective signal sequence would first need to be identified that is compatible with non-native substrates.

Finally, the TVIISS is in a class of its own as it is exclusively encountered in mycobacteria or gram-positive species. In those species, it serves for the secretion of virulence factors[248]. As of now, there has been no attempts at exploiting this system while efforts are still being made to understand its functioning[249]. Although the sequence of the TVIISS substrates responsible for leading them to be secreted through this pathway has been studied[250], the system being exclusively mycobacterial would not be appropriate for this project. We would first need to consider if the machinery is transferable to non-mycobacterial gram-negative species and would implicate the expression of a complex multi-protein system[251].

4.2 Finding a different approach for clustering AMPs

Understanding AMPs is critical to the engineering of novel synthetic peptides that do not possess the drawbacks of natural AMPs. Herein we encountered the particularities of applying general proteomics to AMPs. In a first place, we were faced with an issue concerning the redundancy of AMP databases. The true diversity of the AMPs contained in the DRAMP database is less than expected due to the large quantity of “duplicates” corresponding to many point-mutated iterations of the same parent peptide. To fix this, those groups of peptides could be summarized together into a single consensus sequence. To do so, the sequences in those groups would be aligned using Clustal’s MSA, then a consensus sequence would be generated according to the most frequent amino acid identified at each position[252,253]. The resulting database of consensus sequences could be analyzed using

clustering algorithms to hopefully output larger and more inclusive clusters. These consensus sequences would, however, lead to another issue of sample size. By clustering these sequences together, it is estimated that the number of peptides in our database would be at least halved if not more considering the sheer number of repetitive sequences in all the groups that were sampled. In that case it might be worth trying to combine other AMP databases like the APD3[10] and the DBAASP[9] database to the DRAMP database.

In the end, since this type of clustering method has not been applied to AMPs, it is not surprising that identifying established algorithms and resources to accomplish these goals is challenging. To this end, perhaps modifications would need to be applied to the Linclust algorithm to adapt it better to the kind of dataset that we are trying to study. Firstly, when first generating the k-mer groups, a heuristic approach could be used to combine k-mer groups comprised of only one peptide to the groups whose k-mer is the most similar to the k-mer of that single peptide group. This could be useful as the size distribution of the AMPs in our database (Fig 19) is quite large and the sample size (4996 AMPs) is relatively small compared to usual proteomics datasets. Without this modification, peptides that do not share a k-mer with any other peptides would be immediately singled out and would not end up being clustered.

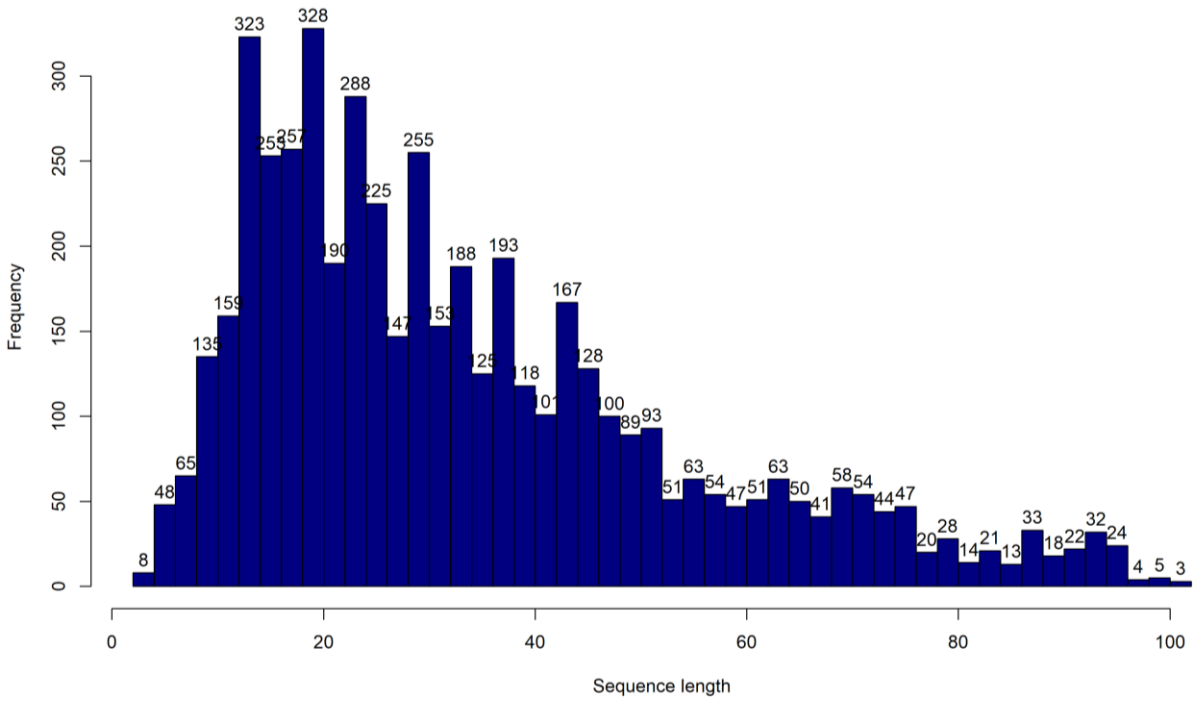


Figure 19. Size distribution of the AMPs contained in DRAMP. The sequence lengths of all the 4996 AMPs were plotted on a histogram.

The final clustering approach used by Linclust could also be modified to suit the type of peptides studied in this project. They first use ungapped local alignment to get rid of sequence associations below a certain cut-off point before moving on to a gapped local alignment in order to obtain the final clusters. It could be argued that ungapped alignments are not suitable for AMPs. Since AMPs are small and non-globular, there would be more value to gapped alignments where the presence of common active domains is favoured over the contextual emplacement of said shared active domains. The non-globular and flexible nature of AMPs makes it unlikely that an active domain would be inactive due to its position within an AMP's sequence.

Alternatively, another promising algorithm by Melman and Roshan[254] that did not rely on sequence alignment used a k-means approach to cluster proteins. Briefly, the peptides

are divided into fragments (k-mers) and an m number of centroid sequences are picked randomly. Next, all fragments are compared against the centroids and are assigned to their closest centroid. The centroid sequence representing each group would then be re-evaluated. The assignment process would repeat until the centroid sequences remain unchanged when re-evaluated. This deep k-means approach unfortunately showed reduced accuracy with smaller proteins, which does not make it suitable for the clustering of AMPs.

Lastly, a novel peptide clustering algorithm could be designed using word embedding. This technique aims to represent strings in a context-dependent vector space[255]. Feature learning is used with an artificial neural network to establish the number of dimensions per strings needed to embed all the properties of the training set to a continuous vector space of lower dimensions. This technique originally used for language learning has since been applied to nucleotide sequences and amino acid sequences[256]. This technique not only accurately represents the properties of these biological sequences but also takes the context (i.e., the properties of the surrounding nucleotides or amino acids) into consideration. These vectors can then be used for complex machine learning analyses such as clustering. In order to embed them, protein sequences can be split into k-mers which are considered to be the "words" that make up the whole sequence ("sentence"). These words are then encoded into an artificial neural network-based vector space. With this, k-mers with similar properties and used in similar contexts are assigned similar positions in the vector space. This process greatly facilitates the clustering of sequences. The Sequence Graph Transform (SGT) function[257] for sequence embedding is particularly attractive for this project as it corrects for the false positive spatial associations caused by sequence length variations in the dataset. The

data representation created by this function is especially useful for clustering approaches and was demonstrated with a k-means approach[257].

4.3 In conclusion

Herein, we accomplished the experimental design of a thorough and consistent proof-of-concept workflow for the design of our proposed AMPSEC. We explored the potential of various surface display technologies and were met with unique roadblocks for each of the design that will necessitate optimization in the hopes of developing a finalized product. EspP, Lpp-OmpA, and MATE all showed great potential as they could all be demonstrated as effective carriers. AIDA was also presented as a very successful alternative in the earlier prototypes of AMPSEC. The SDAs were coupled with a membrane-targeted protease that displayed effective cleaving of the cargo although this happened in a premature fashion in the cytosol of our host bacterium.

Experimenting with clustering approaches to discover structure-target characteristics of AMPs brought to light the challenges involved in the compatibility of this type of short peptide dataset with preestablished clustering algorithms designed for more typical proteomics datasets. The small lengths, small sample size and large sequence length variations reduced the effectiveness of the Linclust algorithm.

Despite this, these priceless tools would serve to increase the efficiency of the AMP research pipeline. With our work, we will also be furthering the knowledge on AMP structure-function as well as our understanding of surface display technologies which have applications in many fields of study beyond AMP design.

References:

1. World Health Organization. 10 global health issues to track in 2021. World Health Organization. Published December 24, 2020. Accessed January 26, 2022. https://www.who.int/news-room/spotlight/10-global-health-issues-to-track-in-2021

2. World Health Organization. Antimicrobial resistance. World Health Organization. Published November 17, 2021. Accessed November 4, 2019. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance

3. World Health Organization. New report calls for urgent action to avert antimicrobial resistance crisis. World Health Organization. Published April 29, 2019. Accessed January 26, 2022. https://www.who.int/news/item/29-04-2019-new-report-calls-for-urgent-action-to-avert-antimicrobial-resistance-crisis

4. Wang CH, Hsieh YH, Powers ZM, Kao CY. Defeating Antibiotic-Resistant Bacteria: Exploring Alternative Therapies for a Post-Antibiotic Era. *Int J Mol Sci*. 2020;21(3):1061. doi:10.3390/ijms21031061

5. Abdelkader K, Gerstmans H, Saafan A, Dishisha T, Briers Y. The Preclinical and Clinical Progress of Bacteriophages and Their Lytic Enzymes: The Parts are Easier than the Whole. *Viruses*. 2019;11(2):96. doi:10.3390/v11020096

6. Wang JW, Kuo CH, Kuo FC, et al. Fecal microbiota transplantation: Review and update. *J Formos Med Assoc*. 2019;118:S23-S31. doi:10.1016/j.jfma.2018.08.011

7. Sartor RB. Probiotic therapy of intestinal inflammation and infections. *Curr Opin Gastroenterol*. 2005;21(1):44-50.

8. Van Hoogmoed CG, Geertsema-doornbusch GI, Teughels W, Quirynen M, Busscher HJ, Van der Mei HC. Reduction of periodontal pathogens adhesion by antagonistic strains. *Oral Microbiol Immunol*. 2008;23(1):43-48. doi:10.1111/j.1399-302X.2007.00388.x

9. Pirtskhalava M, Amstrong AA, Grigolava M, et al. DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. *Nucleic Acids Res*. 2021;49(D1):D288-D297. doi:10.1093/nar/gkaa991

10. Wang G, Li X, Wang Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res*. 2016;44(Database issue):D1087-D1093. doi:10.1093/nar/gkv1278

11. Shi G, Kang X, Dong F, et al. DRAMP 3.0: an enhanced comprehensive data repository of antimicrobial peptides. *Nucleic Acids Res*. 2021;50(D1):D488-D496. doi:10.1093/nar/gkab651

12. Pirtskhalava M, Gabrielian A, Cruz P, et al. DBAASP v.2: an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides. *Nucleic Acids Res*. 2016;44(Database issue):D1104-D1112. doi:10.1093/nar/gkv1174

13. Wang G, Li X, Wang Z. APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res*. 2009;37(Database issue):D933-D937. doi:10.1093/nar/gkn823

14. Kang X, Dong F, Shi C, et al. DRAMP 2.0, an updated data repository of antimicrobial peptides. *Sci Data*. 2019;6(1):148. doi:10.1038/s41597-019-0154-y

15. Haney EF, Mansour SC, Hancock REW. Antimicrobial Peptides: An Introduction. In: Hansen PR, ed. *Antimicrobial Peptides: Methods and Protocols*. Methods in Molecular Biology. Springer; 2017:3-22. doi:10.1007/978-1-4939-6737-7_1

16. Martínez-Núñez MA, López VEL y. Nonribosomal peptides synthetases and their applications in industry. *Sustain Chem Process*. 2016;4(1):13. doi:10.1186/s40508-016-0057-6

17. Hur GH, Vickery CR, Burkart MD. Explorations of Catalytic Domains in Non-Ribosomal Peptide Synthetase Enzymology. *Nat Prod Rep*. 2012;29(10):1074-1098. doi:10.1039/c2np20025b

18. Mcphee JB, Hancock REW. Function and therapeutic potential of host defence peptides. *J Pept Sci*. 2005;11(11):677-687. doi:10.1002/psc.704

19. Yeaman MR, Yount NY. Unifying themes in host defence effector polypeptides. *Nat Rev Microbiol*. 2007;5(9):727-740. doi:10.1038/nrmicro1744

20. Brumfitt W, Salton MRJ, Hamilton-Miller JMT. Nisin, alone and combined with peptidoglycan-modulating antibiotics: activity against methicillin-resistant Staphylococcus aureus and vancomycin-resistant enterococci. *J Antimicrob Chemother*. 2002;50(5):731-734. doi:10.1093/jac/dkf190

21. Lewis K. Riddle of Biofilm Resistance. *Antimicrob Agents Chemother*. 2001;45(4):999-1007. doi:10.1128/AAC.45.4.999-1007.2001

22. Keren I, Kaldalu N, Spoering A, Wang Y, Lewis K. Persister cells and tolerance to antimicrobials. *FEMS Microbiol Lett*. 2004;230(1):13-18. doi:10.1016/S0378-1097(03)00856-5

23. Stewart PS, William Costerton J. Antibiotic resistance of bacteria in biofilms. *The Lancet*. 2001;358(9276):135-138. doi:10.1016/S0140-6736(01)05321-1

24. Chen X, Zhang M, Zhou C, Kallenbach NR, Ren D. Control of Bacterial Persister Cells by Trp/Arg-Containing Antimicrobial Peptides ▿ . *Appl Environ Microbiol*. 2011;77(14):4878-4885. doi:10.1128/AEM.02440-10

25. Schweizer F. Cationic amphiphilic peptides with cancer-selective toxicity. *Eur J Pharmacol*. 2009;625(1):190-194. doi:10.1016/j.ejphar.2009.08.043

26. Mahlapuu M, Håkansson J, Ringstad L, Björn C. Antimicrobial Peptides: An Emerging Category of Therapeutic Agents. *Front Cell Infect Microbiol*. 2016;6. doi:10.3389/fcimb.2016.00194

27. Loeffler JM, Nelson D, Fischetti VA. Rapid Killing of Streptococcus pneumoniae with a Bacteriophage Cell Wall Hydrolase. *Science*. 2001;294(5549):2170-2172. doi:10.1126/science.1066869

28. Phoenix DA, Dennison SR, Harris F. *Antimicrobial Peptides*. John Wiley & Sons; 2012.

29. Bahar AA, Ren D. Antimicrobial Peptides. *Pharmaceuticals*. 2013;6(12):1543-1575. doi:10.3390/ph6121543

30. Huang Y, Huang J, Chen Y. Alpha-helical cationic antimicrobial peptides: relationships of structure and function. *Protein Cell*. 2010;1(2):143-152. doi:10.1007/s13238-010-0004-3

31. Rathinakumar R, Wimley WC. Biomolecular Engineering by Combinatorial Design and High-throughput Screening: Small, Soluble Peptides that Permeabilize Membranes. *J Am Chem Soc*. 2008;130(30):9849-9858. doi:10.1021/ja8017863

32. Li J, Koh JJ, Liu S, Lakshminarayanan R, Verma CS, Beuerman RW. Membrane Active Antimicrobial Peptides: Translating Mechanistic Insights to Design. *Front Neurosci*. 2017;11. doi:10.3389/fnins.2017.00073

33. Tucker AT, Leonard SP, DuBois CD, et al. Discovery of Next-Generation Antimicrobials through Bacterial Self-Screening of Surface-Displayed Peptide Libraries. *Cell*. 2018;172(3):618-628.e13. doi:10.1016/j.cell.2017.12.009

34. Madani F, Lindberg S, Langel Ü, Futaki S, Gräslund A. Mechanisms of Cellular Uptake of Cell-Penetrating Peptides. *J Biophys*. 2011;2011. doi:10.1155/2011/414729

35. Subbalakshmi C, Sitaram N. Mechanism of antimicrobial action of indolicidin. *FEMS Microbiol Lett*. 1998;160(1):91-96. doi:10.1111/j.1574-6968.1998.tb12896.x

36. Castle M, Nazarian A, Yi SS, Tempst P. Lethal Effects of Apidaecin on Escherichia coliInvolve Sequential Molecular Interactions with Diverse Targets. *J Biol Chem*. 1999;274(46):32555-32564. doi:10.1074/jbc.274.46.32555

37. Haney EF, Straus SK, Hancock REW. Reassessing the Host Defense Peptide Landscape. *Front Chem*. 2019;7. doi:10.3389/fchem.2019.00043

38. Hancock REW, Lehrer R. Cationic peptides: a new source of antibiotics. *Trends Biotechnol*. 1998;16(2):82-88. doi:10.1016/S0167-7799(97)01156-6

39. Wang J, Dou X, Song J, et al. Antimicrobial peptides: Promising alternatives in the post feeding antibiotic era. *Med Res Rev*. 2019;39(3):831-859. doi:10.1002/med.21542

40. Rozek A, Friedrich CL, Hancock REW. Structure of the Bovine Antimicrobial Peptide Indolicidin Bound to Dodecylphosphocholine and Sodium Dodecyl Sulfate Micelles,. *Biochemistry*. 2000;39(51):15765-15774. doi:10.1021/bi000714m

41. Hancock REW, Chapple DS. Peptide Antibiotics. *Antimicrob Agents Chemother*. 1999;43(6):1317-1323. doi:10.1128/aac.43.6.1317

42. Kreil G. d-AMINO ACIDS IN ANIMAL PEPTIDES. *Annu Rev Biochem*. 1997;66(1):337-345. doi:10.1146/annurev.biochem.66.1.337

43. Yoshimi K, Hiroyuji M, K.Yosuke N, Kah HK, Anchalee Y, Hiroshi T. Isolation of achatin-i, a neuroactive tetrapeptide having a d-phenylalanine residue, from Achatina ganglia, and its effects on Achatina giant neurones. *Comp Biochem Physiol Part C Comp Pharmacol*. 1991;98(1):97-103. doi:10.1016/0742-8413(91)90186-W

44. Goumon Y, Strub JM, Moniatte M, et al. The C-terminal Bisphosphorylated proenkephalin-A-(209–237)-peptide from Adrenal Medullary Chromaffin Granules Possesses Antibacterial Activity. *Eur J Biochem*. 1996;235(3):516-525. doi:10.1111/j.1432-1033.1996.t01-1-00516.x

45. Rifflet A, Gavalda S, Téné N, et al. Identification and characterization of a novel antimicrobial peptide from the venom of the ant Tetramorium bicarinatum. *Peptides*. 2012;38(2):363-370. doi:10.1016/j.peptides.2012.08.018

46. Mangoni ME, Aumelas A, Charnet P, et al. Change in membrane permeability induced by protegrin 1: implication of disulphide bridges for pore formation. *FEBS Lett*. 1996;383(1-2):93-98. doi:10.1016/0014-5793(96)00236-0

47. Oman TJ, Boettcher JM, Wang H, Okalibe XN, van der Donk WA. Sublancin is not a lantibiotic but an S-linked glycopeptide. *Nat Chem Biol*. 2011;7(2):78-80. doi:10.1038/nchembio.509

48. Shinnar AE, Butler KL, Park HJ. Cathelicidin family of antimicrobial peptides: proteolytic processing and protease resistance. *Bioorganic Chem*. 2003;31(6):425-436. doi:10.1016/S0045-2068(03)00080-4

49. Pancu DF, Scurtu A, Macasoi IG, et al. Antibiotics: Conventional Therapy and Natural Compounds with Antibacterial Activity—A Pharmaco-Toxicological Screening. *Antibiotics*. 2021;10(4):401. doi:10.3390/antibiotics10040401

50. Wade JD, Lin F, Hossain MA, Dawson RM. Chemical synthesis and biological evaluation of an antimicrobial peptide gonococcal growth inhibitor. *Amino Acids*. 2012;43(6):2279-2283. doi:10.1007/s00726-012-1305-z

51. Piers KL, Brown MH, Hancock REW. Recombinant DNA procedures for producing small antimicrobial cationic peptides in bacteria. *Gene*. 1993;134(1):7-13. doi:10.1016/0378-1119(93)90168-3

52. Ramos R, Moreira S, Rodrigues A, Gama M, Domingues L. Recombinant expression and purification of the antimicrobial peptide magainin-2. *Biotechnol Prog*. 2013;29(1):17-22. doi:10.1002/btpr.1650

53. Jenssen H, Hamill P, Hancock REW. Peptide Antimicrobial Agents. *Clin Microbiol Rev*. 2006;19(3):491-511. doi:10.1128/CMR.00056-05

54. Kobayashi S, Takeshima K, Park CB, Kim SC, Matsuzaki K. Interactions of the Novel Antimicrobial Peptide Buforin 2 with Lipid Bilayers: Proline as a Translocation Promoting Factor. *Biochemistry*. 2000;39(29):8648-8654. doi:10.1021/bi0004549

55. Jiang Z, Vasil AI, Hale JD, Hancock REW, Vasil ML, Hodges RS. Effects of Net Charge and the Number of Positively Charged Residues on the Biological Activity of Amphipathic α-Helical Cationic Antimicrobial Peptides. *Biopolymers*. 2008;90(3):369-383. doi:10.1002/bip.20911

56. Chen Y, Guarnieri MT, Vasil AI, Vasil ML, Mant CT, Hodges RS. Role of Peptide Hydrophobicity in the Mechanism of Action of α-Helical Antimicrobial Peptides. *Antimicrob Agents Chemother*. 2007;51(4):1398-1406. doi:10.1128/AAC.00925-06

57. Berthold N, Czihal P, Fritsche S, et al. Novel Apidaecin 1b Analogs with Superior Serum Stabilities for Treatment of Infections by Gram-Negative Pathogens. *Antimicrob Agents Chemother*. 2013;57(1):402-409. doi:10.1128/AAC.01923-12

58. Kim JY, Park SC, Yoon MY, Hahm KS, Park Y. C-terminal amidation of PMAP-23: translocation to the inner membrane of Gram-negative bacteria. *Amino Acids*. 2011;40(1):183-195. doi:10.1007/s00726-010-0632-1

59. Strandberg E, Tiltak D, Ieronimo M, Kanithasen N, Wadhwani P, Ulrich AS. Influence of C-terminal amidation on the antimicrobial and hemolytic activities of cationic α-helical peptides. *Pure Appl Chem*. 2007;79(4):717-728. doi:10.1351/pac200779040717

60. Park IY, Cho JH, Kim KS, Kim YB, Kim MS, Kim SC. Helix Stability Confers Salt Resistance upon Helical Antimicrobial Peptides. *J Biol Chem*. 2004;279(14):13896-13901. doi:10.1074/jbc.M311418200

61. Uteng M, Hauge HH, Markwick PRL, et al. Three-Dimensional Structure in Lipid Micelles of the Pediocin-like Antimicrobial Peptide Sakacin P and a Sakacin P Variant That Is Structurally Stabilized by an Inserted C-Terminal Disulfide Bridge. *Biochemistry*. 2003;42(39):11417-11426. doi:10.1021/bi034572i

62. Rozek A, Powers JPS, Friedrich CL, Hancock REW. Structure-Based Design of an Indolicidin Peptide Analogue with Increased Protease Stability,. *Biochemistry*. 2003;42(48):14130-14138. doi:10.1021/bi035643g

63. Nguyen LT, Chau JK, Perry NA, de Boer L, Zaat SAJ, Vogel HJ. Serum Stabilities of Short Tryptophan- and Arginine-Rich Antimicrobial Peptide Analogs. *PLoS ONE*. 2010;5(9). doi:10.1371/journal.pone.0012684

64. Jing X, Jin K. A gold mine for drug discovery: Strategies to develop cyclic peptides into therapies. *Med Res Rev*. 2019;0(0). doi:10.1002/med.21639

65. Lee DG, Hahm KS, Shin SY. Structure and fungicidal activity of a synthetic antimicrobial peptide, P18, and its truncated peptides. *Biotechnol Lett*. 2004;26(4):337-341. doi:10.1023/B:BILE.0000015472.09542.6d

66. Gupta M, Chauhan VS. De novo design of α,β-didehydrophenylalanine containing peptides: From models to applications. *Biopolymers*. 2011;95(3):161-173. doi:10.1002/bip.21561

67. Yu HY, Tu CH, Yip BS, et al. Easy Strategy To Increase Salt Resistance of Antimicrobial Peptides ▿ . *Antimicrob Agents Chemother*. 2011;55(10):4918-4921. doi:10.1128/AAC.00202-11

68. Aoki W, Ueda M. Characterization of Antimicrobial Peptides toward the Development of Novel Antibiotics. *Pharmaceuticals*. 2013;6(8):1055-1081. doi:10.3390/ph6081055

69. Browne K, Chakraborty S, Chen R, et al. A New Era of Antibiotics: The Clinical Potential of Antimicrobial Peptides. *Int J Mol Sci*. 2020;21(19):7047. doi:10.3390/ijms21197047

70. Guralp SA, Murgha YE, Rouillard JM, Gulari E. From Design to Screening: A New Antimicrobial Peptide Discovery Pipeline. *PLOS ONE*. 2013;8(3):e59305. doi:10.1371/journal.pone.0059305

71. Zhang Y, Li Z, Li L, Rao B, Ma L, Wang Y. A Method for Rapid Screening, Expression, and Purification of Antimicrobial Peptides. *Microorganisms*. 2021;9(9):1858. doi:10.3390/microorganisms9091858

72. Kjeldsen A, Kay JE, Baxter S, et al. The fluorescent protein iLOV as a reporter for screening of high-yield production of antimicrobial peptides in Pichia pastoris. *Microb Biotechnol*. n/a(n/a). doi:10.1111/1751-7915.14034

73. Koch P, Schmitt S, Cardner M, Beerenwinkel N, Panke S, Held M. Discovery of antimicrobials by massively parallelized growth assays (Mex). *Sci Rep*. 2022;12:4097. doi:10.1038/s41598-022-07755-7

74. Nuti N, Rottmann P, Stucki A, Koch P, Panke S, Dittrich PS. A Multiplexed Cell-Free Assay to Screen for Antimicrobial Peptides in Double Emulsion Droplets. *Angew Chem Int Ed*. 2022;61(13):e202114632. doi:10.1002/anie.202114632

75. Schreiber C, Müller H, Birrenbach O, et al. A high-throughput expression screening platform to optimize the production of antimicrobial peptides. *Microb Cell Factories*. 2017;16:29. doi:10.1186/s12934-017-0637-5

76. Li L, Mu L, Wang X, Yu J, Hu R, Li Z. A novel expression vector for the secretion of abaecin in Bacillus subtilis. *Braz J Microbiol*. 2017;48(4):809-814. doi:10.1016/j.bjm.2017.01.009

77. Chen X, Li J, Sun H, et al. High-level heterologous production and Functional Secretion by recombinant Pichia pastoris of the shortest proline-rich antibacterial honeybee peptide Apidaecin. *Sci Rep*. 2017;7:14543. doi:10.1038/s41598-017-15149-3

78. GUO C, HUANG Y, ZHENG H, et al. Secretion and activity of antimicrobial peptide cecropin D expressed in Pichia pastoris. *Exp Ther Med*. 2012;4(6):1063-1068. doi:10.3892/etm.2012.719

79. Ke T, Liang S, Huang J, et al. A novel PCR-based method for high throughput prokaryotic expression of antimicrobial peptide genes. *BMC Biotechnol*. 2012;12:10. doi:10.1186/1472-6750-12-10

80. Nicolay T, Vanderleyden J, Spaepen S. Autotransporter-based cell surface display in Gram-negative bacteria. *Crit Rev Microbiol*. 2015;41(1):109-123. doi:10.3109/1040841X.2013.804032

81. Jose J. Autodisplay: efficient bacterial surface display of recombinant proteins. *Appl Microbiol Biotechnol*. 2006;69(6):607-614. doi:10.1007/s00253-005-0227-z

82. Kuroda K, Ueda M. Cell surface engineering of yeast for applications in white biotechnology. *Biotechnol Lett*. 2011;33(1):1-9. doi:10.1007/s10529-010-0403-9

83. Desvaux M, Dumas E, Chafsey I, Hébraud M. Protein cell surface display in Gram-positive bacteria: from single protein to macromolecular protein structure. *FEMS Microbiol Lett*. 2006;256(1):1-15. doi:10.1111/j.1574-6968.2006.00122.x

84. van Bloois E, Winter RT, Kolmar H, Fraaije MW. Decorating microbes: surface display of proteins on Escherichia coli. *Trends Biotechnol*. 2011;29(2):79-86. doi:10.1016/j.tibtech.2010.11.003

85. Samuelson P, Gunneriusson E, Nygren PÅ, Ståhl S. Display of proteins on bacteria. *J Biotechnol*. 2002;96(2):129-154. doi:10.1016/S0168-1656(02)00043-3

86. Leo JC, Grin I, Linke D. Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philos Trans R Soc B Biol Sci*. 2012;367(1592):1088-1101. doi:10.1098/rstb.2011.0208

87. Bernstein HD. Looks can be deceiving: recent insights into the mechanism of protein secretion by the autotransporter pathway. *Mol Microbiol*. 2015;97(2):205-215. doi:10.1111/mmi.13031

88. Nicolay T, Lemoine L, Lievens E, Balzarini S, Vanderleyden J, Spaepen S. Probing the applicability of autotransporter based surface display with the EstA autotransporter of Pseudomonas stutzeri A15. *Microb Cell Factories*. 2012;11:158. doi:10.1186/1475-2859-11-158

89. Leyton DL, Sevastsyanovich YR, Browning DF, et al. Size and Conformation Limits to Secretion of Disulfide-bonded Loops in Autotransporter Proteins *. *J Biol Chem*. 2011;286(49):42283-42291. doi:10.1074/jbc.M111.306118

90. Quehl P, Schüürmann J, Hollender J, Jose J. Improving the activity of surface displayed cytochrome P450 enzymes by optimizing the outer membrane linker. *Biochim Biophys Acta BBA - Biomembr*. 2017;1859(1):104-116. doi:10.1016/j.bbamem.2016.10.022

91. Zhu Y, Liu Y, Ai M, Jia X. Surface display of carbonic anhydrase on Escherichia coli for CO2 capture and mineralization. *Synth Syst Biotechnol*. 2021;7(1):460-473. doi:10.1016/j.synbio.2021.11.008

92. Ullah J, Chen H, Vastermark A, et al. Impact of orientation and flexibility of peptide linkers on T. maritima lipase Tm1350 displayed on Bacillus subtilis spores surface using CotB as fusion partner. *World J Microbiol Biotechnol*. 2017;33(9):166. doi:10.1007/s11274-017-2327-1

93. Daugherty PS, Olsen MJ, Iverson BL, Georgiou G. Development of an optimized expression system for the screening of antibody libraries displayed on the Escherichia coli surface. *Protein Eng Des Sel*. 1999;12(7):613-621. doi:10.1093/protein/12.7.613

94. Narayanan N, Chou CP. Physiological Improvement to Enhance Escherichia coli Cell-Surface Display via Reducing Extracytoplasmic Stress. *Biotechnol Prog*. 2008;24(2):293-301. doi:10.1021/bp0702121

95. Schlegel S, Löfblom J, Lee C, et al. Optimizing Membrane Protein Overexpression in the Escherichia coli strain Lemo21(DE3). *J Mol Biol*. 2012;423(4):648-659. doi:10.1016/j.jmb.2012.07.019

96. Sichwart S, Tozakidis IEP, Teese M, Jose J. Maximized Autotransporter-Mediated Expression (MATE) for Surface Display and Secretion of Recombinant Proteins in Escherichia coli. *Food Technol Biotechnol*. 2015;53(3):251-260. doi:10.17113/ftb.53.03.15.3802

97. Binder U, Matschiner G, Theobald I, Skerra A. High-throughput Sorting of an Anticalin Library via EspP-mediated Functional Display on the Escherichia coli Cell Surface. *J Mol Biol*. 2010;400(4):783-802. doi:10.1016/j.jmb.2010.05.049

98. Inokuma K, Bamba T, Ishii J, Ito Y, Hasunuma T, Kondo A. Enhanced cell-surface display and secretory production of cellulolytic enzymes with Saccharomyces cerevisiae Sed1 signal peptide. *Biotechnol Bioeng*. 2016;113(11):2358-2366. doi:10.1002/bit.26008

99. Tozakidis IEP, Lüken LM, Üffing A, Meyers A, Jose J. Improving the autotransporter-based surface display of enzymes in Pseudomonas putida KT2440. *Microb Biotechnol*. 2019;13(1):176-184. doi:10.1111/1751-7915.13419

100. Quehl P, Hollender J, Schüürmann J, Brossette T, Maas R, Jose J. Co-expression of active human cytochrome P450 1A2 and cytochrome P450 reductase on the cell surface of Escherichia coli. *Microb Cell Factories*. 2016;15:26. doi:10.1186/s12934-016-0427-5

101. Kranen E, Detzel C, Weber T, Jose J. Autodisplay for the co-expression of lipase and foldase on the surface of E. coli: washing with designer bugs. *Microb Cell Factories*. 2014;13:19. doi:10.1186/1475-2859-13-19

102. Balderas Hernández VE, Salas-Montantes CJ, Barba-De la Rosa AP, De Leon-Rodriguez A. Autodisplay of an endo-1,4-β-xylanase from Clostridium cellulovorans in Escherichia coli for xylans degradation. *Enzyme Microb Technol*. 2021;149:109834. doi:10.1016/j.enzmictec.2021.109834

103. Fleetwood F, Andersson KG, Ståhl S, Löfblom J. An engineered autotransporter-based surface expression vector enables efficient display of Affibody molecules on OmpT-negative E. coli as well as protease-mediated secretion in OmpT-positive strains. *Microb Cell Factories*. 2014;13. doi:10.1186/s12934-014-0179-z

104. Andersson KG, Persson J, Ståhl S, Löfblom J. Autotransporter-Mediated Display of a Naïve Affibody Library on the Outer Membrane of Escherichia coli. *Biotechnol J*. 2019;14(4):1800359. doi:10.1002/biot.201800359

105. Mukherjee S, De Buck J. Autotransporter-based surface expression and complementation of split TreA fragments utilized for the detection of antibodies against bovine leukemia virus. *J Immunol Methods*. 2021;495:113084. doi:10.1016/j.jim.2021.113084

106. Lin CH, Chen JJ, Cheng and CM. Developing a Virus-Binding Bacterium Expressing Mx Protein on the Bacterial Surface to Prevent Grouper Nervous Necrosis Virus Infection. 2021;31(8):1088-1097. doi:10.4014/jmb.2103.03036

107. Skillman KM, Barnard TJ, Peterson JH, Ghirlando R, Bernstein HD. Efficient secretion of a folded protein domain by a monomeric bacterial autotransporter. *Mol Microbiol*. 2005;58(4):945-958. doi:10.1111/j.1365-2958.2005.04885.x

108. Tozakidis IEP, Brossette T, Lenz F, Maas RM, Jose J. Proof of concept for the simplified breakdown of cellulose by combining Pseudomonasputida strains with surface displayed thermophilic endocellulase, exocellulase and β-glucosidase. *Microb Cell Factories*. 2016;15:103. doi:10.1186/s12934-016-0505-8

109. Earhart CF. [30] Use of an Lpp-OmpA fusion vehicle for bacterial surface display. In: *Methods in Enzymology*. Vol 326. Elsevier; 2000:506-516. doi:10.1016/S0076-6879(00)26072-2

110. Gallus S, Peschke T, Paulsen M, Burgahn T, Niemeyer CM, Rabe KS. Surface Display of Complex Enzymes by in Situ SpyCatcher-SpyTag Interaction. *Chembiochem*. 2020;21(15):2126-2131. doi:10.1002/cbic.202000102

111. Jo JH, Han CW, Kim SH, Kwon HJ, Lee HH. Surface display expression of Bacillus licheniformis lipase in Escherichia coli using Lpp'OmpA chimera. *J Microbiol*. 2014;52(10):856-862. doi:10.1007/s12275-014-4217-7

112. Qu W, Xue Y, Ding Q. Display of Fungi Xylanase on Escherichia coli Cell Surface and Use of the Enzyme in Xylan Biodegradation. *Curr Microbiol*. 2015;70(6):779-785. doi:10.1007/s00284-015-0781-2

113. Hui C, Guo Y, Zhang W, et al. Surface display of PbrR on Escherichia coli and evaluation of the bioavailability of lead associated with engineered cells in mice. *Sci Rep*. 2018;8:5685. doi:10.1038/s41598-018-24134-3

114. Vahed M, Ramezani F, Tafakori V, Mirbagheri VS, Najafi A, Ahmadian G. Molecular dynamics simulation and experimental study of the surface-display of SPA protein via Lpp-OmpA system for screening of IgG. *AMB Express*. 2020;10:161. doi:10.1186/s13568-020-01097-1

115. Karami A, Khodi AML and S. Comparison of the Organophosphorus Hydrolase Surface Display Using InaVN and Lpp-OmpA Systems in Escherichia coli. 2014;24(3):379-385. doi:10.4014/jmb.1309.09066

116. Benhar I, Azriel R, Nahary L, et al. Highly efficient selection of phage antibodies mediated by display of antigen as Lpp-OmpA′ fusions on live bacteria11Edited by J. Wells. *J Mol Biol*. 2000;301(4):893-904. doi:10.1006/jmbi.2000.4021

117. Hui C ye, Guo Y, Liu L, et al. Development of a novel bacterial surface display system using truncated OmpT as an anchoring motif. *Biotechnol Lett*. 2019;41(6):763-777. doi:10.1007/s10529-019-02676-4

118. Yang C, Zhao Q, Liu Z, et al. Cell Surface Display of Functional Macromolecule Fusions on Escherichia coli for Development of an Autofluorescent Whole-Cell Biocatalyst. *Environ Sci Technol*. 2008;42(16):6105-6110. doi:10.1021/es800441t

119. Dong H, Sarkes DA, Rice JJ, Hurley MM, Fu AJ, Stratis-Cullum DN. Living Bacteria–Nanoparticle Hybrids Mediated through Surface-Displayed Peptides. *Langmuir*. 2018;34(20):5837-5848. doi:10.1021/acs.langmuir.8b00114

120. Dong H, Sarkes DA, Stratis-Cullum DN, Hurley MM. Direct conjugation of fluorescent quantum dots with E. coli via surface-displayed histidine-containing peptides. *Colloids Surf B Biointerfaces*. 2021;203:111730. doi:10.1016/j.colsurfb.2021.111730

121. Sankaran S, Kiren MC, Jonkheijm P. Incorporating Bacteria as a Living Component in Supramolecular Self-Assembled Monolayers through Dynamic Nanoscale Interactions. *ACS Nano*. 2015;9(4):3579-3586. doi:10.1021/acsnano.5b00694

122. Kogot JM, Zhang Y, Moore SJ, et al. Screening of Peptide Libraries against Protective Antigen of Bacillus anthracis in a Disposable Microfluidic Cartridge. *PLoS ONE*. 2011;6(11):e26925. doi:10.1371/journal.pone.0026925

123. Yang C, Cai N, Dong M, et al. Surface display of MPH on Pseudomonas putida JS444 using ice nucleation protein and its application in detoxification of organophosphates. *Biotechnol Bioeng*. 2008;99(1):30-37. doi:10.1002/bit.21535

124. Li Q, Yu Z, Shao X, He J, Li L. Improved phosphate biosorption by bacterial surface display of phosphate-binding protein utilizing ice nucleation protein. *FEMS Microbiol Lett*. 2009;299(1):44-52. doi:10.1111/j.1574-6968.2009.01724.x

125. Liu D, Geng H, Zhang Z, et al. An Effective Platform for Exploring Rotavirus Receptors by Bacterial Surface Display System. *Virol Sin*. 2019;35(1):103-109. doi:10.1007/s12250-019-00174-5

126. Zhou XF, Zhang CL, Gao XP, et al. A simple and rapid protein purification method based on cell-surface display of SUMO-fused recombinant protein and Ulp1 protease. *AMB Express*. 2020;10. doi:10.1186/s13568-020-00999-4

127. Ko HJ, Park E, Song J, et al. Functional Cell Surface Display and Controlled Secretion of Diverse Agarolytic Enzymes by Escherichia coli with a Novel Ligation-Independent Cloning Vector Based on the Autotransporter YfaL. *Appl Environ Microbiol*. 2012;78(9):3051-3058. doi:10.1128/AEM.07004-11

128. Xu Y, Liu Q, Zhou L, Yang Z, Zhang Y. Surface Display of GFP by Pseudomonas Syringae Truncated Ice Nucleation Protein in Attenuated Vibrio Anguillarum Strain. *Mar Biotechnol*. 2008;10(6):701. doi:10.1007/s10126-008-9108-7

129. Bernstein HD. Type V Secretion: the Autotransporter and Two-Partner Secretion Pathways. *EcoSal Plus*. 2010;4(1). doi:10.1128/ecosalplus.4.3.6

130. Brockmeyer J, Spelten S, Kuczius T, Bielaszewska M, Karch H. Structure and Function Relationship of the Autotransport and Proteolytic Activity of EspP from Shiga Toxin-Producing Escherichia coli. *PLoS ONE*. 2009;4(7). doi:10.1371/journal.pone.0006100

131. Dutta PR, Sui BQ, Nataro JP. Structure-Function Analysis of the Enteroaggregative Escherichia coli Plasmid-encoded Toxin Autotransporter Using Scanning Linker Mutagenesis. *J Biol Chem*. 2003;278(41):39912-39920. doi:10.1074/jbc.M303595200

132. Maurer J, Jose J, Meyer TF. Characterization of the Essential Transport Function of the AIDA-I Autotransporter and Evidence Supporting Structural Predictions. *J Bacteriol*. 1999;181(22):7014-7020.

133. Hritonenko V, Stathopoulos C. Omptin proteins: an expanding family of outer membrane proteases in Gram-negative Enterobacteriaceae (Review). *Mol Membr Biol*. 2007;24(5-6):395-406. doi:10.1080/09687680701443822

134. Steinhauer J, Agha R, Pham T, Varga AW, Goldberg MB. The unipolar Shigella surface protein IcsA is targeted directly to the bacterial old pole: IcsP cleavage of IcsA occurs over the entire bacterial surface. *Mol Microbiol*. 1999;32(2):367-377. doi:10.1046/j.1365-2958.1999.01356.x

135. Samykannu G, Vijayababu P, Antonyraj CB, Narayanan S. Structural investigation of APRs to improve the solubility of outer membrane protease (PgtE) from Salmonella enterica serotype typhi- A multi-constraint approach. *Biochem Biophys Rep*. 2019;21:100693. doi:10.1016/j.bbrep.2019.100693

136. Franco AA, Kothary MH, Gopinath G, et al. Cpa, the Outer Membrane Protease of Cronobacter sakazakii, Activates Plasminogen and Mediates Resistance to Serum Bactericidal Activity. *Infect Immun*. 2011;79(4):1578-1587. doi:10.1128/IAI.01165-10

137. Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnol J*. 2012;7(5):620-634. doi:10.1002/biot.201100155

138. Liew OW, Ching Chong JP, Yandle TG, Brennan SO. Preparation of recombinant thioredoxin fused N-terminal proCNP: Analysis of enterokinase cleavage products reveals new enterokinase cleavage sites. *Protein Expr Purif*. 2005;41(2):332-340. doi:10.1016/j.pep.2005.03.006

139. Choi SI, Song HW, Moon JW, Seong BL. Recombinant enterokinase light chain with affinity tag: Expression from Saccharomyces cerevisiae and its utilities in fusion protein technology. *Biotechnol Bioeng*. 2001;75(6):718-724. doi:10.1002/bit.10082

140. Jong WS, Soprova Z, de Punder K, et al. A structurally informed autotransporter platform for efficient heterologous protein secretion and display. *Microb Cell Factories*. 2012;11:85. doi:10.1186/1475-2859-11-85

141. Sevastsyanovich YR, Leyton DL, Wells TJ, et al. A generalised module for the selective extracellular accumulation of recombinant proteins. *Microb Cell Factories*. 2012;11:69. doi:10.1186/1475-2859-11-69

142. Swan AL, Mobasheri A, Allaway D, Liddell S, Bacardit J. Application of Machine Learning to Proteomics Data: Classification and Biomarker Identification in Postgenomics Biology. *OMICS J Integr Biol*. 2013;17(12):595-610. doi:10.1089/omi.2013.0017

143. Kotsiantis SB. Supervised Machine Learning: A Review of Classification Techniques. In: *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies*. IOS Press; 2007:3-24.

144. Sutton RS, Barto AG. Reinforcement Learning: An Introduction. :352.

145. Dridi S. *Unsupervised Learning - A Systematic Literature Review*.; 2021. doi:10.13140/RG.2.2.16963.12323

146. Xu D, Tian Y. A Comprehensive Survey of Clustering Algorithms. *Ann Data Sci*. 2015;2(2):165-193. doi:10.1007/s40745-015-0040-1

147. Carlsson G, Mémoli F. Characterization, stability and convergence of hierarchical clustering methods. *J Mach Learn Res JMLR*. 2010;11:1425-1470.

148. Bezdek JC, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm. *Comput Geosci*. 1984;10(2):191-203. doi:10.1016/0098-3004(84)90020-7

149. Bock HH. Probabilistic models in cluster analysis. *Comput Stat Data Anal*. 1996;23(1):5-28. doi:10.1016/0167-9473(96)88919-5

150. Ben-Israel A, Iyigun C. Probabilistic D-Clustering. *J Classif*. 2008;25(1):5-26. doi:10.1007/s00357-008-9002-z

151. Wang Y, Wu H, Cai Y. A benchmark study of sequence alignment methods for protein clustering. *BMC Bioinformatics*. 2018;19(19):529. doi:10.1186/s12859-018-2524-4

152. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet*. 2000;16(6):276-277. doi:10.1016/S0168-9525(00)02024-2

153. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008;36(suppl_2):W5-W9. doi:10.1093/nar/gkn201

154. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13):1658-1659. doi:10.1093/bioinformatics/btl158

155. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460-2461. doi:10.1093/bioinformatics/btq461

156. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*. 1970;48(3):443-453. doi:10.1016/0022-2836(70)90057-4

157. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol*. 1981;147(1):195-197. doi:10.1016/0022-2836(81)90087-5

158. Thompson JD, Linard B, Lecompte O, Poch O. A Comprehensive Benchmark Study of Multiple Sequence Alignment Methods: Current Challenges and Future Perspectives. *PLoS ONE*. 2011;6(3):e18093. doi:10.1371/journal.pone.0018093

159. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792-1797. doi:10.1093/nar/gkh340

160. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004;5(1):113. doi:10.1186/1471-2105-5-113

161. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22(22):4673-4680. doi:10.1093/nar/22.22.4673

162. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7(1):539. doi:10.1038/msb.2011.75

163. Notredame C, Higgins DG, Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment11Edited by J. Thornton. *J Mol Biol*. 2000;302(1):205-217. doi:10.1006/jmbi.2000.4042

164. Lassmann T. Kalign 3: multiple sequence alignment of large datasets. *Bioinformatics*. 2020;36(6):1928-1929. doi:10.1093/bioinformatics/btz795

165. Lassmann T, Sonnhammer EL. Kalign – an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics*. 2005;6(1):298. doi:10.1186/1471-2105-6-298

166. Boeva V. Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells. *Front Genet*. 2016;7:24. doi:10.3389/fgene.2016.00024

167. Mohamed SAEH, Elloumi M, D.Thompson J. *Motif Discovery in Protein Sequences*. IntechOpen; 2016. doi:10.5772/65441

168. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. *Nucleic Acids Res*. 2015;43(W1):W39-W49. doi:10.1093/nar/gkv416

169. Rigoutsos I, Floratos A. Combinatorial pattern discovery in biological sequences: The TEIRESIAS algorithm. *Bioinformatics*. 1998;14(1):55-67. doi:10.1093/bioinformatics/14.1.55

170. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*. 1994;2:28-36.

171. Grant CE, Bailey TL. XSTREME: Comprehensive motif analysis of biological sequence datasets. Published online September 3, 2021:2021.09.02.458722. doi:10.1101/2021.09.02.458722

172. Frith MC, Saunders NFW, Kobe B, Bailey TL. Discovering Sequence Motifs with Arbitrary Insertions and Deletions. *PLOS Comput Biol*. 2008;4(5):e1000071. doi:10.1371/journal.pcbi.1000071

173. Bailey TL. STREME: accurate and versatile sequence motif discovery. *Bioinformatics*. 2021;37(18):2834-2840. doi:10.1093/bioinformatics/btab203

174. Bailey TL, Grant CE. SEA: Simple Enrichment Analysis of motifs. Published online August 24, 2021:2021.08.23.457422. doi:10.1101/2021.08.23.457422

175. Neuwald AF, Liu JS, Lawrence CE. Gibbs motif sampling: Detection of bacterial outer membrane protein repeats. *Protein Sci*. 1995;4(8):1618-1632. doi:10.1002/pro.5560040820

176. Al-Enezi JR, Abbod MF, Alsharhan S. Artificial Immune Systems - Models, algorithms and applications. Published online 2010. Accessed March 24, 2022. http://bura.brunel.ac.uk/handle/2438/4643

177. Mulder M, Nezlek GS. Creating protein sequence patterns using efficient regular expressions in bioinformatics research. In: *28th International Conference on Information Technology Interfaces, 2006.* ; 2006:207-212. doi:10.1109/ITI.2006.1708479

178. Xu J, Li F, Leier A, et al. Comprehensive assessment of machine learning-based methods for predicting antimicrobial peptides. *Brief Bioinform*. 2021;22(5):bbab083. doi:10.1093/bib/bbab083

179. Xiao X, Wang P, Lin WZ, Jia JH, Chou KC. iAMP-2L: A two-level multi-label classifier for identifying antimicrobial peptides and their functional types. *Anal Biochem*. 2013;436(2):168-177. doi:10.1016/j.ab.2013.01.019

180. Shao L, Gao H, Liu Z, Feng J, Tang L, Lin H. Identification of Antioxidant Proteins With Deep Learning From Sequence Information. *Front Pharmacol*. 2018;9. Accessed April 28, 2022. https://www.frontiersin.org/article/10.3389/fphar.2018.01036

181. Lv H, Dao FY, Guan ZX, Yang H, Li YW, Lin H. Deep-Kcr: accurate detection of lysine crotonylation sites using deep learning method. *Brief Bioinform*. 2021;22(4):bbaa255. doi:10.1093/bib/bbaa255

182. Lee EY, Lee MW, Fulan BM, Ferguson AL, Wong GCL. What can machine learning do for antimicrobial peptides, and what can antimicrobial peptides do for machine learning? *Interface Focus*. 2017;7(6):20160153. doi:10.1098/rsfs.2016.0153

183. Wang P, Hu L, Liu G, et al. Prediction of Antimicrobial Peptides Based on Sequence Alignment and Feature Selection Methods. *PLoS ONE*. 2011;6(4):e18476. doi:10.1371/journal.pone.0018476

184. Torrent M, Andreu D, Nogués VM, Boix E. Connecting Peptide Physicochemical and Antimicrobial Properties by a Rational Prediction Model. *PLoS ONE*. 2011;6(2):e16968. doi:10.1371/journal.pone.0016968

185. Schneider P, Müller AT, Gabernet G, et al. Hybrid Network Model for "Deep Learning" of Chemical Data: Application to Antimicrobial Peptides. *Mol Inform*. 2017;36(1-2):1600011. doi:10.1002/minf.201600011

186. Steinegger M, Söding J. Clustering huge protein sequence sets in linear time. *Nat Commun*. 2018;9(1):2542. doi:10.1038/s41467-018-04964-5

187. Pinaud L, Ferrari ML, Friedman R, et al. Identification of novel substrates of Shigella T3SA through analysis of its virulence plasmid-encoded secretome. *PLOS ONE*. 2017;12(10):e0186920. doi:10.1371/journal.pone.0186920

188. Malherbe G, Humphreys DP, Davé E. A robust fractionation method for protein subcellular localization studies in Escherichia coli. *BioTechniques*. 2019;66(4):171-178. doi:10.2144/btn-2018-0135

189. Sandrini S, Haigh R, Freestone P. Fractionation by Ultracentrifugation of Gram Negative Cytoplasmic and Membrane Proteins. *BIO-Protoc*. 2014;4(21). doi:10.21769/BioProtoc.1287

190. Balouiri M, Bouhdid S, Harki EH, Sadiki M, Ouedrhiri W, Ibnsouda SK. ANTIFUNGAL ACTIVITY OF BACILLUS SPP. ISOLATED FROM CALOTROPIS PROCERA AIT. RHIZOSPHERE AGAINST CANDIDA ALBICANS. 2015;8(2):6.

191. Charbonneau MÈ, Berthiaume F, Mourez M. Proteolytic Processing Is Not Essential for Multiple Functions of the Escherichia coli Autotransporter Adhesin Involved in Diffuse Adherence (AIDA-I). *J Bacteriol*. 2006;188(24):8504-8512. doi:10.1128/JB.00864-06

192. Maurer J, Jose J, Meyer TF. Autodisplay: one-component system for efficient surface display and release of soluble recombinant proteins from Escherichia coli. *J Bacteriol*. 1997;179(3):794-804. doi:10.1128/jb.179.3.794-804.1997

193. Benz I, Schmidt MA. Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic Escherichia coli. *Infect Immun*. 1989;57(5):1506-1511.

194. Casali N, Konieczny M, Schmidt MA, Riley LW. Invasion Activity of a Mycobacterium tuberculosis Peptide Presented by the Escherichia coli AIDA Autotransporter. *Infect Immun*. 2002;70(12):6846-6852. doi:10.1128/IAI.70.12.6846-6852.2002

195. Jose J, Meyer TF. The Autodisplay Story, from Discovery to Biotechnical and Biomedical Applications. *Microbiol Mol Biol Rev MMBR*. 2007;71(4):600-619. doi:10.1128/MMBR.00011-07

196. Lattemann CT, Maurer J, Gerland E, Meyer TF. Autodisplay: Functional Display of Active β-Lactamase on the Surface of Escherichia coli by the AIDA-I Autotransporter. *J Bacteriol*. 2000;182(13):3726-3733. doi:10.1128/jb.182.13.3726-3733.2000

197. Konieczny MPJ, Suhr M, Noll A, Autenrieth IB, Schmidt MA. Cell surface presentation of recombinant (poly-) peptides including functional T-cell epitopes by the AIDA autotransporter system. *FEMS Immunol Med Microbiol*. 2000;27(4):321-332. doi:10.1111/j.1574-695X.2000.tb01446.x

198. Gawarzewski I, DiMaio F, Winterer E, et al. Crystal structure of the transport unit of the autotransporter adhesin involved in diffuse adherence from Escherichia coli. *J Struct Biol*. 2014;187(1):20-29. doi:10.1016/j.jsb.2014.05.003

199. Suhr M, Benz I, Schmidt MA. Processing of the AIDA-I precursor: removal of AIDAC and evidence for the outer membrane anchoring as a β-barrel structure. *Mol Microbiol*. 1996;22(1):31-42. doi:10.1111/j.1365-2958.1996.tb02653.x

200. Okuno K, Yabuta M, Ohsuye K, Ooi T, Kinoshita S. An analysis of target preferences of Escherichia coli outer-membrane endoprotease OmpT for use in therapeutic

peptide production: efficient cleavage of substrates with basic amino acids at the P4 and P6 positions. *Biotechnol Appl Biochem*. 2002;36(Pt 2):77-84. doi:10.1042/

201. Hwang BY, Varadarajan N, Li H, Rodriguez S, Iverson BL, Georgiou G. Substrate Specificity of the Escherichia coli Outer Membrane Protease OmpP. *J Bacteriol*. 2007;189(2):522-530. doi:10.1128/JB.01493-06

202. Dekker N, Cox RC, Kramer RA, Egmond MR. Substrate Specificity of the Integral Membrane Protease OmpT Determined by Spatially Addressed Peptide Libraries. *Biochemistry*. 2001;40(6):1694-1701. doi:10.1021/bi0014195

203. McCarter JD, Stephens D, Shoemaker K, Rosenberg S, Kirsch JF, Georgiou G. Substrate Specificity of the Escherichia coli Outer Membrane Protease OmpT. *J Bacteriol*. 2004;186(17):5919-5925. doi:10.1128/JB.186.17.5919-5925.2004

204. Klauser T, Pohlner J, Meyer TF. Extracellular transport of cholera toxin B subunit using Neisseria IgA protease beta-domain: conformation-dependent outer membrane translocation. *EMBO J*. 1990;9(6):1991-1999.

205. Klauser T, Pohlner J, Meyer TF. Selective extracellular release of cholera toxin B subunit by Escherichia coli: dissection of Neisseria Iga beta-mediated outer membrane transport. *EMBO J*. 1992;11(6):2327-2335.

206. Raran-Kurussi S, Cherry S, Zhang D, Waugh DS. Removal of Affinity Tags with TEV Protease. In: Burgess-Brown NA, ed. *Heterologous Gene Expression in E.Coli: Methods and Protocols*. Methods in Molecular Biology. Springer; 2017:221-230. doi:10.1007/978-1-4939-6887-9_14

207. Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Weeks SD, Butt TR. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics*. 2004;5(1):75-86. doi:10.1023/B:JSFG.0000029237.70316.52

208. Hew CL, Yang DSC. Protein interaction with ice. *Eur J Biochem*. 1992;203(1-2):33-42. doi:10.1111/j.1432-1033.1992.tb19824.x

209. Kukkonen M, Korhonen TK. The omptin family of enterobacterial surface proteases/adhesins: from housekeeping in Escherichia coli to systemic spread of Yersinia pestis. *Int J Med Microbiol*. 2004;294(1):7-14. doi:10.1016/j.ijmm.2004.01.003

210. Uriu K. Genetic oscillators in development. *Dev Growth Differ*. 2016;58(1):16-30. doi:10.1111/dgd.12262

211. Johnson CH. Bacterial Circadian Programs. *Cold Spring Harb Symp Quant Biol*. 2007;72:395-404. doi:10.1101/sqb.2007.72.027

212. Purcell O, Savery NJ, Grierson CS, di Bernardo M. A comparative analysis of synthetic genetic oscillators. *J R Soc Interface*. 2010;7(52):1503-1524. doi:10.1098/rsif.2010.0183

213. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*. 2000;403(6767):335-338. doi:10.1038/35002125

214. Potvin-Trottier L, Lord ND, Vinnicombe G, Paulsson J. Synchronous long-term oscillations in a synthetic gene circuit. *Nature*. 2016;538(7626):514-517. doi:10.1038/nature19841

215. Fernández-Niño M, Giraldo D, Gomez-Porras JL, Dreyer I, González Barrios AF, Arevalo-Ferro C. A synthetic multi-cellular network of coupled self-sustained oscillators. *PLoS ONE*. 2017;12(6):e0180155. doi:10.1371/journal.pone.0180155

216. Fung E, Wong WW, Suen JK, Bulter T, Lee S gu, Liao JC. A synthetic gene–metabolic oscillator. *Nature*. 2005;435(7038):118-122. doi:10.1038/nature03508

217. Nicchi S, Giuliani M, Giusti F, et al. Decorating the surface of Escherichia coli with bacterial lipoproteins: a comparative analysis of different display systems. *Microb Cell Factories*. 2021;20:33. doi:10.1186/s12934-021-01528-z

218. Gallus S, Mittmann E, Rabe KS. A Modular System for the Rapid Comparison of Different Membrane Anchors for Surface Display on Escherichia coli. *ChemBioChem*. 2022;23(2):e202100472. doi:10.1002/cbic.202100472

219. Feng X, Jin M, Huang W, Liu W, Xian M. Whole-cell catalysis by surface display of fluorinase on Escherichia coli using N-terminal domain of ice nucleation protein. *Microb Cell Factories*. 2021;20:206. doi:10.1186/s12934-021-01697-x

220. Hao K, Chen XH, Qi XZ, Zhu B, Wang GX, Ling F. Display of GCRV vp7 protein on the surface of Escherichia coli and its immunoprotective effects in grass carp (Ctenopharyngodon idella). *Fish Shellfish Immunol*. 2018;72:199-209. doi:10.1016/j.fsi.2017.10.060

221. Dautin N, Barnard TJ, Anderson DE, Bernstein HD. Cleavage of a bacterial autotransporter by an evolutionarily convergent autocatalytic mechanism. *EMBO J*. 2007;26(7):1942-1952. doi:10.1038/sj.emboj.7601638

222. Kleiner-Grote GRM, Risse JM, Friehs K. Secretion of recombinant proteins from E. coli. *Eng Life Sci*. 2018;18(8):532-550. doi:10.1002/elsc.201700200

223. Liu XQ. PROTEIN-SPLICING INTEIN: Genetic Mobility, Origin, and Evolution. *Annu Rev Genet*. 2000;34(1):61-76. doi:10.1146/annurev.genet.34.1.61

224. Nanda A, Nasker SS, Mehra A, Panda S, Nayak S. Inteins in Science: Evolution to Application. *Microorganisms*. 2020;8(12):2004. doi:10.3390/microorganisms8122004

225. Ando T, Tsukiji S, Tanaka T, Nagamune T. Construction of a small-molecule-integrated semisynthetic split intein for in vivoprotein ligation. *Chem Commun*. 2007;(47):4995-4997. doi:10.1039/B712843F

226. Thompson RE, Muir TW. Chemoenzymatic Semisynthesis of Proteins. *Chem Rev*. 2020;120(6):3051-3126. doi:10.1021/acs.chemrev.9b00450

227. Costa TRD, Felisberto-Rodrigues C, Meir A, et al. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol*. 2015;13(6):343-359. doi:10.1038/nrmicro3456

228. Gorasia DG, Veith PD, Reynolds EC. The Type IX Secretion System: Advances in Structure, Function and Organisation. *Microorganisms*. 2020;8(8):1173. doi:10.3390/microorganisms8081173

229. Naskar S, Hohl M, Tassinari M, Low HH. The structure and mechanism of the bacterial type II secretion system. *Mol Microbiol*. 2021;115(3):412-424. doi:10.1111/mmi.14664

230. Bhoite S, Van Gerven N, Chapman MR, Remaut H. Curli Biogenesis: Bacterial amyloid assembly by the Type VIII secretion pathway. *EcoSal Plus*. 2019;8(2):10.1128/ecosalplus.ESP-0037-2018. doi:10.1128/ecosalplus.ESP-0037-2018

231. Van Gerven N, Goyal P, Vandenbussche G, et al. Secretion and functional display of fusion proteins through the curli biogenesis pathway. *Mol Microbiol*. 2014;91(5):1022-1035. doi:10.1111/mmi.12515

232. Sato K, Naito M, Yukitake H, et al. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc Natl Acad Sci U S A*. 2010;107(1):276-281. doi:10.1073/pnas.0912010107

233. Holland IB, Peherstorfer S, Kanonenberg K, Lenders M, Reimann S, Schmitt L. Type I Protein Secretion—Deceptively Simple yet with a Wide Range of Mechanistic Variability across the Family. *EcoSal Plus*. 2016;7(1). doi:10.1128/ecosalplus.ESP-0019-2015

234. Pourhassan N. Z, Smits SHJ, Ahn JH, Schmitt L. Biotechnological applications of type 1 secretion systems. *Biotechnol Adv*. 2021;53:107864. doi:10.1016/j.biotechadv.2021.107864

235. Létoffé S, Ghigo JM, Wandersman C. Secretion of the Serratia marcescens HasA protein by an ABC transporter. *J Bacteriol*. 1994;176(17):5372-5377. doi:10.1128/jb.176.17.5372-5377.1994

236. Boyd CD, Smith TJ, El-Kirat-Chatel S, Newell PD, Dufrêne YF, O'Toole GA. Structural Features of the Pseudomonas fluorescens Biofilm Adhesin LapA Required for LapG-Dependent Cleavage, Biofilm Formation, and Cell Surface Localization. *J Bacteriol*. 2014;196(15):2775-2788. doi:10.1128/JB.01629-14

237. Hanke C, Hess J, Schumacher G, Goebel W. Processing by OmpT of fusion proteins carrying the HlyA transport signal during secretion by theEscherichia coli hemolysin transport system. *Mol Gen Genet MGG*. 1992;233(1):42-48. doi:10.1007/BF00587559

238. Asada T, Takagi D, Nakai M, Abe S, Yuasa K. Secretory production of a camelid single-domain antibody (VHH, nanobody) by the Serratia marcescens Lip system in Escherichia coli. *Biochem Biophys Res Commun*. 2021;549:105-112. doi:10.1016/j.bbrc.2021.02.096

239. Lin DY wei, Huang S, Chen J. Crystal structures of a polypeptide processing and secretion transporter. *Nature*. 2015;523(7561):425-430. doi:10.1038/nature14623

240. Bajunaid W, Haidar-Ahmad N, Kottarampatel AH, et al. The T3SS of Shigella: Expression, Structure, Function, and Role in Vacuole Escape. *Microorganisms*. 2020;8(12):1933. doi:10.3390/microorganisms8121933

241. Ménard R, Sansonetti P, Parsot C. The secretion of the Shigella flexneri Ipa invasins is activated by epithelial cells and controlled by IpaB and IpaD. *EMBO J*. 1994;13(22):5293-5302.

242. Cianfanelli FR, Monlezun L, Coulthurst SJ. Aim, Load, Fire: The Type VI Secretion System, a Bacterial Nanoweapon. *Trends Microbiol*. 2016;24(1):51-62. doi:10.1016/j.tim.2015.10.005

243. Crisan CV, Hammer BK. The Vibrio cholerae type VI secretion system: toxins, regulators and consequences. *Environ Microbiol*. 2020;22(10):4112-4122. doi:10.1111/1462-2920.14976

244. Costa TRD, Harb L, Khara P, Zeng L, Hu B, Christie PJ. Type IV secretion systems: Advances in structure, function, and activation. *Mol Microbiol*. 2021;115(3):436-452. doi:10.1111/mmi.14670

245. Huang L, Boyd D, Amyot WM, et al. The E Block motif is associated with Legionella pneumophila translocated substrates. *Cell Microbiol*. 2011;13(2):227-245. doi:10.1111/j.1462-5822.2010.01531.x

246. Wagner A, Tittes C, Dehio C. Versatility of the BID Domain: Conserved Function as Type-IV-Secretion-Signal and Secondarily Evolved Effector Functions Within Bartonella-Infected Host Cells. *Front Microbiol*. 2019;10:921. doi:10.3389/fmicb.2019.00921

247. Lifshitz Z, Burstein D, Peeri M, et al. Computational modeling and experimental validation of the Legionella and Coxiella virulence-related type-IVB secretion signal. *Proc Natl Acad Sci U S A*. 2013;110(8):E707-E715. doi:10.1073/pnas.1215278110

248. Houben ENG, Korotkov KV, Bitter W. Take five — Type VII secretion systems of Mycobacteria. *Biochim Biophys Acta BBA - Mol Cell Res*. 2014;1843(8):1707-1716. doi:10.1016/j.bbamcr.2013.11.003

249. Stanley SA, Raghavan S, Hwang WW, Cox JS. Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. *Proc Natl Acad Sci*. 2003;100(22):13001-13006. doi:10.1073/pnas.2235593100

250. Phan TH, Ummels R, Bitter W, Houben ENG. Identification of a substrate domain that determines system specificity in mycobacterial type VII secretion systems. *Sci Rep*. 2017;7:42704. doi:10.1038/srep42704

251. Bunduc CM, Fahrenkamp D, Wald J, et al. Structure and dynamics of a mycobacterial type VII secretion system. *Nature*. 2021;593(7859):445-448. doi:10.1038/s41586-021-03517-z

252. Steipe B, Schiller B, Plückthun A, Steinbacher S. Sequence Statistics Reliably Predict Stabilizing Mutations in a Protein Domain. *J Mol Biol*. 1994;240(3):188-192. doi:10.1006/jmbi.1994.1434

253. Porebski BT, Buckle AM. Consensus protein design. *Protein Eng Des Sel*. 2016;29(7):245-251. doi:10.1093/protein/gzw015

254. Melman P, Roshan UW. K-means-based Feature Learning for Protein Sequence Classification. :6.

255. Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. *ArXiv13013781 Cs*. Published online September 6, 2013. Accessed April 27, 2022. http://arxiv.org/abs/1301.3781

256. Asgari E, Mofrad MRK. Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics. *PLOS ONE*. 2015;10(11):e0141287. doi:10.1371/journal.pone.0141287

257. Ranjan C, Ebrahimi S, Paynabar K. Sequence Graph Transform (SGT): A Feature Embedding Function for Sequence Data Mining. *ArXiv160803533 Cs Stat*. Published online October 4, 2021. Accessed April 27, 2022. http://arxiv.org/abs/1608.03533

Appendix I: List of primers and genes

Primer #	Name	Sequence (5'-3')
HMIO292	AIDA_I847A_F	GCG GGT AAT ACT CTT ACC GTG TCA AA
HMIO293	AIDA_846_R	ACT TTC TTT TGT AGG ATT AAG AAT GAT AT
HMIO399	AIDA_ I847A_ΔLinker_Fw	AAAGAAAGTGCGGGTAATACTCTTACCGTGTCAAATTATAC
HMIO400	AIDA_ΔLinker_Fw	AAAGAAAGTATAGGTAATACTCTTACCG
HMIO401	AIDA_ΔLinker_Rv	TTTGTCATCGTCATCTTTATAATCGC
HMIO425	pMK90_OmpTcsMyc_Fw	TCATCTCAGAAGAGGATCTGGTGAATAACAATGGAAGCATTGTC
HMIO426	pMK90_OmpTcsMyc_Rv	GTTTTTGTTCCACTTTACGCGCTTTGTCATCGTCATCTTTATAATCGC
HMIO431	OmpT_Fw(pSU2_1)	AGAGAGGGATCCAATGGAGAACTTTTATGCGG
HMIO432	OmpT_Rv(pSU2_1)	AGAGAGTCTAGACTAATTAAAATGTGTACTTAAGACCAGC
HMIO463	pMK90_Pln-423_Rv	ATTGCCATAATATTTTGCAAATGCATTTCCGATT
HMIO464	pMK90_Pln-423wt_Fw	GGCCATGGCAAATGCGATTATAAAGATGACGATGACAAAG
HMIO465	pMK90_Pln-423S27DK36N_Fw	GGCCATGGCAATTGCGATTATAAAGATGACGATGACAAAG
HMIO485	Pln423_FLAGRemove_Fw	GCGCGTAAAGTGGAACAAAAACT
HMIO487	Pln423wt_FLAGRemove_Rv	GCATTTGCCATGGCCAAAATTC
HMIO488	Pln423m_FLAGRemove_Rv	GCAATTGCCATGGCCAAAATTC

HMIO507	AIDA_GG_Fw	GAGAGAGGTCTCCTCTGGTGAATAACAATGGAAGCATTG
HMIO508	AIDA_GG_Rv	TCTCTCGGTCTCGTCTCACTTAGAAGCTGTATTTTATCCCCAG

Gene name	**Sequence (5'-3')**
AIDA	GTGAATAACAATGGAAGCATTGTCATTAATAACAGCATTATAAACGGGAATATTACGAATGATGCT GACTTAAGCTTTGGTACAGCAAAGCTGCTCTCTGCTACAGTGAATGGTAGTCTTGTTAATAACAAAA ATATCATTCTTAATCCTACAAAAGAAAGTATAGGTAATACTCTTACCGTGTCAAATTATACTGGGAC ACCGGGAAGTGTTATTTCTCTTGGTGGTGTGCTTGAAGGAGATAATTCACTTACGGACCGTCTGGTG GTGAAAGGTAATACCTCTGGTCAAAGTGACATCGTTTATGTCAATGAAGATGGCAGTGGTGGTCAG ACGAGAGATGGTATTAATATTATTTCTGTAGAGGGAAATTCTGATGCAGAATTCTCTCTGAAGAACC GCGTAGTTGCCGGAGCTTATGATTACACACTGCAGAAAGGAAACGAGAGTGGGACAGATAATAAG GGATGGTATTTAACCAGTCATCTTCCCACATCTGATACCCGGCAATACAGACCGGAGAACGGAAGT TATGCTACCAATATGGCACTGGCTAACTCACTGTTCCTCATGGATTTGAATGAGCGTAAGCAATTCA GGGCCATGAGTGATAATACACAGCCTGAGTCTGCATCCGTGTGGATGAAGATCACTGGAGGAAGAA GCTCTGGTAAGCTTAATGACGGGCAAAATAAAACAACAACCAATCAGTTTATCAATCAGCTCGGGG GGGATATTTATAAATTCCATGCTGAACAACTGGGTGATTTTACCTTAGGGATTATGGGAGGATACGC GAATGCAAAAGGTAAAACGATAAATTACACGAGCAACAAAGCTGCCAGAAACACACTGGATGGTT ATTCTGTCGGGGTATACGGTACGTGGTATCAGAATGGGGAAAATGCAACAGGGCTCTTTGCTGAAA CTTGGATGCAATATAACTGGTTTAATGCATCAGTGAAAGGTGACGGACTGGAAGAAGAAAAATATA ATCTGAATGGTTTAACCGCTTCTGCAGGTGGGGGATATAACCTGAATGTGCACACATGGACATCACC TGAAGGAATAACAGGTGAATTCTGGTTGCAGCCTCATTTGCAGGCTGTCTGGATGGGGGTTACACC GGATACACATCAGGAGGATAACGGAACGGTGGTGCAGGGAGCAGGGAAAAATAATATTCAGACAA AAGCAGGTATTCGTGCATCCTGGAAGGTGAAAAGCACCCTGGATAAGGATACCGGGCGGGAGTTCC GTCCGTATATAGAGGCAAACTGGATCCATAACACTCATGAATTTGGTGTTAAAATGAGTGATGACA GCCAGTTGTTGTCAGGTAGCCGAAATCAGGGAGAGATAAAGACAGGTATTGAAGGGGTGATTACTC

	AAAACTTGTCAGTGAATGGCGGAGTCGCATATCAGGCAGGAGGTCACGGGAGCAATGCCATCTCCG GAGCACTGGGGATAAAATACAGCTTC
EspP	GCGAATAAGGAAGCTACACGTAACGCCGCAGCTTTATTTTCGGTTGATTACAAGGCTTTCTTGAATG AAGTCAATAACTTAAATAAAAGAATGGGGGACCTGCGGGATATCAATGGAGAGGCCGGTGCCTGG GCACGTATTATGTCCGGTACCGGGTCGGCCTCTGGGGGATTTAGTGATAACTATACTCACGTACAAG TCGGTGTTGACAAGAAACATGAGCTGGACGGGTTGGACCTGTTCACAGGCTTCACGGTGACCCATA CGGACAGTTCTGCTTCCGCTGATGTGTTCAGTGGTAAGACGAAGTCGGTGGGGGCTGGGCTGTACG CTTCCGCGATGTTTGATTCCGGGGCTTATATAAATTTAATTGGTAAGTATGTGCATCACGATAACGA ATATACTGCTACTTTTGCCGGGCTTGGTACCCGTGATTACAGTACACATTCATGGTACGCAGGTGCA GAAGCGGGATACAGATACCATGTAACCGAAGATGCGTGGATAGAACCCCAGGCGGAGTTGGTATAT GGCAGCGTGTCGGGGAAGCAATTTGCGTGGAAGGATCAAGGTATGCACTTATCTATGAAAGACAAG GACTATAATCCATTGATCGGTAGAACCGGCGTGGACGTGGGAAAATCGTTCTCAGGCAAGGACTGG AAGGTTACCGCCAGAGCGGGTCTTGGATACCAGTTTGATCTGCTTGCAAATGGGGAAACGGTCCTT AGAGACGCTTCTGGCGAAAAGAGAATAAAGGGTGAGAAGGATAGTCGGATGCTTATGAGTGTAGG GCTTAACGCCGAAATTCGGGATAATGTAAGATTCGGTCTGGAATTTGAGAAATCCGCATTCGGAAA ATATAATGTGGATAATGCCGTGAACGCAAACTTCAGATATAGCTTT
MATE	CTGACTAATAATGGAACTCTGATGACGGGCATGTCAGGGCAACAGGCAGGGAATGTGCTGGTAGTA AAGGGGAATTATCATGGCAATAATGGACAACTGGTGATGAACACCGTTTTAAATGGGGATGATAGC GTTACTGACAAGTTAGTAGTAGAAGGTGATACATCTGGGACTACCGCCGTGACTGTAAACAACGCG GGCGGCACCGGCGCTAAGACTTTAAATGGGATCGAGCTTATCCATGTGGATGGCAAATCAGAGGGT GAGTTCGTACAAGCAGGGCGGATTGTGGCGGGTGCGTATGATTACACGCTGGCACGCGGACAAGGA GCGAATTCGGGAAATTGGTACTTAACTTCTGGGAGCGATTCACCAGAACTTCAACCAGAGCCAGAT CCAATGCCCAACCCAGAACCTAACCCGAACCCTGAACCCAACCCAAATCCCACTCCAACGCCCGGG CCAGATTTGAATGTGGACAACGACTTGCGGCCGGAGGCGGGTTACATCGCCAACCTGGCTGCAGCA AACACCATGTTCACAACACGTCTGCATGAGCGCCTTGGTAACACGTATTACACAGATATGGTAACTG GCGAACAAAAGCAAACAACCATGTGGATGAGACACGAGGGGGGGCATAATAAATGGAGAGATGGC TCAGGTCAACTGAAGACACAATCCAACCGCTATGTTCTTCAGTTGGGGGGAGACGTAGCCCAATGG TCCCAAAATGGGTCTGATAGATGGCATGTTGGTGTAATGGCTGGGTACGGAAACTCGGATTCAAAA

	ACTATCTCTAGTCGGACTGGATACAGAGCTAAAGCATCAGTGAACGGGTATTCCACTGGATTATAC GCTACGTGGTATGCAGACGATGAATCAAGAAACGGTGCTTACCTTGACAGTTGGGCCCAGTATTCTT GGTTCGATAATACTGTAAAGGGTGATGACCTGCAAAGTGAAAGCTACAAATCCAAGGGGTTCACTG CATCTTTAGAAGCAGGCTATAAACATAAATTGGCGGAATTCAACGGATCTCAGGGCACCCGGAACG AATGGTACGTGCAACCGCAAGCGCAGGTGACTTGGATGGGTGTAAAAGCTGATAAGCACCGGGAAT CGAATGGTACATTGGTCCATTCCAATGGCGACGGCAATGTTCAAACGCGGCTGGGCGTCAAGACAT GGTTAAAATCACACCATAAAATGGATGACGGTAAGTCAAGAGAATTTCAGCCGTTTGTAGAAGTGA ATTGGCTGCACAACAGCAAGGACTTCAGCACTAGTATGGACGGTGTCTCGGTGACACAGGATGGTG CACGTAACATCGCGGAAATTAAAACTGGGGTTGAACAGTTGAATGCAAATCTGAATGTGTGGGGTA ATGTAGGAGTGCAGGTCGCAGACCGCGGATATAACGATACGAGCGCAATGGTCGGGATTAAGTGGC AGTTT
eCPX	ATGAAAAAGATCGCATGCTTATCGGCATTGGCTGCAGTATTGGCATTCACTGCAGGTACTTCTGTTG GAGGACAAAGCGGACAGTCAGGTGACTATAACAAAAACCAATACTATGGTATTACAGCAGGGCCA GCTTATCGTATAAACGACTGGGCGTCAATTTATGGAGTCGTAGGGGTTGGGTATGGCAAGTTTCAGA CCACAGAGTATCCTACATATAAACATGATACCTCTGACTATGGGTTCTCTTACGGTGCAGGTCTTCA GTTTAATCCCATGGAGAATGTTGCGCTGGATTTCTCTTACGAGCAAAGTAGAATCCGCTCCGTTGAC GTGGGTACATGGATTCTGTCCGTCGGTTATCGCTTTGGGTCGAAATCCCGTCGTGCCGCGACCTCGA CCGTTACAGGTGGGTATGCCCAATCCGATGCTCAAGGGCAAATGAACAAGATGGGGGGCTTTAATT TGAAATATCGGTATGAAGAAGATAATTCACCTTTAGGGGTGATAGGTTCTTTCACATATACTGAAAA AAGCAGAACAGCCTCTGGTGGCCAATCGGGGCAGGGTCTG
InaQN	ATGAATTTAGATAAGGCTTTAGTTTTGCGGACATGTGCTAATAATATGGCCGACCACTGCGGATTGA TTTGGCCGGCGAGTGGTACAGTGGAATCCAAGTATTGGCAAAGTACACGTCGCCATGAGAACGGTC TTGTTGGCTTGCTTTGGGGTGCAGGCACATCAGCATTCCTGTCTGTTCATGCAGACGCGCGGTGGAT CGTACGCGAGGTTGCAGTCGCCGATATTATCAGCTTAGAAGAACCCGGAATGGTTAAATTTCCTCGG GCTGAGGTGGTTCATGTTGGGGACAGAATCAGCGCGAGTCATTTCATTTCGGCCAGACAAGCCGAC CCCGCGTCTACGCCTACTCCCACACCTACACCTATGACTGCAGCTACCCCACCTCCTACACCTGCGA CAGCCAACGTGACGTTGCCGGTCGCCGAACAGGCTTCCCACGAGGTTTTCGATGTTGCGCTGGTATC AGCCGCTGCGCCTCCCGTAAACACATTGCCTGTGACAACGCCTCAGAACTTGCAGACCGGTCTG

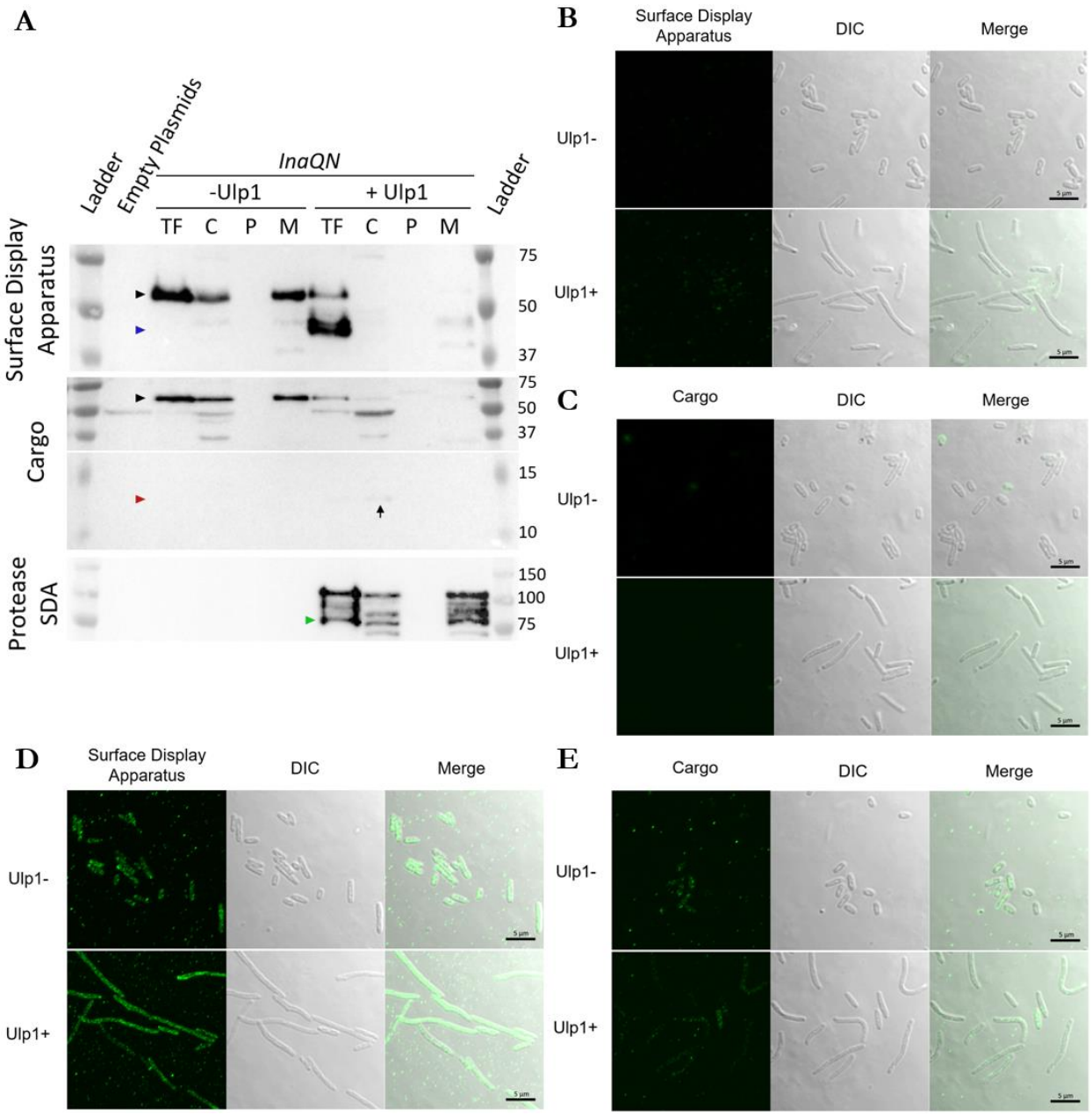
InaVN	ATGAATCTCGACAAGGCGTTGGTGCTGCGTACCTGTGCAAATAACATGGCCGATCATTGCGGCCTTA TATGGCCCGCTTCCGGGACGGTTGAATCCAAATACTGGCAGTCAACCAGGCGTCACGAGAATGGTC TGGTGGGTTTACTGTGGGGCGCTGGAACCAGCGCTTTTCTAAGCGTGCATGCCGATGCGCGATGGAT TGTCTGTGAAGTCGCCGTTGCAGACATCATCAGTCTGGATGAGCCGGGAATGGTCAAGTTTCCGCGC GCTGAGGTGGTTCATGTCGGCGACAGGATCAGCGCGTCACACTTCATTTCGGCGCGTCAGGCCGAC CCTGCATCAACGCCCCCGCCCACGTCAATGACCACGCCCCCGCCCACGCCTGCGGCAGCACATGTC ACATTACCTGTGGCGGCAAGTGTCACGTTACCGGTGGCCGAACAGGCCAGTCATGAAGTGTTCGAT GTGGCGTTGGTCATCGCAGCGGCCCCCTCAGTAAATACCCTGCCGGTAACGACGCCGCAGAATTTG CAGACCGCCACTTACGGCAGCACGTTGAGTGGCGACAACCACAGTCGGCTCATTGCCGGTTATGGC AGTAACGAGACCGCTGGCAACCACAGTGATCTGATT
Lpp-OmpA	ATGAAGGCGACCAAACTGGTGCTGGGCGCGGTTATCCTGGGTAGCACCCTGCTGGCGGGTTGCAGC AGCAACGCGAAGATCGACCAGGGCATTAACAACAACGGTCCGACCCACGAAAACCAACTGGGTGC GGGTGCGTTTGGTGGCTACCAAGTGAACCCGTATGTTGGCTTTGAAATGGGTTACGATTGGCTGGGT CGTATGCCGTATAAAGGCAGCGTGGAGAACGGTGCGTACAAGGCGCAGGGCGTTCAACTGACCGCG AAACTGGGTTACCCGATCACCGACGATCTGGACATTTATACCCGTCTGGGTGGCATGGTGTGGCGTG CGGACACCAAAAGCAACGTTTACGGCAAGAACCACGATACCGGTGTGAGCCCGGTTTTTGCGGGTG GCGTGGAATATGCGATCACCCCGGAGATTGCGACCCGTGGTATTCCGGGTAAGCTTGGGTCCCTGC AGGGTCTG
YfaL	GGTACGACGGAGCTTTCAAACGTCACGGTTAACGGCAATTTGACTAATACTAGCGGTGCGGTTAGC CTGCAAAACGGCGTCGCTGGCGACACCCTTACCGTAAATGGAGATTATACAGGAGGAGGGACCTTG CTTTTAGACTCGGAGCTGAATGGAGATGACAGCGTCTCGGATCAGCTTGTAATGAATGGAAACACG GCGGGAAACACCACCGTTGTTGTAAACAGTATCACGGGCATTGGTGAACCTACATCTACTGGAATT AAGGTGGTTGACTTCGCTGCAGATCCTACGCAGTTTCAGAACAACGCCCAATTCTCGTTAGCCGGAT CTGGGTATGTGAATATGGGCGCATACGACTACACATTGGTGGAGGATAACAACGATTGGTATTTAC GTTCTCAAGAAGTGACACCGCCGTCGCCTCCCGATCCTGATCCAACACCTGATCCGGACCCAACGCC AGATCCTGATCCAACTCCGGACCCGGAACCTACGCCCGCATATCAGCCTGTTTTGAATGCAAAGGTA GGTGGATATTTGAATAACCTTCGTGCAGCTAATCAGGCATTCATGATGGAACGCCGCGATCACGCTG GGGGTGACGGACAAACATTGAATCTTCGTGTGATCGGGGGAGATTATCATTACACGGCCGCAGGTC

	AGTTGGCACAACACGAAGATACTAGCACGGTGCAGCTTAGCGGAGACTTATTTTCAGGACGCTGGG
	GCACCGACGGCGAGTGGATGCTTGGAATTGTGGGTGGGTATAGCGATAACCAGGGAGATTCTCGTT
	CGAACATGACCGGTACTCGCGCGGACAACCAGAATCACGGTTACGCAGTCGGGCTGACGTCTAGCT
	GGTTTCAGCATGGGAACCAGAAACAAGGGGCTTGGCTTGATTCATGGCTGCAGTATGCCTGGTTTA
	GCAACGACGTCAGCGAACAAGAGGATGGTACAGATCATTATCATAGCTCTGGCATTATCGCGTCAC
	TTGAAGCAGGATATCAATGGCTTCCCGGTCGTGGCGTTGTGATTGAACCTCAAGCACAGGTAATCTA
	CCAAGGCGTGCAGCAGGATGATTTTACGGCGGCAAACCGTGCGCGCGTTTCTCAGTCGCAGGGAGA
	TGATATCCAGACTCGTCTTGGATTACACAGTGAATGGCGTACTGCTGTTCACGTTATTCCCACGCTT
	GACTTGAATTACTACCATGACCCGCACTCGACAGAAATCGAAGAAGACGGGTCTACTATTTCGGAT
	GATGCGGTTAAGCAACGCGGTGAGATCAAGGTAGGCGTGACGGGAAATATCAGTCAGCGTGTATCC
	TTGCGTGGAAGTGTGGCGTGGCAAAAAGGTTCTGACGACTTCGCACAGACTGCAGGATTCTTATCA
	ATGACTGTGAAATGG
	CATGGCACCCCGCAGAACATTACCGATCTGTGCGCGGAATATCATAACACCCAGATTTATACCCTGA
	ACGATAAAATTTTTAGCTATACCGAAAGCCTGGCGGGCAAACGCGAAATGGCGATTATTACCTTTA
CtxB	AAAACGGCGCGATTTTTCAGGTGGAAGTGCCGGGCAGCCAGCATATTGATAGCCAGAAAAAAGCG
	ATTGAACGCATGAAAGATACCCTGCGCATTGCGTATCTGACCGAAGCGAAAGTGGAAAAACTGTGC
	GTGTGGAACAACAAAACCCCGCATGCGATTGCGGCGATTAGCATGGCGAAC
	GGCGAAAGCTTGTTTAAAGGTCCGCGCGACTACAACCCGATCAGCTCCTCAATTTGCCATCTGACGA
	ACGAAAGTGATGGACATACAACGAGCCTGTATGGCATCGGTTTTGGGCCGTTCATAATCACAAATA
	AACACTTGTTCAGACGTAATAATGGAACATTGTTAGTGCAATCACTGCATGGAGTCTTTAAGGTCAA
	GGACACGACGACTCTTCAACAGCACCTTGTCGATGGGAGAGATATGATCATTATCCGGATGCCAAA
	GGATTTCCCGCCTTTTCCGCAGAAGTTGAAGTTCCGTGAGCCACAACGCGAGGAGCGGATATGTTTG
TEV	GTAACCACTAACTTTCAGACAAAGAGCATGTCGTCTATGGTATCAGATACGTCGTGTACGTTCCCAT
	CGAGTGATGGCATCTTCTGGAAGCACTGGATTCAGACGAAAGACGGCCAATGTGGCTCGCCATTGG
	TGTCCACGCGTGACGGCTTCATAGTGGGCATTCATAGCGCAAGTAACTTTACTAATACCAATAACTA
	CTTCACGTCAGTGCCAAAGAATTTCATGGAACTGCTGACTAATCAAGAAGCCCAGCAGTGGGTATC
	CGGATGGAGACTTAACGCGGACTCGGTGCTTTGGGGGGGTCATAAGGTTTTTATGAATAAGCCAGA
	AGAACCCTTCCAGCCCGTCAAGGAGGCTACTCAGTTGATGAATGAGTTGGTTTACTCCCAG

Ulp1	CTGGTCCCTGAGCTTAATGAAAAGGATGACGACCAGGTTCAGAAGGCACTTGCGTCTCGGGAGAAT ACCCAGTTGATGAACAGAGACAATATTGAGATTACAGTTCGGGATTTTAAGACCCTTGCACCTCGCC GTTGGTTAAACGACACCATAATAGAATTTTTTATGAAATATATAGAGAAATCTACCCCGAACACAGT TGCGTTCAACTCGTTTTTTTATACCAACTTGTCCGAGCGTGGGTACCAGGGAGTTCGCCGTTGGATG AAGCGTAAGAAGACCCAAATTGATAAATTGGACAAGATCTTTACTCCGATAAACTTAAATCAAAGT CACTGGGCTCTTGGCATAATCGACCTGAAAAAGAAGACGATAGGCTATGTTGATTCCCTTTCAAACG GTCCTAATGCTATGTCATTTGCTATACTGACTGATTTACAGAAGTATGTCATGGAGGAATCCAAGCA CACCATTGGGGAGGATTTCGATCTGATTCATCTTGATTGCCCACAGCAGCCCAACGGCTATGACTGC GGAATTTATGTCTGCATGAATACGCTGTATGGATCAGCAGACGCTCCACTGGATTTTGACTACAAAG ACGCGATTCGGATGCGTAGATTCATCGCACACCTGATCCTTACTGACGCTTTGAAA
SUMO	TCCGACTCAGAAGTCAATCAAGAAGCTAAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTCAC ATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTCCTTTAA GAAGGCTGATGGAAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCTTAAGATTCTTGTACG ACGGTATTAGAATTCAAGCTGATCAGGCCCCTGAAGATTTGGACATGGAGGATAACGATATTATTG AGGCTCACCGCGAACAGATTGGAGGT
TEVcs	GAAAACCTGTATTTTCAGGGC
AIDA SP	ATGAATAAGGCCTACAGTATCATATGGAGCCACTCCAGACAGGCCTGGATTGTGGCCTCAGAGTTA GCCAGAGGACATGGTTTTGTCCTTGCAAAAAATACACTGCTGGTATTGGCGGTTGTTTCCACAATCG GAAATGCATTTGCA
YfaLSP	ATGCGCATTATTTTTCTGCGCAAAGAATATCTGAGCCTGCTGCCGAGCATGATTGCGAGCCTGTTTA GCGCGAACGGCGTGGCGGCG
Myc	GAACAAAAACTCATCTCAGAAGAGGATCTG
FLAG	GATTATAAAGATGACGATGACAAA
Ollas	AGCGGTTTTGCAAACGAACTGGGTCCGCGCTTAATGGGCAAA

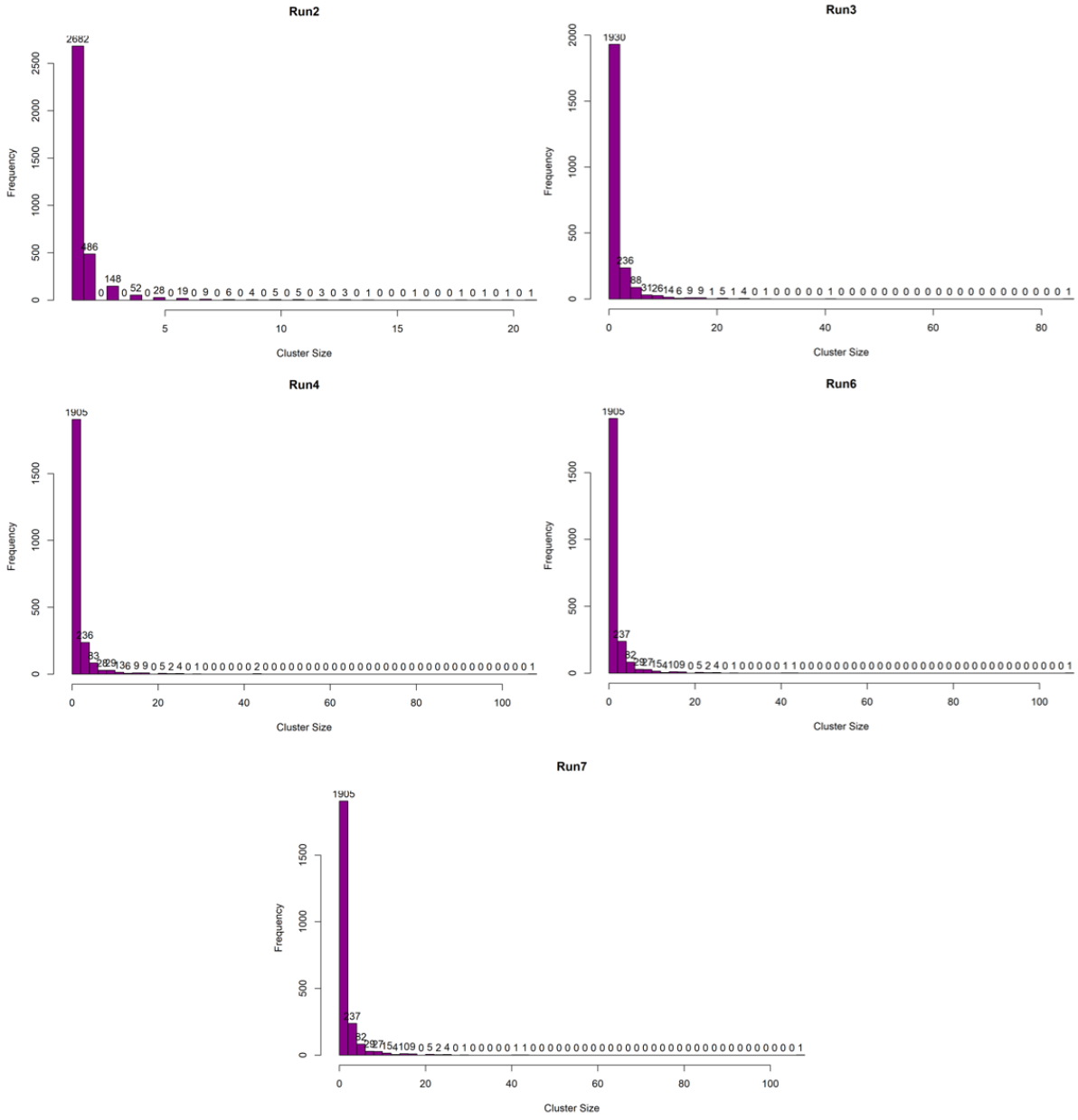
plac-lacO-RBS	TTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACT

Appendix II: Supplementary figures



Supplementary figure 1. Functional study of the InaQN/Ulp1 secretion system. A) Three separate WBs of cell fractions. TF = Whole cell lysate with media, C = Cytoplasm, P = Periplasm, M = Membranes. The arrow points to a band of interest that only appeared faintly. The black arrowhead points at the full-sized SDA (~55 kDa). The blue arrowhead points at the cleaved TMD of the SDA (~42 kDa). The red arrowhead points at the cleaved cargo (~13 kDa). The green arrowhead points at the full-size protease SDA (~77 kDa).

Numbers on the right side show the weights of the ladder in kDa. **B-E)** Fluorescence microscopy imaging of immunostained fixed. In **D)** and **E)**, the bacteria were permeabilized to allow penetration of the antibodies. See Fig. 10 for more details.

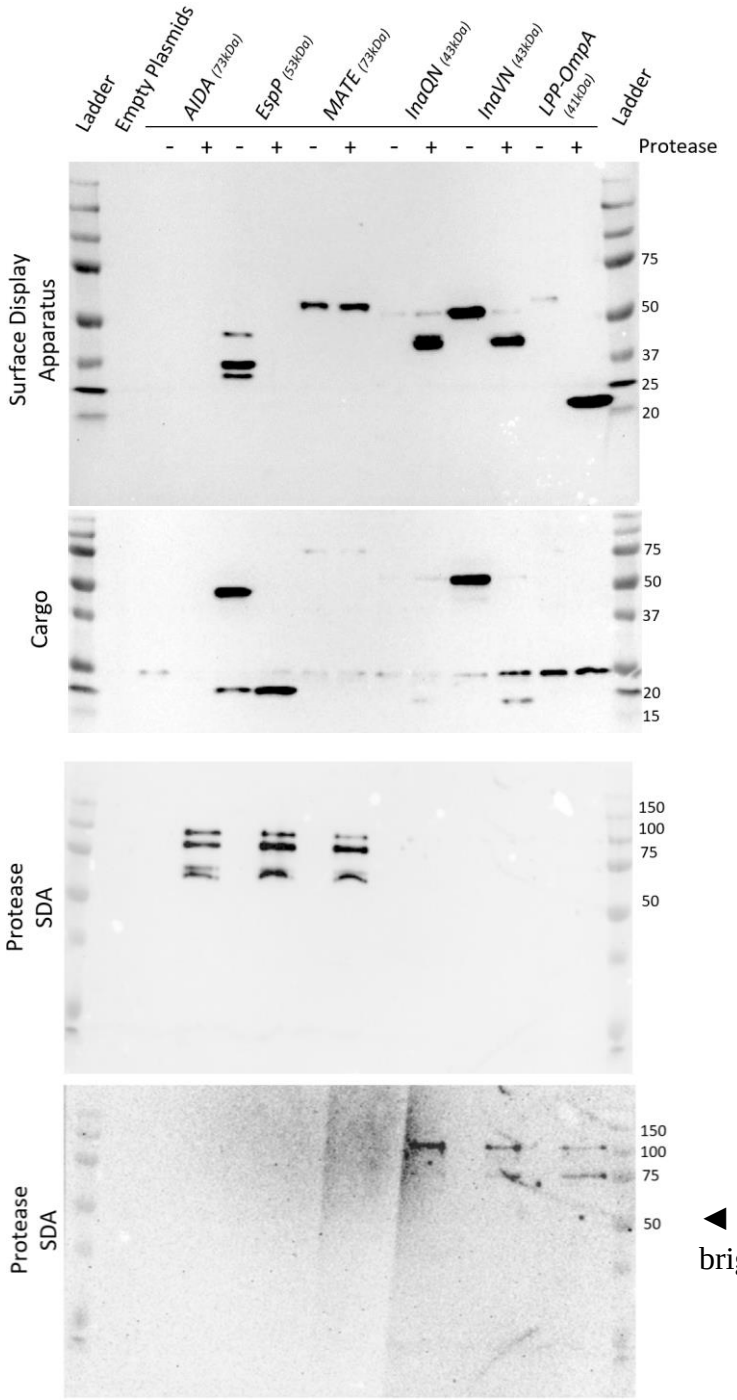


Supplementary figure 2. Cluster size distribution using alternative Linclust parameters tested. 4996 FASTA sequences from the DRAMP database were inputted using coverage mode 1. Refer to Table 4 for the parameters used.

Appendix III: Raw data

Uncropped Wester blots:

Figure 9:



◄ Longer exposition time with the brighter bands covered

Figure 10:

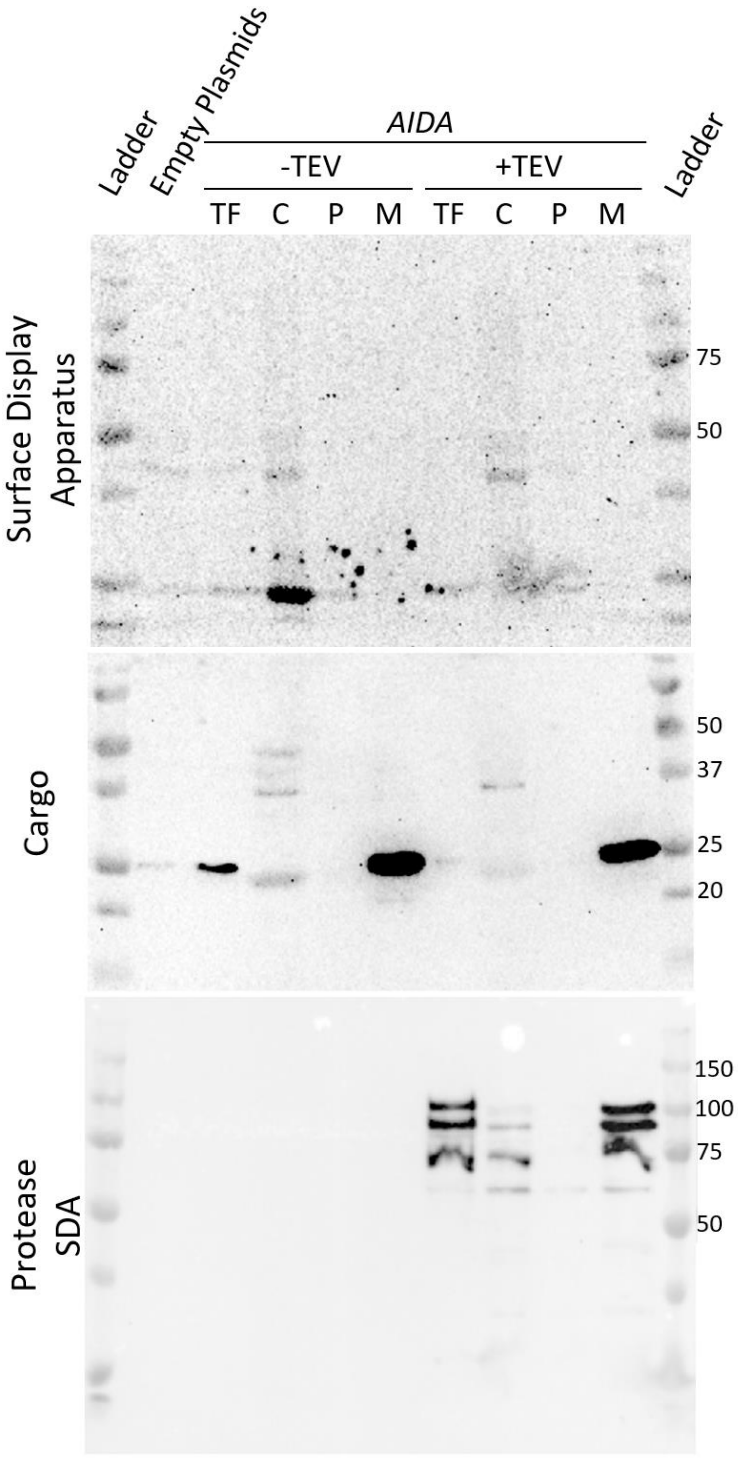

Ladder
Empty Plasmids
AIDA
-TEV
+TEV
TF C P M TF C P M
Ladder
Surface Display Apparatus
75
50
Cargo
50
37
25
20
Protease SDA
150
100
75
50

Figure 11:

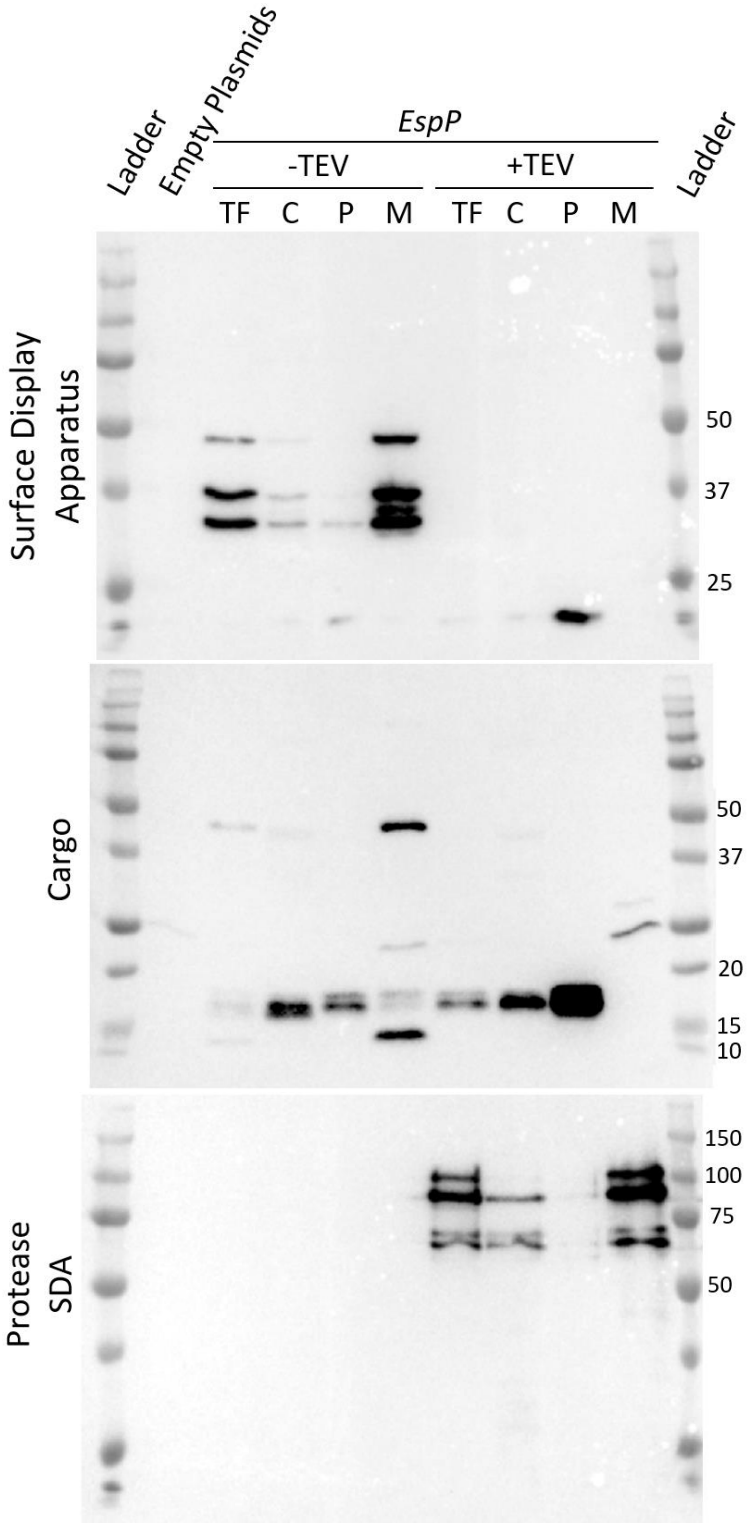

Ladder
Empty Plasmids
EspP
-TEV
+TEV
TF C P M TF C P M
Ladder
Surface Display Apparatus
50
37
25
Cargo
50
37
20
15
10
Protease SDA
150
100
75
50

Figure 12:

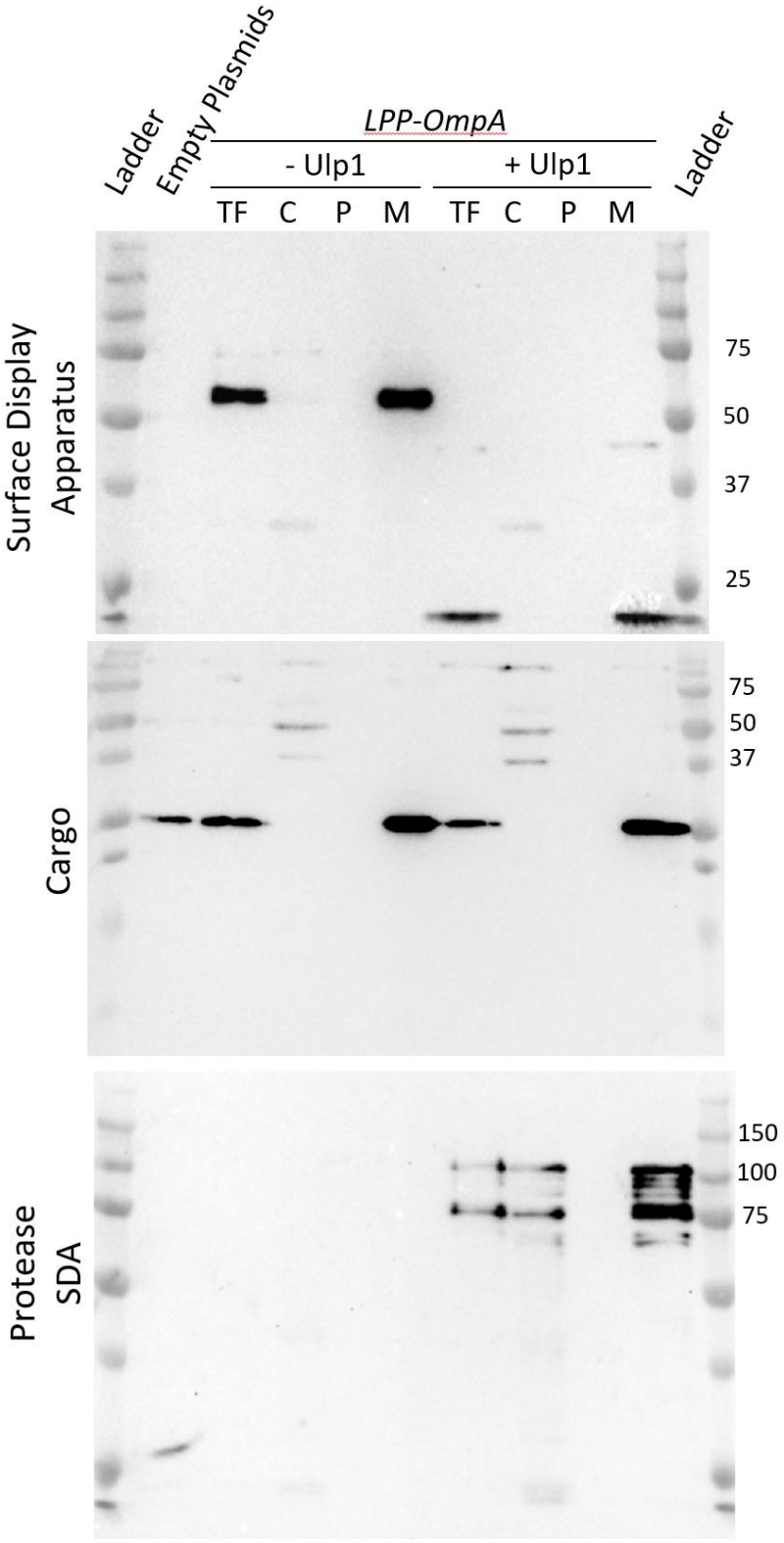

Ladder
Empty Plasmids
LPP-OmpA
- Ulp1
+ Ulp1
TF C P M TF C P M
Ladder
Surface Display Apparatus
75
50
37
25
Cargo
75
50
37
Protease SDA
150
100
75

Figure 13:

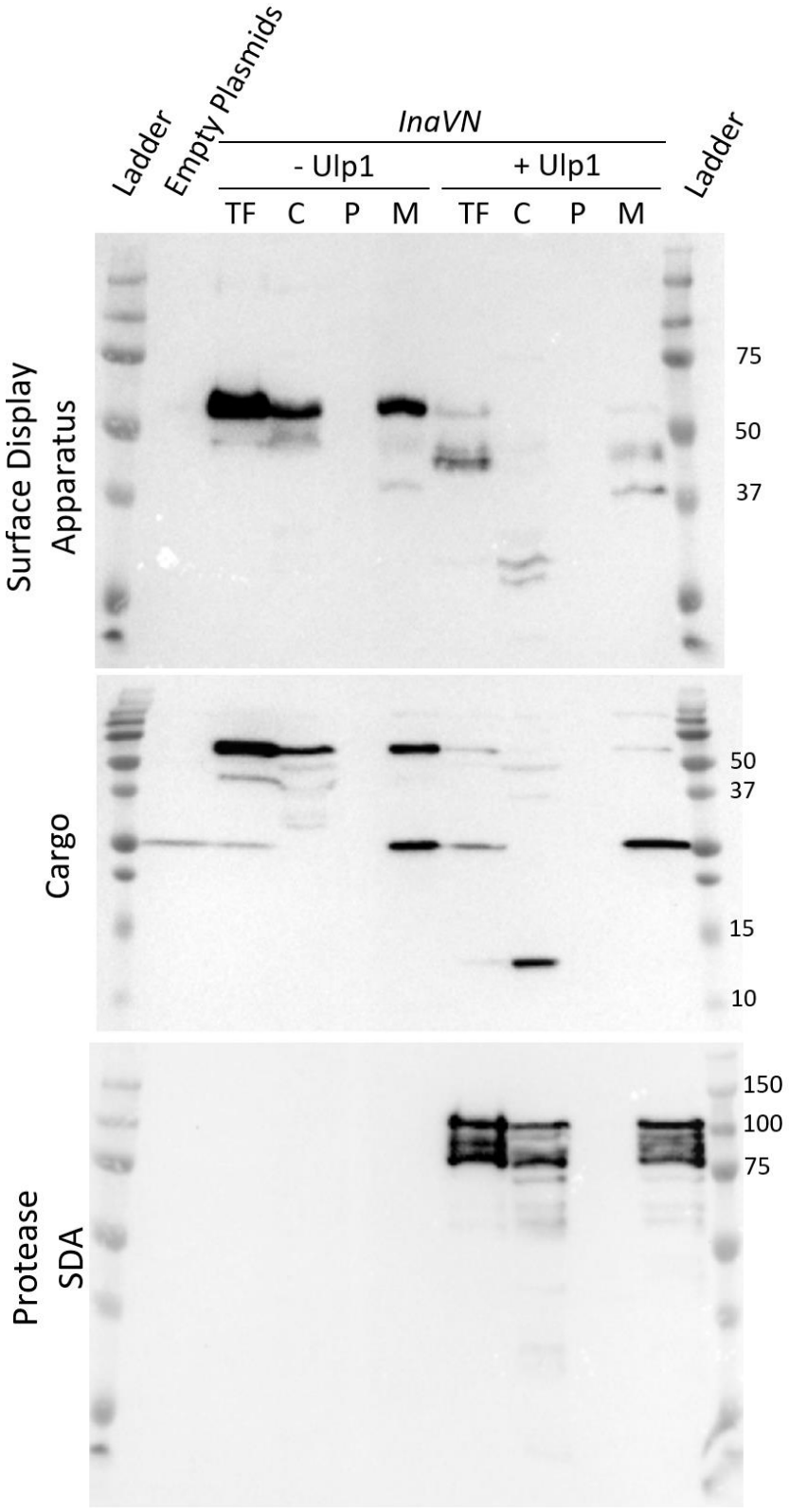

Ladder
Empty Plasmids
InaVN
- Ulp1
+ Ulp1
TF C P M TF C P M
Ladder
Surface Display Apparatus
75
50
37
Cargo
50
37
15
10
Protease SDA
150
100
75

Figure 14:

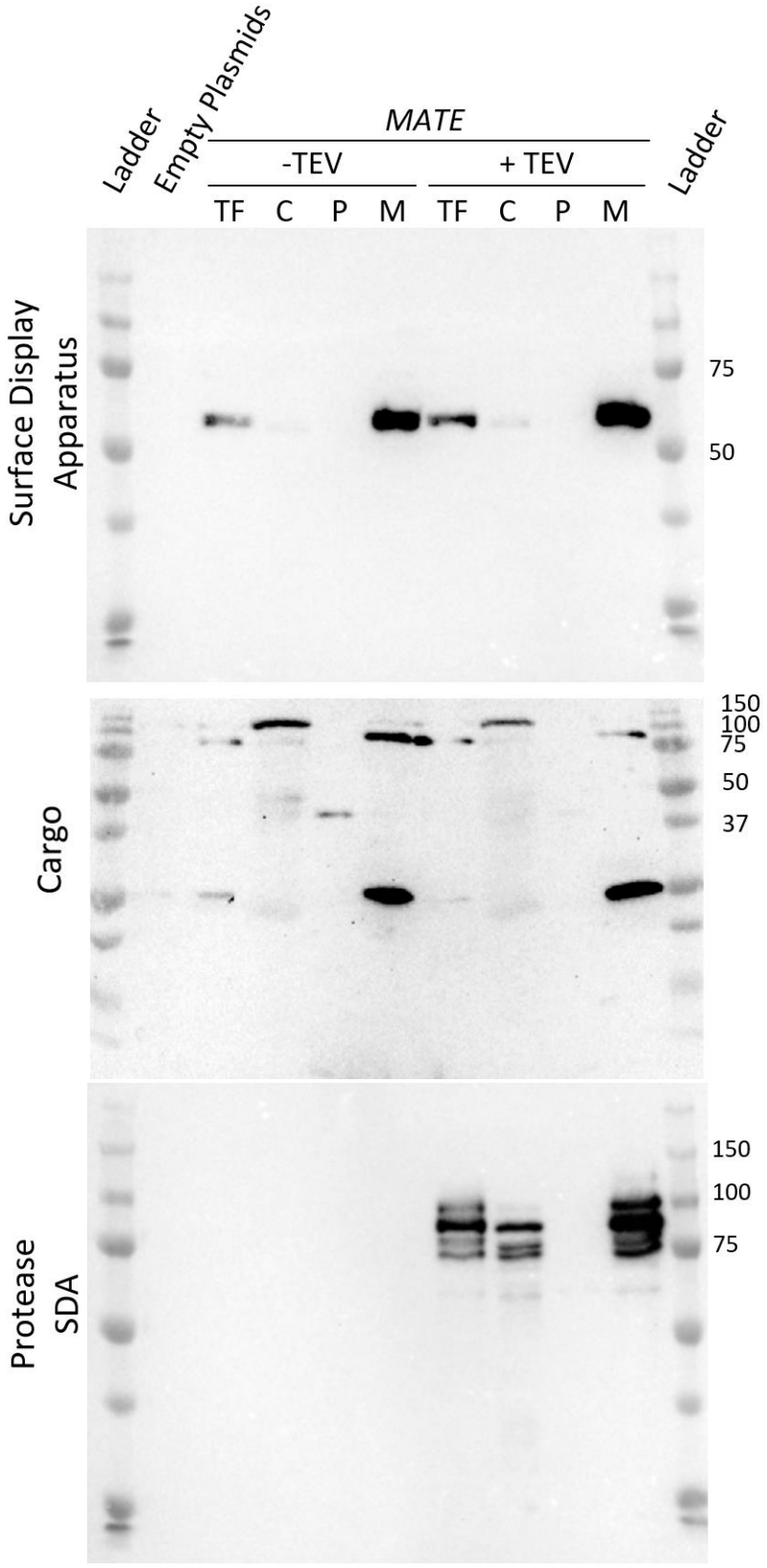
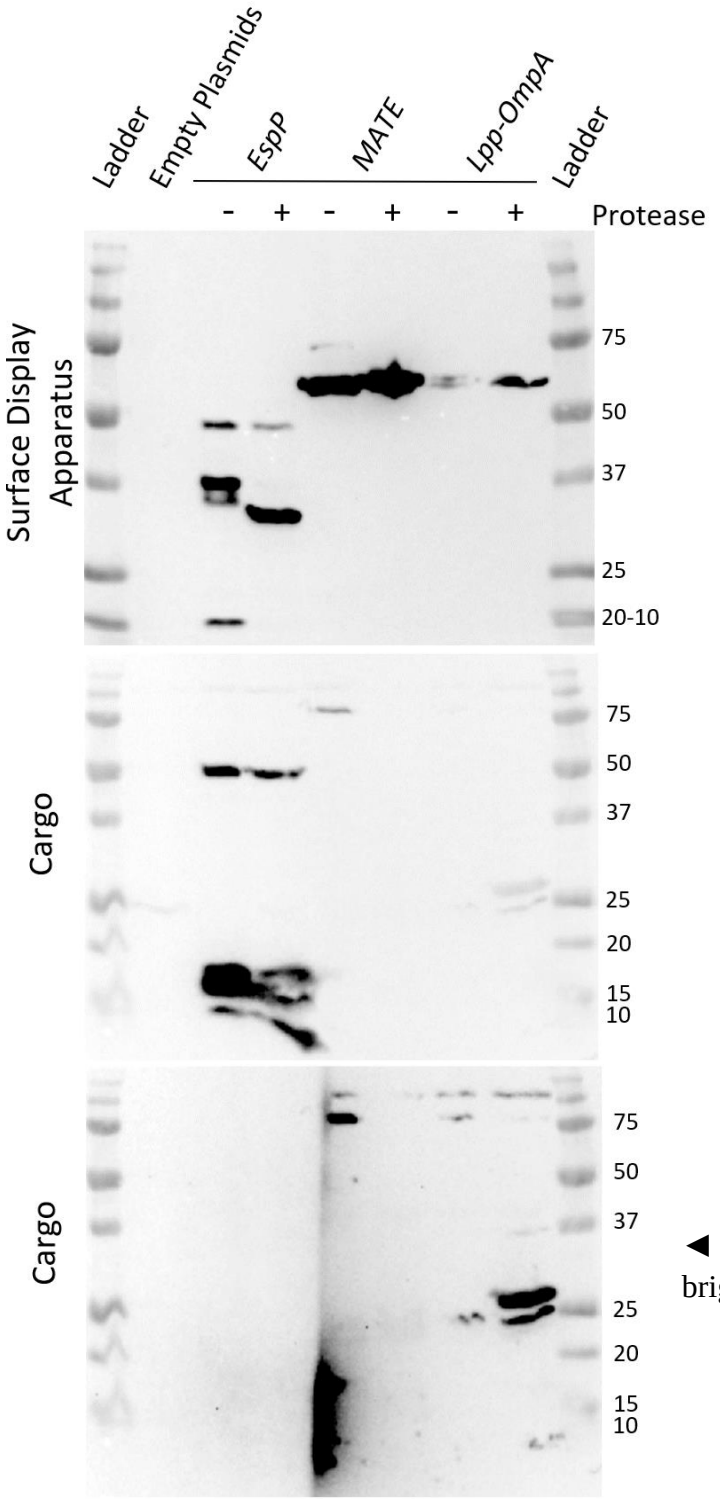

Ladder
Empty Plasmids
MATE
-TEV
+ TEV
TF C P M TF C P M
Ladder
Surface Display Apparatus
75
50
Cargo
150
100
75
50
37
Protease SDA
150
100
75

Figure 16:

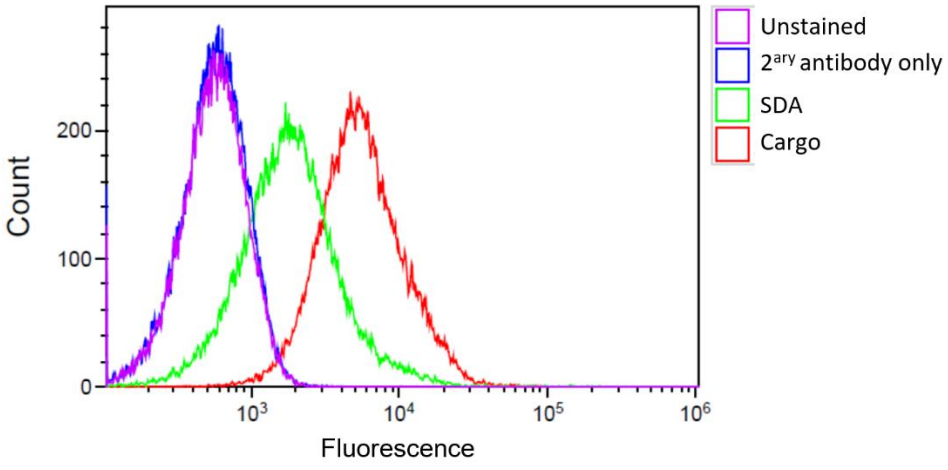


◄ Longer exposition time with the brighter bands covered

Sample flow cytometry graphs:

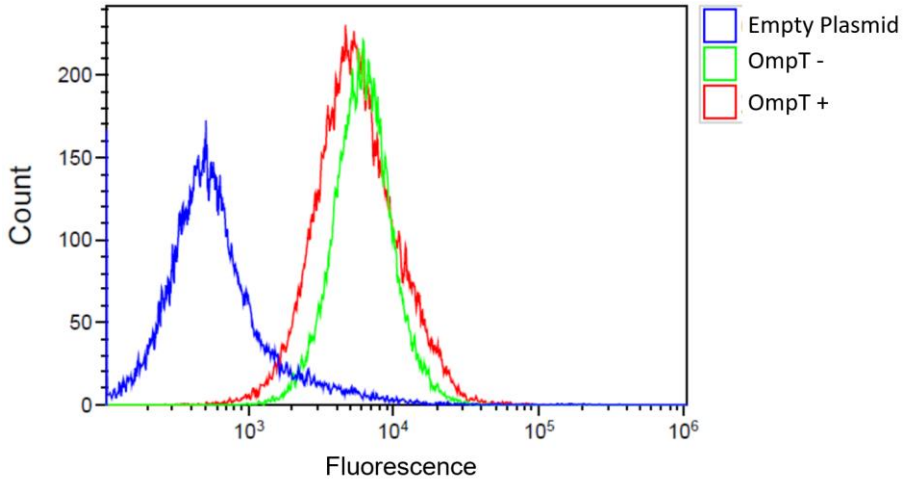
Figure 6 D)-E):

Each strain is analysed with two controls: an unstained control and a secondary antibody control where the primary antibody is not used. SDA = Anti-myc primary antibody; Cargo = Anti-FLAG primary antibody. The listed valued are the median fluorescence. Here is a fluorescence histogram for the OTII secreting strain with OmpT expression:

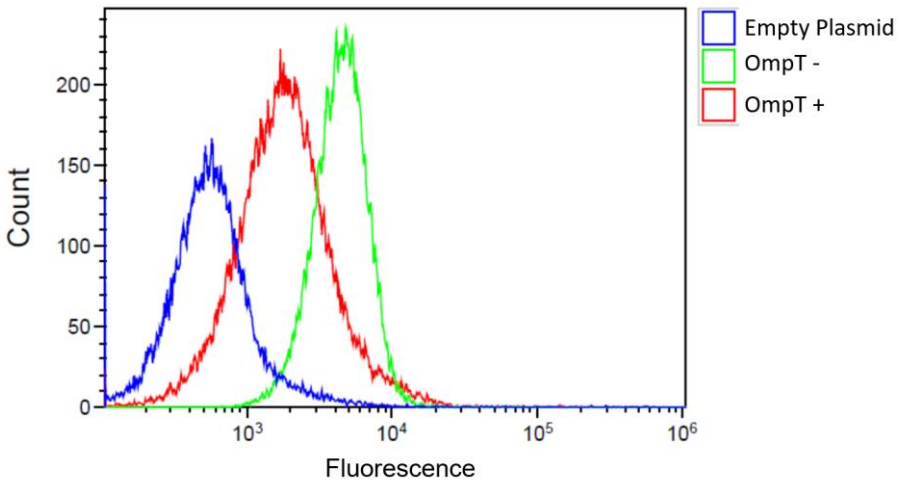


Marker	X-Med
All	564.1
All	572.6
All	1,784.0
All	5,327.5

The secreting and non-secreting strains are then compared to each other along with an empty plasmid control. The following are two fluorescence histograms for the OTII constructs following the SDA (top) and the Cargo (bottom) respectively:



Marker	X-Med
All	530.9
All	6,093.4
All	5,327.5



Marker	X-Med
All	554.6
All	4,458.5
All	1,784.0